Sedimentation Equilibrium Analysis of Protein Interactions with Global Implicit Mass Conservation Constraints and Systematic Noise Decomposition

Jennifer Vistica [a], Julie Dam [b], Andrea Balbo [a], Emine Yikilmaz [c], Roy A. Mariuzza [b], Tracey A. Rouault [c], and Peter Schuck [a]*

[a] *Protein Biophysics Resource, Division of Bioengineering & Physical Science, ORS, OD, National Institutes of Health, Bethesda, Maryland, USA,* [b] *Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, Maryland, USA,* [c] *Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, USA*

Running Title: SEDIMENTATION EQUILIBRIUM

Keywords: protein interactions, heterogeneous association, protein complexes, analytical ultracentrifugation, molar mass distribution

Category: Physical Techniques

* Corresponding Author:

Dr. Peter Schuck
National Institutes of Health
Bldg. 13, Rm. 3N17
13 South Drive
Bethesda, MD 20892-5766,USA
Phone: 301 435-1950
Fax: 301 480-1242
Email: pschuck@helix.nih.gov

ABSTRACT

Sedimentation equilibrium is a powerful tool for the characterization of protein self-association and heterogeneous protein interactions. Frequently, it is applied in a configuration with relatively long solution columns and with equilibrium profiles being acquired sequentially at several rotor speeds. The present study proposes computational tools, implemented in the software SEDPHAT, for the global analysis of equilibrium data at multiple rotor speeds, multiple concentrations, and multiple optical detection methods. The detailed global modeling of such equilibrium data can be a non-trivial computational problem. It was shown previously that mass conservation constraints can significantly improve and extend the analysis of heterogeneous protein interactions. Here, a method for using conservation of mass constraints for the macromolecular redistribution is proposed in which the effective loading concentrations is being calculated from the sedimentation equilibrium profiles. The approach is similar to that described by Roark (Biophys. Chem. 5 (1976) 185-196), but its utility is extended by determining the bottom position of the solution columns from the macromolecular redistribution. For analyzing heterogeneous associations at multiple protein concentrations, additional constraints are introduced that relate the effective loading concentrations of the different components or their molar ratio in the global analysis. Equilibrium profiles at multiple rotor speeds also permit the algebraic determination of radial-dependent baseline profiles, which can govern interference optical ultracentrifugation data, but usually also occur, to a smaller extent, in absorbance optical data. Finally, the global analysis of equilibrium profiles at multiple rotor speeds with implicit mass conservation and computation of the bottom of the solution column provides an unbiased scale for determining molar mass distributions of non-interacting species. The properties of these tools are studied with theoretical and experimental data sets.

Introduction

The major focus of modern sedimentation equilibrium (SE) analytical ultracentrifugation is the determination of the stoichiometry and thermodynamic equilibrium binding constants of protein interactions. It is a classical technique of physical biochemistry (1, 2), which is very versatile and is being applied to a wide variety of biological systems. One of the appealing features of SE analytical ultracentrifugation is that it observes the proteins in dilute buffer solution, without requiring interactions with any matrix or surface, or a comparison with any standards. Usually no protein modification is necessary. Recent reviews, and introductory texts on principle and practice can be found in (3-12).

During the last decades, the use of SE for characterizing reversible protein interactions was greatly facilitated by increased computational power. For interacting macromolecules, although the concentration gradients generated in the gravitational field produce a continuous range of populations of free species and complex, the analysis usually still requires the combination of information from multiple experiments at different loading concentrations, molar ratios, and/or multiple rotor speeds (13). In ideal SE, macromolecular mixtures assume a multi-exponential radial concentration distribution. Each exponential corresponds to a species in solution, with the molar mass being proportional to the exponent, or curvature, and each species concentration proportional to the amplitude of the exponential signal component. As a consequence, the data analysis requires unraveling the multi-exponential experimental data sets into contributions of exponentials corresponding to each free and complex species, usually in the context of a global analysis to determine the stoichiometry and binding constant. This is a notoriously difficult problem (14). In particular for heterogeneous protein interactions, the error surface of the corresponding modeling problem usually exhibits many local minima, and global optimization can result in physically unrealistic parameter combinations and can be ill-conditioned (15-17). This problem can sometimes be avoided by only extracting a weight-average molar mass from each experimental gradient, and by

separately modeling the isotherm of the average molar mass as a function of the known loading composition (for example, see (18)). However, this approach could potentially neglect a substantial amount of information if it were applied to long column SE data, the focus of the present work. On the other hand, one conceptual disadvantage of the global multi-exponential decomposition is that the (ordinarily at least approximate) knowledge of the amount of loaded material is commonly not utilized, instead only relying on the total measured signal distributions in SE. This can be a significant deficit, in particular if the macromolecular components cannot be spectrally distinguished, and/or if some of the species do not appreciably contribute to the measured signal, which leads to an underdetermined analysis problem.

Integration of the SE profiles over the entire solution column provides estimates of the total mass of soluble protein, which can be compared with the total amount loaded (1). Several authors have considered conservation of mass (MC) as a constraint to improve the data analysis. Fujita has pointed out the difficulty that for reacting systems in SE, mass action law does not hold for the mean concentrations (19). In 1976, Roark used analytical expressions for mass conservation (MC) to transform the local concentrations to effective loading concentrations in conjunction with corrections of the binding constants. This permitted the use of effective loading concentrations as fitting parameters and MC to constrain the global analysis of data acquired at different rotor speeds. This method, termed 'component-constrained' multi-speed fit, greatly enhanced the data analysis (13). Nichol et al. have incorporated MC into a data transformation and extrapolation method for the characterization of heterogeneous interactions (20). In 1982, in studies of ligand-induced association of tRNA, Pörschke and co-workers used analytical MC expressions to eliminate free parameters in the global multi-exponential modeling of SE profiles of a ligand-induced monomer-dimer association (21) and an isodesmic association (22). In 1986, a similar approach was described by Lewis for the characterization of ricin subunit association, termed 'method of implicit constraints' (23). Similarly, the known total loading concentration of a protein component not contributing to the measured

absorbance signal was used in a study of its interaction with a detected protein component (24). Such analytic substitution of concentration parameters was also applied later by Lewis and co-workers to the analysis of a mixed association of HIV reverse transcriptase (15). Besides reducing the number of free parameters, it was noted that MC constraints also improve the shape of the error surface to have a well-defined minimum, eliminating fits with physically unrealistic parameter combinations.

While this method was very successful in these studies, the drawback of this approach in general is that the mass balance consideration hinges on the knowledge of the precise location of the boundaries of the solution column, which unfortunately, due to optical artifacts, are not directly visible. In particular the bottom position of the solution column is critical. Most of the protein will accumulate at this interface and inaccuracies in the estimated bottom position will greatly affect the mass balance, especially when using relatively steep gradients, which frequently are preferable due to their higher information content (25). Maybe partially because of this, MC does experimentally not appear to be always fulfilled (17, 21). Other reasons for violations of MC in the SE experiment could be slight loss of material due to adsorption onto the centerpiece or windows (26), slow partial aggregation and pelleting at the bottom of the cell, limited solubility, limitations in the wavelength accuracy of the absorption optical system, or simply insufficiently precise estimates of the loaded protein concentration (21). The question whether MC is apparently fulfilled or violated is also related to the assignment of the baseline signal, which, even in the absorption optical system, may not be negligible due to technical imperfections of the absorbance scanner or possibly unmatched or variable buffer absorbance.

Philo has introduced a method for using MC constraints that do not necessarily need to be strictly fulfilled but gradually penalize a model for increasing violation of mass conservation, thus termed ‘soft’ mass conservation (17). This elegant method addresses the problem mentioned above that the MC condition may not be precisely true, but it will reshape the global optimization surface of the multi-exponential SE model such that physically unreasonable combinations of fitting parameters are

avoided. The MC constraint is used here similarly as is the prior knowledge in regularization terms for ill-conditioned inverse problems (27, 28). The penalty term can be continuously made more stringent or be relaxed, even abandoned completely when the parameters have converged to a minimum with physically reasonable values (17). This method was used extensively for receptor-ligand interactions (29-31). Recently, the strategy to use the degree of violations of MC for selecting among alternative models for SE data was also used by Arkin and Lear in a separate post-fitting step (32).

In the present paper, we propose an extension of the original method of Roark (13) using implicit MC constraints specifically for the global analysis of SE profiles of the same protein mixture at different rotor speeds. It could be termed a 'soft' MC in a different sense, because it does not require knowledge of loading concentration and permits initial loss, such as adsorption to the cell components. It is based on the assumption that the total mass of observed soluble material at the first rotor speed does not change at all other rotor speeds. To overcome the problem of strict MC to possibly introduce bias from errors in the estimated bottom position of the solution column and the baseline signal, it does not assume knowledge of either of these parameters; they are estimated in the process of globally modeling the data. Again, it is utilized that for the equilibria attained for the same solution column at different rotor speeds, the bottom position (and usually the baseline) does not change. This approach is studied with experimental and simulated data, and the resulting parameter correlations and their robustness with regard to actual mass conservation are examined.

Three additional aspects of MC analysis of SE data at multiple rotor speeds are introduced: First, to extend the utility to interference optical SE data as well as to account for optical imperfections of absorption data, we demonstrate that a radial-dependent baseline profile can be calculated from such data sets. A mathematical approach similar to the algebraic noise decomposition of sedimentation velocity (33) and pseudo-absorbance (34) transport experiments is taken. It is shown that with suitable experimental data sets, the combination with MC analysis will result in a stable analysis with well-determined macromolecular parameters, and with realistic baseline profiles. Second, it is shown that

the MC constraints can be used to generate a rational scale for the determination of molar mass distributions from SE profiles at multiple rotor speeds. This may be useful in a model-free analysis of SE of mixtures of non-interacting macromolecular species. Finally, the present paper is intended to serve as an introduction to the use of the software SEDPHAT for SE analysis, as an extension of the previous report on using SEDPHAT for global sedimentation velocity analysis of self-associating proteins (35).

Theory and Modeling

Sedimentation Equilibrium

In SE, a single ideally sedimenting species of molar mass M assumes a Boltzmann distribution of the well-known form

$$A_r = c_{r_0} e d \exp\left\{ M(1-\bar{v}\rho)\frac{\omega^2}{2RT}\left(r^2 - r_0^2\right)\right\} \quad (1)$$

(1), where r denotes the distance from center of rotation, r_0 an arbitrary reference radius, ω the angular velocity and T the absolute temperature of the rotor, R the gas constant, $\bar{v}$ the partial-specific volume, ρ the solvent density, ε the extinction coefficient (or analogous signal increment), d the optical pathlength, and c_{r0} the concentration at the reference radius. The concentration at the reference radius is related to the loading concentration c^{load} by

$$\begin{aligned} c^{load} &= c_{r_0} c \\ c &= \frac{1}{V^{load}} \int_V \exp\left\{ M(1-\bar{v}\rho)\frac{\omega^2}{2RT}\left(r^2 - r_0^2\right)\right\} dV \end{aligned} \quad (2)$$

where V^{load} denotes the volume of the solution column, and χ is a dimensionless factor that depends on the geometry of the solution column, the rotor speed, the reference radius, and the molar mass of the solute. This relationship may be used either to predict c_{r0} for a given loading concentration, or *vice*

versa, to calculate an 'effective' loading concentration after determining c_{r0} through a fit of the data. For the standard sector-shaped solution column extending from the meniscus at radius m to the bottom at radius b, an analytical expression of Eq. (2) is known

$$c = \left(M(1-\bar{v}\,r)\frac{w^2}{2RT}(b^2-m^2) \right)^{-1} \times \left[e^{M(1-\bar{v}\,r)\frac{w^2}{2RT}(b^2-r_0^2)} - e^{M(1-\bar{v}\,r)\frac{w^2}{2RT}(m^2-r_0^2)} \right] \tag{3}$$

which can be used to express c_{r0} as a function of the loading concentration. For the rectangular six-channel and for the external loading centerpiece numerical integration of Eq. (2) is required (for the latter also corrections for the use of an oil layer at the base of the cell may be required).

For a mixture of two interacting, but otherwise ideally sedimenting species A and B, we can express the SE signal as

$$A_r = \sum_{i,j}\left[c_{ij,r_0}(c_A^{tot}, c_B^{tot})(ie_A + je_B)\,d \times \exp\left\{ \left[iM_A(1-\bar{v}_A r) + jM_B(1-\bar{v}_B r) \right] \frac{w^2}{2RT}\left(r^2 - r_0^2\right) \right\} \right] \tag{4}$$

(assuming the absence of pressure dependence of the interaction and the absence of hyper- or hypochromicity). The concentration $c_{ij,r0}$ of the complex A_iB_j at the reference radius r_0 is dependent on the total loading concentrations of the species A and B, c_A^{tot} and c_B^{tot}, respectively. It can be calculated by considering mass action law locally at each radial point and mass conservation for the entire solution column

$$\begin{aligned} c_{ij,r_0} &= K_{ij}(c_{A,r_0})^i (c_{B,r_0})^j \\ c_A^{tot} &= c_{A,r_0} c_A + \sum_{ij} c_{ij,r_0} i c_{ij} \\ c_B^{tot} &= c_{B,r_0} c_B + \sum_{ij} c_{ij,r_0} j c_{ij} \end{aligned} \tag{5}$$

for i,j ≥1; $K_{10} = K_{01} = 1$; with c_{A,r_0} and c_{B,r_0} denoting the concentrations of free A and B, respectively, at the reference radius; K_{ij} denoting the macroscopic equilibrium association constant for formation of a complex with stoichiometry (i,j); and χ_A, χ_B, and χ_{ij} denoting the factor χ evaluated according to Eq. (2) for species of molar mass of A, B, and mixed complexes (i,j) with molar mass (iM_A+jM_B), respectively.

Molar mass distributions

For non-interacting multi-component mixtures with unknown number of species, the SE profile can be related to a molar mass distribution

$$A_r = \int_{M_{\min}}^{M_{\max}} c_{r_0}(M)\varepsilon d \exp\left\{M(1-\bar{v}\rho)\frac{\omega^2}{2RT}\left(r^2 - r_0^2\right)\right\} dM \qquad (6)$$

In this form, the shape of the distribution $c_{r0}(M)$ does not well represent the molar mass distribution of the sample. It is strongly dependent on the choice of the reference radius, and can be significantly skewed either toward the low or the high molar mass range. Using Equation (2), we can arrive at a more rational distribution

$$c(M) = \frac{c_{r_0}(M)}{\chi(M,m,b)} \qquad (7)$$

in which the distribution is scaled in equivalent loading concentrations of species with different molar mass. This distribution is now only dependent on the meniscus and the bottom position of the solution column. Besides the rational concentration scale, it has the advantage that it can be modeled globally to data from different rotor speeds, with the bottom position an adjustable parameter.

A computational method for calculating the molar mass distribution with regularization has been described by Wiff and Gehatia (36). In SEDPHAT, we adopt a similar procedure with Tikhonov-Phillips regularization (37) as described by Provencher (38) which we have applied recently to

sedimentation problems (28, 39), and kinetic analysis of binding to surface-immobilized sites in optical biosensors (40). The regularization procedure addresses the problem that the direct inversion of Equation (6) is highly ill-conditioned, by calculating the simplest or smoothest distribution that is consistent with the experimental data. The numerical procedure is described in detail in (28). In the implementation in SEDPHAT, Equations (6) and (7) can be evaluated repeatedly in a global fit of different data channels from the same or from different cells while iteratively optimizing the meniscus and bottom position of each solution column.

Systematic Noise Decomposition

We have shown previously that in dynamic transport experiments the systematic radial-dependent signal offsets can be distinguished from the macromolecular migration if a mathematical model for the macromolecular redistribution is available. In the case of sedimentation velocity, the Lamm equation is such a model (as are sedimentation coefficient distributions ls-g*(s) (39), c(s) (28, 41), and the back-transform of g(s*) (35)), and the measured signal can be decomposed into that from sedimentation combined with a time-invariant radial-dependent (TI) and radial-invariant time-dependent (RI) signal offset (33). The decomposition is an algebraic step and part of the global model. It has been applied to the modeling of sedimentation velocity data from interference, absorbance and pseudo-absorbance data (33, 34, 41). Although SE is not a transport experiment observing molecular migration as a function of time, it can be understood in an analogous form as molecular redistribution as a function of rotor speed. Therefore, similar algebraic techniques can be used in a global model of experimental data from multiple rotor speeds, if the Lamm equation is replaced by suitable SE models.

In general, the least-squares problem of modeling SE data $a_{r,\omega}$ acquired at radius r and rotor speed ω, which are superimposed by radial-dependent baseline offsets b_r can be written as

$$\underset{\{c,\,K,\,p\},\,b_r}{\mathrm{Min}} \sum_{r,w} \left[a_{r,w} - \left(b_r + A_{r,w}(\{c,\,K,\,p\}) \right) \right]^2 \qquad (8)$$

where $A_{r,\omega}$ denotes the SE model at radius r and rotor speed ω. It depends on the set of parameters {c,K,p} usually representing concentrations, equilibrium constants, or other parameters (e.g., see Eq. (4)). Analogous to the derivation in (41), it can be shown that the best-fit baseline offsets are given by

$$b_r(\{c, K, p\}) = \bar{a}_r - \bar{A}_r(\{c, K, p\})$$
$$\bar{a}_r = (1/N_w)\sum_w a_{r,w}\ , \quad \bar{A}_r(\{c, K, p\}) = (1/N_w)\sum_w A_{r,w}(\{c, K, p\}) \tag{9}$$

with N_ω denoting the total number of rotor speeds available. The least-squares problem for the calculation of the remaining parameters {c,K,p} reduces to

$$\underset{\{c, K, p\}}{\text{Min}} \sum_{r,w} \left[\left(a_{r,w} - \bar{a}_r \right) - \left(A_{r,w}(\{c, K, p\}) - \bar{A}_r(\{c, K, p\}) \right) \right]^2 \tag{10}$$

(41). A similar procedure can be used to determine additional baseline offsets b_ω that do not change with radius, but are different for different rotor speeds (in analogy to the radial-invariant 'RI'-noise of sedimentation velocity) (33). It should be noted that the information content exploited in Eq. (10) is conceptually equivalent to the difference between the signals at the different rotor speeds, but without the noise amplification associated with the differencing of data.

When working at higher rotor speeds, the centrifugal force slightly deforms the rotor and shifts the cell components to higher radii. In particular for the absorbance optical system where the radial-dependent offsets usually originate from small imperfections in the cell assembly, this may cause rotor-speed dependent displacement of the radial-dependent offsets b_r. Since this displacement is precisely predictable and a constant property of the rotor, and since this does not change the property of the b_r being linear parameters, corrections for this effect can be made in the algebraic method for determining b_r, as outlined in the Appendix.

When no mass conservation is used and the concentration parameters in the SE model are simply the concentrations at the reference radius for non-interacting species, such as c_{r0} in Eq. (1), the calculation of the systematic radial-dependent offsets also proceeds slightly different from the case of

sedimentation velocity. The difference stems from the concentration parameter in a sedimentation velocity model being a global multiplicative parameter that remains the same for all scans at each time-point. In contrast, in a global fit of Eq. (1) to SE profiles of macromolecular redistribution at different rotor speeds, the concentration parameter is local to each data set. The algebraic expressions for calculating the concentration parameters, either with a global constant baseline or with radial-dependent baseline offsets, are derived in the Appendix.

As in transport experiments, the estimated baseline offsets b_r and b_ω are dependent on the model $A_{r,\omega}$ (41). In principle, they can correlate in the non-linear regression with the parameters {c,K,p}, which can be minimized by choosing a suitably large range of rotor speeds and by applying MC constraints (see below). The extent of correlation can be assessed through inspection of the covariance matrix.

Global modeling with the software SEDPHAT

SEDPHAT is a general software package for the analysis of hydrodynamic and thermodynamic data from analytical ultracentrifugation and dynamic light scattering. It can be downloaded from www.analyticalultracentrifugation.com, where a help-system and documentation can also be found. Some of the basic principles of data organization and global modeling was described previously in the context of using sedimentation velocity for the characterization of protein self-association (35). The present communication describes some of the capabilities for SE. Data are organized in different data channels (termed 'experiments'), which can be of the type of a single equilibrium profile or of the type 'multi-speed equilibrium'. The latter refers to a set of scans acquired at different rotor speeds from the same cell using the same optical configuration of data acquisition. (A third data type implemented for SE analysis in SEDPHAT is the isotherm of signal-average molar mass as a function of composition, which may be available, for example, from a series of short-column or micro-fractionation SE

experiments (42)(8)(43).) Channels of data with different optical systems may be analyzed simultaneously in a global analysis. For each channel, the experimental parameters such as buffer density, partial-specific volume, centerpiece and rotor type, optical pathlength, meniscus and bottom position, extinction coefficients (or signal increments) of the different components, the expected baseline type (constant throughout, time or rotor-speed dependent, or radial dependent), and the expected instrumental noise are stored. Baselines, meniscus, bottom, extinction coefficients, and/or concentrations can be floated in the data analysis as parameters local to this data channel, and/or they can be shared unknown parameters among different data sets. As algorithm for the non-linear optimization either the Simplex or the Marquardt-Levenberg method (44) can be chosen.

SE models for non-interacting species (discrete or continuous), protein self-association, hetero-association, and mixed associations are available. The models for associating components can incorporate thermodynamic non-ideality and corrections for the compressibility of water (45), and they can be combined with non-interacting species, either in the form of incompetent fractions or of unrelated contaminating species. The choice of the model determines the type and number of global parameters which may include molar mass, partial-specific volume, binding constants, and fractions of incompetent material. The analysis can be switched from concentration units at a reference radius (taken as the bottom of the solution column) to units of 'effective loading concentrations' imposing mass conservation, the latter using Equation (2) to integrate the concentration profiles from meniscus to bottom analytically or numerically, dependent on the centerpiece type. Mass conservation for SE can be 'soft' or 'hard', dependent on if the effective loading concentration and/or the bottom position of the cell are treated as floating parameters to be optimized while modeling the data. When using mass conservation for the analysis of heterogeneous interactions between two components A and B, the effective loading concentration of B can be either specified directly or be specified via the concentration ratio B/A. As the effective loading concentrations, this ratio may also be a parameter shared among different data channels.

Several models are available for globally analyzing mixtures of non-interacting species: a) different mixtures of the same species, where the partial concentration of each species is completely unrelated between data sets; b) mixtures of species that may have different populations or signal contributions in different experiments (different data channels) but with mass balance constraint for the redistribution of species within one cell (i.e., for data channels of the type 'multi-speed equilibrium'); c) a discrete distribution, which may give different total signals in different data channels, but with mass conservation and the same relative populations of the species throughout; d) a continuous distribution, using the same mass conservation assumption as the discrete distribution but without specifying the number of species; e) a hybrid model with discrete species that are common to all experiments combined with multiple continuous distributions that are local to each data set.

The numerical strategy for the mass conservation models is as follows: The effective molar loading concentrations are fitting parameters. For each experiment and each species, the values of χ are calculated by either numerical or analytical integration of the sample volume from meniscus to bottom. For the non-interacting species, Equation (2) will then give directly the concentrations at the reference radius for each experiment. For the associating components, Equation (5) is solved iteratively to calculate the reference concentrations of the free protein, based on the total effective loading concentrations and the binding constants. For self-associating systems, this requires only a one-parametric root finding algorithm (SEDPHAT applies the Van Wijngaareden-Dekker-Brent method (44)). For hetero-associations and mixed associations of two proteins A and B forming reversible complexes at stoichiometries not larger than 2:2, the reference concentration of free B can be calculated directly for any value of free A via a quadratic equation, thus still requiring only a one-dimensional non-linear root-finding procedure for determining the reference concentration of free A that satisfies Equation (5). From these, the reference concentrations of free B and all complexes follow, and Equation (4) can be used to calculate the signal distributions, using the appropriate signal

increments or extinction coefficients of each species for each data channel.

Experimental:

SE experiments were performed with an Optima XL-A/I (Beckman Coulter, Fullerton, CA) with double-sector or six-channel centerpieces and using both the absorbance optical and/or the interference optical data acquisition system. Because of the high sensitivity of the interference optical system against mechanical deformation of the cell components, the cells were mechanically stabilized ('aged') before the SE experiment (Jeff Lary, personal communication): Following the procedure of Ansevin et al (46), after assembling the cell it was first filled with water, inserted in a rotor, spun at a high rotor speed (e.g. 50,000 rpm) for approximately 1 hour, taken out of the rotor and retorqued. This was repeated until no change in the torque of the cell was observed. The cell was then subjected to alternating 1 hour periods at high and low centrifugal fields, taking a set of interference scans each time before changing the speed. This can be programmed to take place automatically overnight with the feature 'equilibrium method' of the instrument software of the centrifuge. After several cycles, the fringe displacement profiles at the high field stabilize, which can be followed by the software MATCH (kindly provided by Jeff Lary from the National Analytical Ultracentrifugation Facility, Storrs, CT). This indicates that the cells are 'aged', i.e., mechanically stabilized to give a reproducible baseline profile. The water blanks were then measured at the different rotor speeds to be used in the SE experiment. Finally, the water was removed and the macromolecular sample and buffer reference were inserted into the cell without disassembly. After the experiment the sample was removed and usually water was reinserted into the cell to verify that the baseline profile remained constant during the experiment.

Results

As a first example for the application of mass conservation constraints and global analysis in SEDPHAT, Figure 1 shows the SE profiles of SPE-C (47), a protein that exhibits no self-association and in the absence of other proteins sediments as a single ideal non-interacting species. The data shown are from the same cell in equilibrium at a range of different rotor speeds. The data sets acquired at 280 nm at all rotor speeds form one 'multi-speed equilibrium' data channel (Figure 1A), and a second data channel is formed by the equilibrium absorbance profiles at 230 nm (Figure 1C). A good fit is achieved with a global model with mass conservation between the different rotor speeds, with the bottom position determined by non-linear regression (but constrained to be the same in both data channels). A single floating 'effective' loading concentration was determined by the global fit, with the extinction coefficient at 280 nm a fixed parameter, and at 230 nm a floating parameter, respectively. The baselines at each wavelength were also unknowns, but constrained to be independent of rotor speed (i.e., for each wavelength, there is one baseline parameter shared for all equilibrium profiles). The global fit resulted in an rms error of 0.006 OD for the 280 nm data channel, and 0.012 OD for the 230 nm data, and in a molar mass estimate of 23.6 kDa, which is nearly identical to the theoretical mass 24.3 kDa. (Also included in the global fit were interference optical data of the same cell (Figure 3), which will be discussed below.) A slightly worse fit was achieved when the geometry of the 6-channel centerpiece was not taken into account.

This example illustrates that a good fit of the experimental data can be found that obeys mass conservation, without using a fixed prior assumption of the bottom position, baseline, or loading concentration. We found this to be true with the majority (but not all) of proteins studied in our laboratory by SE. As will be discussed below, the analysis is not very sensitive to a small loss of material during the experiment. With a sufficient sample volume and range of rotor speeds, the effective loading concentration, meniscus and bottom of the solution column, and the baseline can be well determined by the data.

The data in Figure 1 also illustrates the well-known advantage of using more than one wavelength (or optical system) for data acquisition: While the data at 280 nm produce smaller gradients and permit scanning somewhat closer to the bottom of the cell, the data at 230 nm have significantly higher signal-to-noise ratio in the vicinity of the meniscus. Since the solute concentration is obviously the same for the scans at 280 nm and 230 nm, the extinction coefficient at 230 nm is very well determined by the data and the additional information can be seamlessly integrated in the analysis. It is known that the wavelength reproducibility of the monochromator in the absorbance optical system of the Optima XL-A is limited, and the scans may deviate by 1 - 2 nm from the desired wavelength. This is not significant when scanning at a maximum or minimum of the solute extinction profile, but can be a concern in the steep extinction gradient of proteins at 230 nm. However, we have found deviations by 1 nm not to be problematic in global fits; this could be a result of a sharp emission peak from the lamp in combination with a relatively wider bandwidth of the monochromator.

There are two motivations for imposing mass conservation constraints: first, to reduce the number of concentration parameters which simplifies the error surface in non-linear regression, and second, to improve the error estimates of the derived parameters due to the use of prior knowledge and to make the analysis more stringent. The second aspect can be studied with simulated data of a multi-component system. Figure 2A shows simulated SE of a mixture of species with molar masses of 25 kDa (20% relative loading concentration), 50 kDa (30%) and 100 kDa (50%), at a series of rotor speeds. To test the effect of mass conservation, we attempted the analysis with an impostor model which considers only a 50 kDa and a 100 kDa species, neglecting the 20% population of the smallest species. Without mass conservation constraints, this leads to χ^2_r of the fit of 1.9. If the mass conservation between the different rotor speeds is imposed (with an effective loading concentration, a single baseline, and the bottom position as floating parameters), the χ^2_r of the impostor fit increases to 3.1, with clearly systematic residuals (Figure 2B). This demonstrates that despite the introduction of

the baseline and bottom as additional floating parameters (which converge very close to the correct position), a fit with soft mass conservation is more stringent and has a higher potential to discriminate the correct model.

Similarly, if the correct model is applied, the error estimates of the derived parameters decrease when the MC constraint is applied. This was examined with simulated equilibrium profiles (in the same configuration as in Figure 2) of an interacting system of two proteins (40 kDa and 70 kDa) at equimolar 5 μM loading concentration, which form a reversible complex with K_D = 10 μM. When analyzing the data without MC and using a shared (radially constant) baseline, the covariance matrix gives an error estimate for log10(K_D) of 0.0134. In the otherwise unchanged model but with MC, the error estimate was 0.0057, and in a model with MC and floating bottom position it was 0.0076. With a shared radial-dependent baseline profile (see below) a value of 0.0182 was obtained, with radially constant offsets for each rotor speed it was 0.0090, and with a combination of a shared radial-dependent baseline profile and constant offsets for each rotor speed, the error estimate increased to 0.0190. Although error estimates from the covariance matrix may not reflect true errors of the determined parameters, they describe the local curvature of the error surface in the minimum, and therefore can be used to compare the rigidity of different models. These results illustrate how MC constraints improve the precision of the analysis.

However, it should be noted that this improvement does not overcome the fundamental problem of determining unknown molar mass values in multi-component mixtures (or more generally, to determine the exponents in noisy multi-exponential data). For example, the data in Figure 2A can be modeled with a χ_r^2 close to unity with two species of 34 kDa and 94 kDa, with or without mass conservation constraints (data not shown). Fortunately, in most applications of SE, the molar mass values are sufficiently well known from amino acid composition or mass spectroscopy. One possibility for data analysis of mixtures with unknown molar mass values and an unknown number of

species is a continuous molar mass distribution c(M). As outlined in the theory section, a rational scale of c(M) uses the transformation of reference c_{r0} values to effective loading concentrations (Equation (2)), which implies mass conservation and knowledge of the bottom position of the solution column. The inset in Figure 2A shows the calculated molar mass distributions for the multi-component mixture. The bold solid line is the molar mass distribution calculated from a global fit of the data treating the bottom position as a floating parameter (the best-fit value converged to 7.198 cm, compared to 7.2 cm underlying the simulation) and with an unknown baseline common to all data. For comparison, the inset also shows the c(M) distributions from the separate analyses of the scans at each rotor speed (with a floating baseline, but keeping the bottom position fixed at the value of 7.2 cm). Although the fundamental limitations of analyzing noisy exponentials and of regularization cannot be overcome, the global analysis shows higher resolution than the individual analyses. It should be noted that the c(M) analysis of SE can only be applied to non-interacting species.

One could conceive that by keeping the total effective loading concentration an unknown parameter, the combination with an unknown bottom position and floating baseline would represent a system of parameters flexible enough to compensate for any experimental violations of mass conservation. Since this would still permit eliminating the reference concentrations as fitting parameters and therefore simplify the computation, and since the loading concentration, baseline and bottom position are not the parameters of primary interest, this would not necessarily represent a drawback. However, the concern would be that wrongly imposing mass conservation constraints could potentially bias the analysis and provide incorrect parameter estimates perhaps also for the macromolecular parameters. This possibility was studied using simulated data.

Figure 2C shows simulated SE profiles of a macromolecule of 75 kDa at rotor speeds of 10,000 rpm, 12,000 rpm, 15,000 rpm and 20,000 rpm, assuming decreasing effective loading concentrations of 0.5, 0.45, 0.4, and 0.35, respectively, mimicking the continuous loss of material going from lower to higher rotor speeds (e.g. through the existence of a critical concentration at which the protein

aggregates and enters a surface-film). The best-fit with the impostor assumption of mass conservation and a baseline offset shared among the profiles leads to a χ^2_r of 4.4, with a molar mass estimate 3.9 % above the correct value, a best-fit baseline offset of 0.023 and a bottom position shifted by 0.014 cm. However, the residual are poor and show systematic variation (Figure 2D). Fits with a single, radial-dependent baseline profile (see below) yields a slightly lower but still significantly elevated χ^2_r of 3.1 (with an error in the molar mass of 2%). Table 1 shows the results of similar impostor fits to a series of different simulated systems, all not fulfilling mass conservation. They are qualitatively in agreement with the results shown in Figure 2C/D/E. (In more detail, they suggest that the assumption of rotor-speed dependent baselines decreases the discrimination whether mass balance is observed or not.). This demonstrates that, to some extent, small experimental violation of mass conservation can be compensated by floating the baseline and bottom position, especially with data of low signal-to-noise ratio. However, in general, the relative error in the macromolecular parameters were smaller than the relative loss of material between rotor speeds. Further, the quality of a global fit will suffer significantly from substantial violations of mass conservation, and can therefore be taken as a criterion for the validity of the mass conservation analysis. These results may depend on having long solution columns and sufficiently high rotor speeds (similar to those used in the present communication), such that most of the information of the equilibrium profiles can reside in their curvature.

Since the soft mass conservation introduced in the present paper requires equilibrium profiles at different rotor speeds, it opens the possibility of an algebraic determination of the time-invariant baseline offsets in interference optical data. Obviously, the signal from a single data set does not permit one to distinguish the artificial offsets due to the optical system from the macromolecular signal contributions. However, if the artificial signal offsets do not change with rotor speed, equilibrium data from a sequence of rotor speeds may contain sufficient information to extract the signal from the macromolecular distributions. In principle, this may be possible with or without mass conservation

constraints, but the use of mass conservation will clearly add additional stringency to the model and will permit more reliable identification of the signal offsets. Interestingly, compared to rotor-speed dependent but radially constant offsets, a radial-dependent baseline profile shared by the different scans generally introduces less flexibility into the model with regard to the discrimination if mass conservation is really fulfilled (Table 1). This possibility to calculate baseline profiles is illustrated in practice in Figure 3. Panel A shows the interference optical signals of the same cell as Figure 1, taken in SE at the same rotor speeds and alongside with the absorbance data. The data were included in the global fit of Figure 1 as an additional 'multi-speed equilibrium' data channel, with a radial-dependent baseline (local to this data channel). The bold solid line in Figure 3A is the calculated baseline, and Figure 3B shows the net signal from macromolecular SE alone. Figure 3C shows the residuals with an rms error of 0.0056 fringes, which is within the usual noise of the optical system. (Some very low level of remaining apparently non-random noise is observed, but in our hands this is common for modeling interference SE data; the residual randomness and the rms errors compare well with those observed in other laboratories (48) or (49)). The consistency of the interference data with the absorbance data in Figure 1 supports the validity of the algebraic signal decomposition. The analysis of the IF data alone by a continuous molar mass distribution model reveals a relatively sharp peak at 23 kDa, close to the sequence molar mass of 24.3 kDa.

The preceding analysis demonstrates that SE fringe shift profiles at different rotor speeds can be modeled globally assuming that the well-known systematic baseline offsets of IF data do not change with rotor speed, and therefore can be calculated algebraically. The question if the calculated baseline profiles resemble the profiles experimentally measured with water blanks, and to what extent the latter can be experimentally shown as independent of rotor speed was studied in more detail. Figure 4 shows SE profiles of the octameric rotavirus protein tsE (50) at a wide range of rotor speeds. In the absence of magnesium, this molecule is a very stable octamer of molar mass 302 kDa. Accordingly the equilibrium data can be modeled well as a single species with best-fit molar mass of 304 kDa (the c(M)

analysis exhibits a peak at ~ 302 kDa). Mass conservation is fulfilled, with an rms error of the fit (0.0066 fringes) close to the noise of the data acquisition. That there is no loss of protein is consistent with previous observations in sedimentation velocity experiments, where the protein could be virtually completely resuspended by shaking up the cells after sedimentation for 3 hours at 45,000 rpm (99% recovery of loading signal, unpublished observation).

Figure 4A shows the calculated baseline profile (bold line), which can be compared with the measured profiles from water blanks (Figure 5). It should be noted that the measured blanks are only defined up to a constant radial-independent offset. The gradients described by the calculated baseline profile (solid line in Figure 5A) and the measured water blanks before (+) and after (o) the SE experiment (averaged over all rotor speeds) are very consistent and show similar overall features. It could be argued that the consistency between the calculated and the measured blank is favored because data at very low rotor speeds were included. However, very similar baseline profiles were calculated based on the analysis of the three highest rotor speeds only (squares in Figure 5A, in this case the rms error of the fit improves to 0.0038 fringes). Some differences at the radii closest to the bottom position can be observed between the profiles calculated from the full data set and the highest speeds (solid lines and squares). A similar slight difference is also observed at these radii when comparing the blanks before (+) and after (o) the experiment, indicating a minor instability in the experimental baseline. Despite these small deviations, very satisfactory consistency between the experimental and calculated baseline profiles is found overall, with an rms deviation of only 0.0031 fringes (after correction for the radial-independent offsets). The data of Figure 5 provide another control if mass balance was fulfilled: Simulations mimicking the experimental data but with a 5% loss of material in successively higher rotor speeds (Table 1) would produce a nearly linearly increasing baseline profile estimate with maximum amplitude of ~ 0.018 fringes, or an amplitude of 0.05 fringes for 10% successive loss, respectively (data not shown). From comparison of the calculated and measured baseline profiles, such artificial systematic gradients can be clearly ruled out, supporting the notion that

conservation of mass is fulfilled to a precision better than 5%.

To address the question of how the experimental baselines vary with rotor speed, a superposition of the water blanks acquired before the equilibrium experiment is shown in Figure 5B. The rms deviation between all scans is 0.0041 fringes. Again, although there are clearly some differences in the details, the overall pattern of the signal gradient is very similar. In fact, the detailed changes in the patterns at ~ 7.05 to 7.10 cm appear to be a result of limited reproducibility rather than a systematic change with rotor speed, as the highest and lowest trace in that range are from the 3,000 and 5,000 rpm blanks, respectively. This shows that at the relatively small centrifugal fields used for these SE experiments, the baseline profiles can, in a first approximation, be considered rotor-speed invariant. We have tested this assumption with rotor speeds up to 25,000 rpm (with mechanically stabilized and 'aged' cells) and found it to be satisfactory (usually with rms deviations < 0.01 fringes). At least within that range of rotor speeds our results support the use of a single average baseline profile, and its algebraic calculation as part of the SE analysis with mass conservation constraints.

An interesting question for mass conservation methods is how well the bottom position of the solution column can be determined, either experimentally or computationally. The high quality of the global fit shown in Figures 1 and 3, the correct molar mass value (within the uncertainty of the buoyancy), and the in general relatively high stability of bacterial superantigens support that, in this experiment, the mass of sedimenting protein was indeed conserved when redistributing at the different rotor speeds. However, simulated data mimicking the experimental conditions and global analysis model of the data shown in Figure 1 and 3, but assuming a loss of 5% for successively higher rotor speeds, can still be fit well with a model assuming mass conservation (Table 1). While the error in the molar mass only amounts to 0.4%, the bottom position would migrate from a simulated 7.156 cm to the best-fit position of 7.173 cm (with strongly decreasing quality of fit for non-optimal values). Because of this uncertainty, for the given data the best-fit bottom position does not appear to be a good way to measure the true bottom position of the cell. When compared to the onset of optical artifacts in

the interference optical scans (the regions of IF artifacts at the lowest rotor speed are indicated by the dotted lines in the upper right corner of Figure 3A), the calculated bottom at 7.164 cm appears to be too large (vertical dashed lines in Figure 3A and 3B). However, when the bottom position is constrained to the onset of artifacts in the interference optical system, no acceptable fit fulfilling mass conservation is achieved (15fold increase in χ^2 of the fit). Another example is the data set of the relatively stable protein tsE in Figure 4, which exhibits higher curvature and higher signal-to-noise ratio. Here, the calculated radius appears much better determined by the data, and despite the independent support for mass conservation (via the shape of the baseline profiles, see above), it is quite insensitive to small violations in the conservation of mass (Table 1). Although more reliable, it is still located at a radius higher than the onset of optical artifacts (vertical line in Figure 4B). Therefore, we question if the onset of optical artifacts in the experimental scans represents a good estimate for the true bottom position. Factors we hypothesize may contribute to experimental uncertainty include the finite optical resolution in both the IF and the absorbance system, reflections close to the bottom, the shadow caused by the window gasket or the cell barrel, the planar bottom of the centerpiece, tilting of the axis of rotation, and limitations in the accuracy of the radial calibration. In our experience, although the calculated bottom positions are sometimes consistent with the onset of optical artifacts in the optical traces (for example, the transmitted intensity data), the optical traces do not always guarantee a good estimate for the bottom position. Fortunately, at least when using long solution columns and a range of rotor speeds including meniscus depletion (26), we found the detailed conservation of mass and the detailed bottom position not to be crucial for the derived values of the macromolecular parameters (Table 1).

As described above, it is possible to include radial-dependent baseline into the global SE model if data at several rotor speeds are available. Obviously, this will introduce more flexibility into the model. For example, the impostor model of the simulated data in Figure 2 leads to a χ_r^2 of the fit of 1.6

if a radial-dependent baseline is permitted as opposed to a value of 3.1 with a constant baseline. On the other hand, real experimental data in most cases do exhibit at least some systematic radial-dependent baseline profiles, as is apparent from sedimentation velocity analyses. Their consideration can make the model more realistic and lead to a better fit, which in turn can lead to smaller error estimates.

This aspect is explored in the following example of a protein in a monomer-dimer self-association equilibrium, observed with the absorbance optical system (Yikilmaz *et al*., manuscript in preparation). In this experiment, three different cells were loaded with threefold dilutions of the protein and SE was attained at three different rotor speeds (Figure 6). In order to acquire data over a wide concentration range, absorbance traces at wavelengths of 230 nm, 250 nm, and 280 nm were collected. The relative extinction coefficients were predetermined from a preliminary global analysis of the absorbance traces acquired at an intermediate loading concentration (similar to the example in Figure 1). Alternatively, data from multiple wavelengths of the same cells were included in the global analysis (Figure 6A) including an extinction coefficient as a floating parameter. For the data shown in Figure 6, a reasonable global fit is obtained with soft mass conservations (χ^2_r = 2.98, assuming a random noise of the data acquisition of 0.005 OD at all wavelengths), floating bottom positions and baselines, leading to effective loading concentrations consistent with the loading of the cells. The best-fit dimerization constant is 7.7 [6.9 – 8.9] μM. This shows that for this interacting system, the soft mass conservation constraints lead to well-determined binding constants.

A close inspection of fit in Figure 6 shows that the residuals appear slightly systematic in several details, such as spikes in the meniscus region in Panel A. This suggests that systematic radial-dependent imperfections of the optical system are partially responsible for the deviations from the theoretical best-fit distributions. Similar observations are frequently made in sedimentation velocity experiments, where a large number of absorbance traces are collected and small systematic, time-

invariant radial-dependent absorbance contributions are routinely observed (e.g., (33, 51)). In sedimentation velocity, they are routinely discriminated from the signals of macromolecular migration by algebraic noise decomposition. As shown for the IF optical system above, SE profiles acquired at several rotor speeds permit similar computational techniques for their identification, in particular if the equilibrium analysis is stabilized by mass conservation constraints. When this modeling approach is applied to the absorbance data from Figure 6, i.e. including a radial-dependent baseline with an otherwise identical model, a significantly better fit is achieved. As shown in Figure 7, the calculated radial-dependent baselines follow the systematic patterns indicated above, and the remaining residuals appear more randomly distributed. This is very clear in the low absorbance data, for example, in comparison of Figure 7C/D with Figure 6 C/D. Overall, this leads to a highly significant 2.3fold reduction of the χ_r^2 of the fit (χ_r^2 = 1.30). Interestingly, despite the increased flexibility of the fit introduced by the radial-dependent baseline, the error interval for the binding constant is 7.1 – 9. 3 μM (best-fit estimate 8.1 μM) which is similar to the error interval obtained with the model in the absence of radial-dependent baseline.

This example demonstrates that accounting for systematic signal offsets caused by optical imperfections makes the model statistically more rigid, which, in this case, largely compensates for introducing the additional parameters. Similar results were obtained when an unknown fraction of irreversible dimer was included in the fit (data not shown). However, we did not get satisfactory results with the calculated radial baseline profiles in experiments that did not include the shallow gradient established at the lowest rotor speed of 8,000 rpm. Analysis of SE data in a different experiment using a set of rotor speeds of 12,000 rpm, 15,000 rpm, and 20,000 rpm resulted in very large systematic slopes in the calculated baseline profiles (data not shown). This was accompanied by a 10fold increase in the error estimate (calculated by covariance matrix) of the association constant when radial-dependent baseline was permitted. We conclude, therefore, that the algebraic

determination of radial-dependent baseline profiles for absorbance data cannot be applied uncritically without examining the shape of the baseline profile and the flexibility introduced into the model (see discussion).

At higher rotor speeds, the mechanical forces on the rotor cause a deformation that displaces the centrifugal cell assembly to slightly higher radii (rotor stretching). For the data shown in Figure 1 which cover rotor speeds from 15,000 rpm to 30,000 rpm, the movement of the cell is 0.0065 cm. Figure 8 illustrates how this effect can be accounted for in the systematic signal offsets. Since it is based on a pre-determined elasticity module for the rotor (we experimentally determined values of 8.6×10^{-12} cm/rpm^2 for the 4-hole rotor An-60 Ti and 9.6×10^{-12} cm/rpm^2 for the 8-hole rotor An-50 Ti, respectively), this does not introduce any additional unknowns. As shown in Figure 8, the small displacements can be accurately modeled. While the introduction of radial-dependent offsets reduces the rms error by 20% (as compared to the analysis of Figure 1) consideration of the rotor stretching leads to a further reduction by 21%, and leads to a more detailed profile of the offsets for shorter spatial scales. It should be noted that significantly larger effects can be expected at higher rotor speeds (on the other hand, virtually no difference was found for fits with or without consideration of rotor stretching for the experiments at the lower rotor speeds of Figure 7).

Interestingly, we have not observed similar improvements in the fit for the interference optical data. Most likely this is due to the fact that the entire optical configuration (in particular the mirrors) contributes to the systematic radial-dependent offsets, not only the cell components in the rotor. It is unclear to what relative extent the different sources contribute to the baseline profile, and it may be dependent on the particular state of the instrument and cell assembly used. This may limit the precision of the estimated radial-dependent offset on shorter spatial scales, but would not affect the calculated radial-dependence on longer spatial scales, which can be more dominating in interference optical data. (However, limiting factor for the algebraic modeling of the offsets at high rotor speeds

may be the mechanical stability of the cell assemblies.) Also, we have not observed significant improvements in the quality of the mass conservation models when the rotor-speed dependent small displacement of the meniscus and bottom of the cells were taken into consideration.

Heterogeneous interactions provide the opportunity to extend the strategy of soft mass balance constraints. Obviously, constraints can be implemented separately for each component requiring the conservation of mass during the redistribution at different rotor speeds. As suggested by Philo (17), additional constraints can be imposed for the relationship of the effective loading concentrations between the components and/or between different cells. The precise effective loading concentrations at equilibrium may not be known for any component, but in many situations it may be a conservative assumption that any fractional loss of material or concentration error in either component would be the same in different experiments (i.e., different cells). For example, if the cells were loaded by holding constant the concentration of one component while varying the second component, the effective concentration of the first component may be constrained to be the same in all cells. Alternatively, if both components are varied in different cells but held at a constant molar ratio (for example, by loading a dilution series of the stock of a mixture), a soft constraint can be implemented that the molar ratio of the effective total loading concentrations be the same in each cell. In the notorious case of heterogeneous interactions with proteins of similar size, frequently only a lower limit for the binding constants can be determined if the concentrations of both component are allowed to float independently in the fit. In this case, if the molar ratio of loading concentrations of the two components can be fixed and treated as prior knowledge, the dependency between the binding constant and the remaining concentration parameters is removed, permitting the determination of the binding constant.

Figure 9 shows an example for a heterogeneous interaction of a NK receptor dimer (30 kDa) binding two MHC molecules (44 kDa) at symmetric sites (52). The extinction coefficient and buoyant molar mass for each protein were determined in separate experiments by using SEDPHAT global

analysis of data from the same cell acquired at different wavelengths, floating molar mass and extinction coefficient. The sedimentation profiles of each protein could be modeled very well as a single species obeying mass conservation among the data at the different rotor speeds when the bottom position was treated as an unknown parameter and determined by nonlinear regression (with the procedure described above). For the analysis of the interaction, three cells were loaded in a dilution series, producing SE gradients covering a large concentration range. Each cell was scanned at least at two wavelengths (230 nm, 250 nm, and/or 280 nm). For the mixtures, mass conservation of each cell among the different rotor speeds was imposed (with unknown bottom position), as well as a constant, but unknown molar ratio of the two components among the different cells. A radial-dependent baseline was permitted in order to account for the imperfections of the data acquisition system. As illustrated by the residuals in Figure 9, an excellent fit was achieved with nearly random residuals, and a best-fit molar ratio of ~ 1.7. Because of the similar molar mass of the receptor and its ligand and the weak binding (estimated 80 [40 – 200] μM), the molar ratio was not very precisely determined by the data. When the molar ratio of the effective loading concentrations of the two proteins was constrained to the value determined prior to centrifugation (loading was aimed at equimolar concentrations), a slight increase in the rms deviation of the fit occurred (F = 1.099). Overall, the data show a weak (but significant) interaction between the proteins with largely similar binding constants independent on the details of the treatment of baseline parameters and molar ratio. However, in the present context, the data demonstrate that soft mass conservation constraints may be imposed in a flexible way on the analysis of heterogeneous protein interactions, if the loading concentration and bottom position of the cells are not treated as fix prior knowledge. Without any mass balance constraint, the analysis contains 18 parameters for reference concentrations, which are correlated and are difficult to optimize in the non-linear regression. Imposing mass conservation between the different rotor speeds of each cell reduces the number of total concentration parameters to 6, and the span of the estimated error interval

for K_A decreases by at least a factor of 3.2 (assuming the absence of cooperativity between the two sites.) Linking the molar ratio of the two protein components in the different cells to a constant, but unknown value, reduces the concentration parameters to 4 and leads to a further decrease of the error interval by a factor of 1.05. Finally, treating the molar ratio as constant prior knowledge leaves 3 concentration parameters only, and produces a smaller error interval by another factor of 1.7.

Discussion

In the present paper we have suggested an extension of the original approach by Roark (13) to introduce implicit mass conservation (MC) constraints in the global modeling of SE data. It requires that SE be established sequentially with the same sample at multiple rotor speeds. From global modeling of these profiles information can be extracted on the total amount of soluble macromolecules (i.e., the effective loading concentration), the position of the bottom of the solution column, and the parameters of the baseline signal. This avoids the potential introduction of errors from imprecise knowledge of the loading concentration, initial adsorption to components of the ultracentrifugal cell assembly, and bias from the graphical estimation of the bottom position which in our hands make the strict implementation of MC frequently problematic (see below). Introduction of implicit MC can offer several advantages in the analysis of SE.

Most importantly, it can substantially reduce the number of free parameters in the global model. This is true, in particular, for global analyses of heterogeneous associations, which in the absence of MC can represent a notoriously difficult modeling problem, exhibiting a very complicated error surface with many local minima at unrealistic parameter combinations (15, 17). MC considerably simplifies the global error surface and enables rapid convergence. This property can be utilized even if the amount of soluble material is not completely conserved throughout the experiment: As pointed out by Philo, it can be used in a first step to guide the local concentration parameters to excellent starting

values for an unconstrained fit (otherwise a formidable problem). In contrast to Philo's elegant soft MC implementation (17), which can be regarded as a regularization term for an otherwise possibly ill-conditioned problem, in the present method the stringency of the MC constraint cannot be continuously adjusted. However, it is interesting to note in this context the flexibility introduced by treating the bottom of the solution column as an unknown. The present approach does tolerate some minor violations of MC without significant amplification of error (see Table 1), which should make it useful to initialize a second-stage unconstrained fit even if MC is not strictly fulfilled.

MC can improve the error estimates of the derived parameters and can help with rejecting impostor models. The extent of how much the precision of the analysis can be improved depends significantly on the type and number of baseline parameters that need to be introduced. (Usually, for absorbance optical data, only a single baseline should be required for each set of rotor speeds.) Since the method of global implicit MC presented here introduces the effective loading concentration and the bottom of the solution column as additional unknown parameters, the benefit in the error estimates is slightly lower than those of the original strict MC as implemented, for example, by Pörschke (21) or Lewis (23). It will also be slightly less than in the original multi-speed method by Roark (13). However, this treatment can be more conservative, considering the uncertainty which the prior estimates of these parameters may exhibit themselves. Nevertheless, SEDPHAT can be easily used to analyze data with 'hard' MC constraints by not floating the effective loading concentration and by fixing the bottom position in the analysis.

For heterogeneous interactions where one species does not contribute to the observed signal, MC constraints make possible the analysis of binding constants (24), where otherwise only the stoichiometry of the complex could be determined. A sedimentation analysis of such interactions is also possible in the MC implementation in SEDPHAT, of course requiring the effective loading concentration of the 'invisible' component not to be floated. (As SEDPHAT calculates SE and mass action law in molar units, treating each component's extinction coefficients as local parameters [which

may be unknowns, possibly shared among data subsets], it is suitable for general multi-wavelength and multi-signal analyses.) Another situation where a conventional SE analysis by multi-exponential decomposition is impossible is the heterogeneous interactions of two species with similar molar mass. In such a case, because the local concentrations of both free species are indistinguishable, the binding constant cannot be assigned. This can also be overcome by imposing MC constraints on the model. Most useful in this regard should be the possibility to constrain molar ratios of loading concentrations (17, 23). This is implemented in SEDPHAT, in a variation that still permits to treat the effective loading concentrations as unknowns. Similarly, the effective loading concentration of one component may be a parameter (possibly unknown) shared by several data sets. Obviously, which of these constraints is applicable will depend on the design of the experiment (e.g., if different experiments are dilutions of the same premixed stock solution, or if one component is loaded at constant concentration in a titration series with the second component). In the example of Figure 9, the global model of data from different cells could be constrained to have the same molar ratio of loading concentrations of both components. In this case, treating the molar ratio as unknown, which may frequently be more conservative unless the true concentrations of both proteins are accurately known, mainly reduced the number of unknown parameters. Fixing the molar ratio additionally reduced the error estimate by a factor of 1.7, and this treatment would also be generally required for a well-conditioned analysis if two proteins have the same molar mass.

Another benefit from MC constraints for heterogeneous interactions is that it provides a simple framework to consider incompetent subpopulations of loading material in a global model. Finally, the implicit MC can be useful to calculate radial-dependent baseline profiles in a well-conditioned global analysis, as discussed in more detail below. Although non-ideality was not considered in the present paper, it could possibly be introduced by appropriate modification of Eqs. 1 and 2.

For non-interacting mixtures of macromolecules, the global implicit MC provides an unbiased scale for continuous molar mass distributions from SE. The use of MC as a concentration scale in

continuous model-free molar mass distributions was used before by Wiff and Gehatia (36), and the analysis of equilibrium profiles from different rotor speeds was introduced by Scholte (for discrete distributions and without regularization) (53, 54). However, the problem of determining the correct bottom position, which has considerable influence on the calculated molar mass distribution, was not addressed. The extension in SEDPHAT to include the bottom as a fitting parameter allows an unbiased estimate from the data, scaling the continuous molar mass distribution calculated with Tikhonov-Phillips regularization approach (37).

The approach described here requires SE in relatively long solution columns (we have used 4 – 5 mm) to be measured at least at two but preferably at three rotor speeds. It works best if the rotor speeds span a range including shallow and steepregions, thus combining information from the traditional low-speed (25) and high-speed equilibrium methods (26). For good experimental baseline information, in our experience it is advantageous if the highest rotor speed generates (at least close to) meniscus depletion (26). These observations are in line with the discussion of experimental conditions by Roark, who characterized 6 mm columns as optimal (13). He has also pointed out the advantage of combining, in a global fit, the multi-speed analyses of solutions with different initial loading composition (as shown also in the examples of the present paper) and recommended to start with the lowest rotor speed in order to avoid gel-formation at the bottom of the solution column (13). In our experience, with configurations as shown in the examples, it usually takes less time for equilibrium to be attained at the second and following rotor speeds than for the first (not considering initial overspeeding (25)).

Obviously, not all protein samples will be stable enough during the extended time of the multi-speed experiment. Slow degradation of the material may occur and cause violations of mass conservation. (It should be noted that SEDPHAT has the capability to do flexible conventional global SE analyses without MC, in the form of local or global analysis of single or multiple data sets at different concentrations but only one rotor speed; such a conventional global analysis can still benefit

from a preliminary MC analysis to initialize the local reference concentrations, as pointed out by Philo (17)). However, in our hands MC could be applied in most protein systems recently studied in our laboratory. A significant advantage of using implicit MC in a global multi-speed analysis is that it does not compare the measured total mass of soluble protein with the 'known' concentrations before the start of the experiment. This excludes errors in estimating protein concentrations by different methods, initial adsorption to the window and centerpieces, and large rapidly sedimenting aggregates (26) from MC consideration. With regard to the stringency of MC, Roark has estimated that a loss of material of 2% may be tolerated (13). However, since we have introduced the bottom position as a free parameter, slightly higher loss of material could occur before generating large and systematic residuals or suspicious baseline profiles in the global model. Importantly, the macromolecular parameters of interest were found not to be very sensitive to small violations in MC. In that regard, the enhanced flexibility of the model could be an advantage.

An important question when considering mass balance is the precise location of the bottom of the solution column. Experimentally, it is far from trivial to locate, partly due to the finite resolution of the optical system and optical artifacts close to the interface. The same is true for the meniscus, but because in SE the proteins deplete the meniscus but accumulate at the bottom, the latter is much more critical. This issue has been discussed in detail by Schachman and colleagues (25). Because of the experimental difficulties, we have treated the detailed bottom position as a floating parameter to be determined in the data analysis using MC constraints. This is, perhaps, the most significant extension of the original method by Roark (13). The free parameter can be constrained to a radial range that is evident from the optical traces and from upper bounds on the geometry of the centerpieces and rotor, within the experimental uncertainty. In some experimental configurations, such as in Figure 1/3, this parameter may not be very well determined by the data. However, this is not a drawback of the method, since it is not a parameter of primary interest, and the most significant information on the macromolecular properties resides in the curvature of the equilibrium gradients, not their projected end

point or their integral. On the other hand, in the example of Figure 4, the calculated bottom position was very well determined and relatively insensitive to hypothetical MC violations (which several indicators suggested to be absent). In both situations, the calculated bottom position was not consistent with the onset of optical artifacts in the equilibrium scans, while fixing the bottom to the latter experimental estimates did not lead to a satisfactory fit. Nevertheless, we have observed cases where the onset of artifacts was more consistent with the calculated bottom position (data not shown). These results show the difficulty in the experimental determination of the precise bottom position and suggest that circumventing it by global modeling of multi-speed data is a reasonable approach.

The precise bottom location (as well as the question to what degree MC is violated) appears less critical in Philo's method of soft MC constraints, where MC plays a role mainly as regularization to reshape the error surface of the global model and can be adjusted not to bias the final results (17). Another method that has been discussed in this context, which is completely independent of the question of MC violations or of the bottom position is Minton's method of 'conservation of signal' (55). It also acts to simplify the computational problem of fitting the data, but arrives there without any mass balance consideration of loaded or suspended material. The term 'conservation of signal' refers to the scaling of the partial signals from each species to match the total measured signal for each profile, calculated by integration over the range of modeled data. Such a 'conservation' is ordinarily understood as part of a least-squares fit, i.e., simply for the predicted profiles to closely match the data points. While the method can simplify a computational step in modeling the data, it has the drawback of not using the Euclidian norm for eliminating the reference concentration, which makes the method inconsistent with least-squares methods (56). 'Conservation' in this context does not stand for a quantity to be compared between different scans or with prior knowledge (55). However, if the 'conservation-of-signal' method would be extended and applied in a sense of conservation of the mass visible in a given radial window at different rotor speeds, then the neglect of material in the segments of the solution column not directly observed would create significant problems (17).

Using the conservation of mass as criterion in the analysis also raises the issue of the baseline absorbance. While there is a school of thought of categorically not treating the baseline as an unknown, we believe that an optimal analysis has to take experimental imperfections into account, which can include small differences in the light transmitted through the reference and the sample sector in the absorbance optics. While that may not be a good practice and lead to well-known parameter correlation with short solution columns and shallow gradients, this is obviously not problematic in meniscus depletion or close to meniscus depletion equilibrium gradients. It is crucial, therefore, to specify the experimental configuration and the fitting model in the discussion of how to treat baseline offsets. With long solution columns, conditions close to meniscus depletion, and the absorbance optical system, good results can generally be achieved treating the baseline as an unknown parameter. In the present case, the use of several rotor speeds permits to constrain the baseline to be the same for these profiles. The combination with MC adds further stringency in its unambiguous determination. Unfortunately, in some cases there is clear evidence of shifting baselines between the rotor speeds (Figure 9C), which can be caused by instable buffer components. According to our results (Table 1), including individually floating baselines for each profile can degrade the potential to discriminate among different models and should be avoided. This is consistent with the recommendations of Philo on this subject (17).

With Rayleigh interference optical data, the situation is much different, since baseline offsets are radial-dependent and considerable effort has to be spent on the experimental determination and elimination (46, 57). Also, due to the nature of fringe displacement data, no baseline can be assigned *a priori* and arbitrary radial-independent offsets have to be considered for each scan. Fortunately, the higher resolution and linearity of this optical system permit higher gradients, which can compensate for the lack of an absolute baseline (e.g. in meniscus depletion). Because it generally allows one to scan closer to the bottom of the solution column, mass balance can be calculated more precisely and the MC constraints appear stronger. In the present paper, we have proposed to exploit the small rotor-

speed dependence of the radial-dependent baseline profiles (at the usual rotor speeds of SE for large macromolecules, < 30,000 rpm), to make the assumption that in a first approximation the profiles are rotor-speed invariant, and to use algebraic techniques to computationally extract them from the global analysis of profiles at multiple rotor speeds. ~~This implies that, again in a first order approximation, beyond the mechanical stability of the cell, rotor stretching will also not be considered.~~ In the communication of Ansevin et al. on improved ultracentrifuge cells, the authors have reported significant changes of the baseline profile with rotor speed (46). However, the latter study was concerned with rotor speeds in excess of 40,000 rpm, and in earlier studies at lower rotor speeds, rotor speed independent blank profiles were reported (26). That the baseline profiles are much more robust at lower rotor speeds is consistent with our findings, as illustrated by the data shown in Figure 5B. Corrections for small radial displacement of the radial-dependent baseline profiles from rotor stretching can be made, although it is in general unclear to what extent the interference data are affected by this (i.e., how much of the profile of the baseline signal is caused by imperfections of the cell components in the rotor *versus* the remaining static elements of the optics, such as mirrors and lenses).

The approach of calculating the baseline profile from global analysis of multiple-speed SE is experimentally and computationally simpler than that developed previously of using an over-determined Lamm equation model to fit the approach to equilibrium (57). Clearly, the previous method would be the method of choice for equilibrium data from a single rotor speed, where the algebraic decomposition of baseline and macromolecular signals cannot be applied. For data from multiple rotor speeds, however, the new method provides some advantages: While the approach-to-equilibrium method works best at higher transport rates, i.e. with steeper gradients, it is more advantageous for the multi-speed MC method presented here to start with more shallow gradients, for which the approach-to-equilibrium method is less precise. Further, the approach-to-equilibrium method excludes the use of initial overspeeding, which can be advantageous to shorten the time

required to attain the first equilibrium. As shown in Figure 5, we found the algebraically calculated baseline profiles to be consistent with the measured water blanks. The algebraic calculation of the radial-dependent baseline is similar to that used routinely for sedimentation velocity (33), and it is similarly dependent on the accuracy of the model for the macromolecular redistribution. The information content exploited for the macromolecules is conceptually equivalent to that of an analysis of $\Delta c/\Delta\omega$, a situation analogous to time-difference data in sedimentation velocity (41).

The same technique may also be applied to remove imperfections in the baseline profile of absorbance optical data, which in many cases can improve the quality of fit significantly. This would suggest that to a significant extent, the noise in the absorbance data are not statistical, but partially systematic. This idea would be supported by the fact that sedimentation velocity data with the absorbance optics clearly show low levels of systematic radial-dependent baseline patterns, possibly originating from imperfections in the windows. As indicated by the data in table 1, introducing radial-dependent baseline profiles does not necessarily introduce a very large flexibility in the model. For the configurations of the experimental and simulated data shown here, floating constant baseline offsets independently for each rotor speed generally introduced larger uncertainties as compared to a single radial-dependent baseline profile shared by all scans. For interference optical data, unfortunately both the local offsets and the radial-dependent profiles have to be considered. But in conjunction with mass conservation, with a large radial range of the equilibrium profiles, and with a large range of rotor speeds, we have had generally good results with this method. In this context, it is of interest to note that MC considerations have been traditionally used in order to calculate fringe offsets (58).

However, it is very important to note that radial-dependent baselines cannot be applied uncritically, in particular to absorbance optical data with lower signal-to-noise ratio: for example, when using a wrong model, if MC is violated and with an insufficient radial range, artifactual baseline profiles may result, which can be identified by a systematic slope across the cell. Such a systematic

sloping baseline should normally not occur in absorbance optics. Besides this criterion, when using absorbance data, the fits with and without considering baseline profiles may differ slightly in quality and error estimates, but they should not give substantially different results. Further, when the error estimates of the determined macromolecular parameters significantly increase, the consideration of radial-dependent baseline profiles is probably not indicated. In our hands, for a reliable calculation of the radial baseline profile it seemed to be important to include a large radial range and rotor speeds generating shallow gradients. The importance of low-speed equilibrium profiles for MC considerations was pointed out earlier by Schachman et al. (25). Where it is possible to collect both absorbance and interference optical data of the same cell, a global analysis (for example, using one extinction coefficient or signal increment as a floating variable to scale the signals) can be advantageous, because it would decrease the overall flexibility of the model that might be introduced by the requirement for calculating radial-dependent baseline offsets in the interference data channel. Nevertheless, such combination can take advantage of the larger radial range and higher linearity of the interference optical system in SE (Figure 1/3). When using sector-shaped cells, it may also be advantageous to perform a sedimentation velocity experiment after remixing the solution by carefully shaking the rotor with cell assemblies. This can be analyzed, for example, with the c(s) method (28), which is suitable for such short column lengths. Beyond independent experimental information about the radial baseline profile (which, except for rotor stretching, should be consistent to the one calculated from multi-speed SE), the peak structure of the c(s) may indicate if degradation has occurred.

The methods outlined above have been implemented in the software SEDPHAT, which is designed as a general global analysis tool for sedimentation velocity (35), sedimentation equilibrium, and dynamic light scattering. The sedimentation equilibrium models currently include non-interacting species as discrete and continuous distributions, protein self-association, heterogeneous and mixed associations of two protein components. It is suitable for flexible global analysis of data acquired at different wavelengths and different optical systems, and it can treat the signal increments, as well as

density increments (59, 60) as unknowns to be determined from the data. All models can be executed with or without mass conservation, and with or without global, local, or radial-dependent baseline offsets. The present study has focused mainly on two aspects when globally analyzing long-column SE data from multiple rotor speeds: we have proposed that in this case radial-dependent baseline profiles can be calculated, and we have examined the use of global implicit MC constraints in conjunction with unbiased computation of the bottom of the solution column. In comparison with Philo's method of soft MC, the presented approach is clearly less general and not as useful if MC is far from being fulfilled. However, we believe it should provide additional utility if MC is fulfilled or almost fulfilled.

Appendix

In the following are outlined the algebraic determination of the radial-dependent offsets with corrections for rotor stretching, and an approach how linear concentration parameters for non-interacting species without mass conservation constraints can be calculated.

In the simplest form, the least-squares problem of modeling SE data $a_{r,\omega}$ acquired at radius r and rotor speed ω which are superimposed by radial-dependent baseline offsets b_r can be written as

$$\underset{\{c,K,p\},b_r}{\text{Min}} \sum_{r,w} \left[a_{r,w} - \left(b_r + A_{r,w} \right) \right]^2 \tag{11}$$

where $A_{r,\omega}$ denotes the SE model at radius r and rotor speed ω (Eq. (8)). For clarity of presentation, the dependence of $A_{r,\omega}$ on a set of additional parameters {c,K,p} is not explicitly stated but implied; these can be optimized in a non-linear regression, where the algebraic elimination of the linear parameters is part of each evaluation of a particular set {c',K',p'}. We want to consider the fact that the radial-dependent offsets b_r experience a rotor-speed dependent shift $\delta_r(\omega)$, which is proportional to the square of the rotor speed $d_r(w) = e(w^2 - w_1^2)$ (we can arbitrarily take the radii at the lowest rotor speed ω_1 as our frame of reference). Because we have data points only at discrete radial intervals Δ_0 (all scans are interpolated to the same equidistant grid of radii), it is useful to calculate the shift in units of Δ_0

$$\Delta_w = e\left(w^2 - w_1^2\right) \Big/ \Delta_0 \tag{12}$$

rounded to the nearest integral number. In this way, rotor stretching appears as a discrete shift in the index of the radius values, and we are not concerned with shifts smaller than the radial precision Δ_0 of the data points. Using the radial-dependent baseline $b_{\rho,1}$ at the lowest rotor speed ω_1 as a reference, this allows us to express the radial-dependent baseline at the higher rotor speeds as

$$b_{r,w} = b_{r-\Delta_w,1} \quad , or \quad b_{r,1} = b_{r+\Delta_w,w} \tag{13}$$

Eq. (11) now becomes

$$\underset{b_{r,1}}{Min}\sum_{w}\sum_{r=\Delta_w}^{N_r+\Delta_w}\left(a_{r,w}-A_{r,w}-b_{r,w}\right)^2 \tag{14}$$

where we have also shifted the analysis range 1…N_r of the radial points to follow the displacement of the rotor. As they are linear parameters, we can calculate the best-fit values of $b_{\rho,1}$ by differentiating (14) with regard to $b_{\rho',1}$. Using the relationship

$$\partial b_{r,w}/\partial b_{r',1}=d_{r,\,r'+\Delta_w} \tag{15}$$

with the Kronecker symbol $\delta_{x,y}$ ($\delta_{x,y}$ = 1 for x =y, 0 else) leads to

$$0=-2\sum_{w}\sum_{r=\Delta_w}^{N_r+\Delta_w}\left(a_{r,w}-A_{r,w}-b_{r,w}\right)d_{r,\,r'+\Delta_w} \tag{16}$$

. It follows

$$\begin{aligned}\sum_{w}\sum_{r=\Delta_w}^{N_r+\Delta_w} b_{r,w}d_{r,\,r'+\Delta_w} &= \sum_{w}\sum_{r=\Delta_w}^{N_r+\Delta_w} b_{r-\Delta_w,1}d_{r,\,r'+\Delta_w}\\ &= \sum_{w}\sum_{r=\Delta_w}^{N_r+\Delta_w}\left(a_{r,w}-A_{r,w}\right)d_{r,\,r'+\Delta_w}\end{aligned} \tag{17}$$

and, executing the summation over the radial points,

$$\begin{aligned}b_{r',1} &= \frac{1}{N_w}\sum_{w}a_{r'+\Delta_w,w}-\frac{1}{N_w}\sum_{w}A_{r'+\Delta_w,w}\\ &=: \bar{a}_r^*-\bar{A}_r^*\end{aligned} \tag{18}$$

with the averages $\bar{a}_r^*$ and $\bar{A}_r^*$ replacing the averages of the data and model over each all scans at each radius in Eq. 2 in (33).

The determination of additional rotor-speed dependent baselines β_ω in the problem

$$\underset{b_{r,1},b_w}{Min}\sum_{w}\sum_{r=\Delta_w}^{N_r+\Delta_w}\left(a_{r,w}-A_{r,w}-b_{r,w}-b_w\right)^2 \tag{19}$$

can proceed completely analogously to the strategy used for sedimentation velocity analysis (33). By normalizing

$$\frac{1}{N_r}\sum_{r=1}^{N_r} b_{r,1} = 0 \tag{20}$$

(compare Eq. 7 in (33)), we arrive at

$$\begin{aligned} b_s &= \frac{1}{N_r}\sum_{r=\Delta_s}^{N_r+\Delta_s} a_{r,s} - \frac{1}{N_r}\sum_{r=\Delta_s}^{N_r+\Delta_s} A_{r,s} \\ &=: \bar{a}_s^* - \bar{A}_s^* \end{aligned} \tag{21}$$

and insertion of this result in Eq. (19) gives a problem of the same structure as Eq. (14), with the analogous solution for the radial-dependent offset

$$b_{r',1} = \frac{1}{N_w}\sum_w (a_{r'+\Delta_w,w} - \bar{a}_w^*) - \frac{1}{N_w}\sum_w (A_{r'+\Delta_w,w} - \bar{A}_w^*) \tag{22}$$

Finally, we consider the case where we want to determine radial-dependent offsets and the concentrations $c_{k,\omega}$ of different independently sedimenting species distributing as $A^{(k)}_{r,\omega}$

$$\underset{b_{r,1},c_{k,w}}{Min} \sum_w \sum_{r=\Delta_w}^{N_r+\Delta_w} \left(a_{r,w} - b_{r,w} - \sum_k c_{k,w} A_{r,w}^{(k)} \right)^2 \tag{23}$$

In the absence of mass conservation constraints, there are no relationships between the concentrations of each species at the different rotor speeds, which makes the $c_{k,\omega}$ linear parameters that can be directly calculated algebraically without non-linear regression. The same strategy as used in Eq. (16) for eliminating $b_{\rho,1}$ leads here to

$$b_{r,1} = \bar{a}_r^* - \sum_k \frac{1}{N_w}\sum_w c_{k,w} A^{(k)}_{r+\Delta_w,w} \tag{24}$$

(for clarity in the following, without changing any results the index ρ' was renamed to ρ and ω to *w*). Inserting this into (23) and using the transformation of radial scale (13) leads to the remaining problem

$$\underset{c_{k,w}}{Min}\sum_{w}\sum_{r=\Delta_w}^{N_r+\Delta_w}\left[\left(a_{r,w}-\bar{a}^{*}_{r-\Delta_w}\right)-\left(\sum_{k}c_{k,w}A^{(k)}_{r,w}-\sum_{k}\frac{1}{N_w}\sum_{w}c_{k,w}A^{(k)}_{r-\Delta_w+\Delta_w,w}\right)\right]^2 \tag{25}$$

This can be rewritten as

$$\underset{c_{k,w}}{Min}\sum_{w}\sum_{r=\Delta_w}^{N_r+\Delta_w}\left[\left(a_{r,w}-\bar{a}^{*}_{r-\Delta_w}\right)-\sum_{w}\sum_{k}c_{k,w}\left(A^{(k)}_{r,w}d_{w,w}-\frac{1}{N_w}A^{(k)}_{r-\Delta_w+\Delta_w,w}\right)\right]^2 \tag{26}$$

and transformed into a $N_r \times N_\omega$ linear equation system for the $c_{k,w}$ by differentiation with regard to the concentrations $c_{k',w'}$

$$\begin{aligned}\sum_{w}\sum_{r=\Delta_w}^{N_r+\Delta_w}&\left[\left(a_{r,w}-\bar{a}^{*}_{r-\Delta_w}\right)\times\left(A^{(k')}_{r,w'}d_{w,w'}-\frac{1}{N_w}A^{(k')}_{r-\Delta_w+\Delta_{w'},w'}\right)\right]\\&=\sum_{w}\sum_{k}c_{k,w}\sum_{w}\sum_{r=\Delta_w}^{N_r+\Delta_w}\left[\left(A^{(k)}_{r,w}d_{w,w}-\frac{1}{N_w}A^{(k)}_{r-\Delta_w+\Delta_w,w}\right)\times\left(A^{(k')}_{r,w'}d_{w,w'}-\frac{1}{N_w}A^{(k')}_{r-\Delta_w+\Delta_{w'},w'}\right)\right]\end{aligned} \tag{27}$$

which can be solved algebraically constraining the concentrations to positive values using the NNLS algorithm by Lawson & Hansen (61) adapted to normal equations. The additional incorporation of rotor-speed dependent baselines β_ω is straightforward. Eq. (27) can be slightly simplified if the radial-dependent baselines are not affected by rotor stretching, as can be the case for interference optical data.

Table 1

simulated data mimicking		% mass loss per rotor speed	baseline offsets	χ^2_r	Δr_b (cm)	simulated macromolecular parameters and error introduced in impostor analysis		
Figure 2C [1]	simulated	~ 10%	constant 0	1		species 25 kDa c_{load} = 0.1	species 50 kDa c_{load} = 0.15	species 100 kDa c_{load} = 0.25
	global fit	0	single const.	3.45	0.027	– 3% [2]	– 1% [2]	+ 4% [2]
	global fit	0	radial dep.	2.71	0.036	– 6% [2]	+ 5% [2]	– 1% [2]
	global fit	0	multiple const.	1.74	0.021	– 3% [2]	– 3% [2]	+ 6% [2]
	global fit	0	radial dep. + multiple const.	1.49	0.028	0	– 1% [2]	+ 1% [2]
Figure 2C [3]	simulated	~10%	constant 0	1		single species Mw 75 kDa		
	global fit	0	single const.	5.17	0.017	fix		
	global fit	0	radial dep.	3.17	0.023	fix		
	global fit	0	multiple const.	2.96	0.016	fix		
	global fit	0	radial dep. + multiple const.	2.36	0.021	fix		
Figure 2C [4]	simulated	~10%	constant 0	1		single species Mw 75 kDa		
	global fit	0	single const.	4.4	0.014	+3.9 %		
	global fit	0	radial dep.	3.12	0.016	+2 %		
	global fit	0	multiple const.	2.93	0.018	+1.1 %		
	global fit	0	radial dep. + multiple const.	2.31	0.016	+4 %		
Figure 2C [5,6]	simulated	5%	constant 0	1		(heterogeneous association of 40 kDa and 70 kDa species) K_D 10 μM (fix molar masses)		
	global fit	0	single const.	1.92 (2.93)	0.009 (0.008)	7.6 μM (7.9 μM) [6]		
	global fit	0	radial dep.	1.54 (2.42)	0.024 (0.004)	31 μM (3.2 μM) [6]		
	global fit	0	multiple const.	1.35 (2.05)	0.008 (0.007)	7.8 μM (8.0 μM) [6]		
	global fit	0	radial dep. + multiple const.	1.00 (1.76)	0.008 (0.006)	4.2 μM (7.7 μM) [6]		
Figure 1A/C + 3B [7]	simulated	5%	constant 0	1		single species Mw 23.3 kDa		
	global fit	0	as in exp. data	1.11	0.017	– 0.4 %		
Figure 4 [8]	simulated	5%	constant 0	1		single species Mw 300 kDa		
	global fit	0	radial dep. + multiple const., as in exp. data	1.11	0.0019	+ 1.2 %		
Figure 4 [8]	simulated	10%	constant 0	1		single species Mw 300 kDa		
	global fit	0	radial dep. + multiple const., as in exp. data	2.28	0.0038	+ 3.2 %		

Legend for Table 1: Effect of the violation of mass conservation on the data analysis. Sets of SE profiles were simulated mimicking configurations of the experiments shown in Figures 2C, 1AC/3B, and 4, but including a hypothesized fractional loss of mass during redistribution at each rotor speed (going from lowest to highest centrifugal field). The resulting sets of 'multi-speed equilibrium' profiles were analyzed using the incorrect assumption of mass conservation. For each experimental configuration and for different rates of loss of material, the table lists decrease in the quality of fit due to violation of mass conservation (increase in χ^2_r), the difference between simulated and best-fit bottom position (Δr_b), and the relative errors introduced to the best-fit estimates of the macromolecular parameters. The errors resulting from violated mass conservation are shown for analyses with different baseline assumptions: a single baseline shared by the profiles at all rotor speeds (single const.), a radially dependent baseline profile common to all rotor speeds (radial dep.), multiple baselines constant in radius but different at different rotor speeds (multiple const.), or a shared radially dependent baseline profile combined with rotor speed dependent constant offsets (single radial dep. + multiple const.). Details of the simulation: (1) At rotor speed of 10,000 rpm: total simulated loading concentration 0.5 (with 20% of the 25kDa, 30% of the 50 kDa, and (30%) and 50% of the 100 kDa species) with a baseline offset of 0; other rotor speeds with same relative ratio of species, but at a total loading concentration of 0.45 and a baseline at 0.01 (12,000 rpm), 0.40 and 0.02 (15,000 rpm), 0.35 and 0.025 (20,000 rpm), respectively. (2) Values are errors estimated for the relative population (in percent of the total) of the different species. (3) Simulated for identical conditions as in previous row, but assuming only a single macromolecular species of known molar mass 75 kDa. (4) Same simulation as above, but treating the molar mass as an unknown in the analysis. (5) Simulated heterogeneous association in the same geometry and rotor speeds as Figure 2C, with a 70 kDa species and a 40 kDa species, both equimolar at 5 μM, reversibly associating to form a 110 kDa species with a K_D of 10 μM. (6) Values in parentheses correspond to the same simulation with fitting range slightly extended to

larger maximal radii from 7.121 cm to 7.140 cm. (7) Simulated conditions resemble the experiments in Figures 1A/C and 3B with regard to rotor speeds, geometry of the solution column, loading concentrations, extinction coefficients, and number and type of optical traces, and required baselines. Gaussian noise of 0.005 OD or fringe displacement, respectively, was added to the theoretical data. (8) Conditions as in the experiment shown in Figure 4, simulated with both 5% or 10% loss of material between equilibria, respectively.

References

1. T. Svedberg, K. O. Pedersen, Die Ultrazentrifuge, Theodor Steinkopff, Dresden,1940.
2. H. K. Schachman, Ultracentrifugation in Biochemistry, Academic Press, New York,1959.
3. A. P. Minton, Quantitative Characterization of reversible molecular associations via analytical ultracentrifugation, Anal. Biochem. 190 (1990) 1-6.
4. T. M. Schuster, J. M. Toedt, New revolutions in the evolution of analytical ultracentrifugation, Curr. Opin. Structural Biology 6 (1996) 650-658.
5. P. Hensley, Defining the structure and stability of macromolecular assemblies in solution: the re-emergence of analytical ultracentrifugation as a practical tool, Structure 4 (1996) 367-373.
6. T. M. Laue, Analytical centrifugation: equilibrium approach., Current Protocols in Protein Science (1999) 20.3.1-20.3.13.
7. T. M. Laue, W. F. I. Stafford, Modern applications of analytical ultracentrifugation., Annu. Rev. Biophys. Biomol. Struct. 28 (1999) 75-100.
8. G. Rivas, W. Stafford, A. P. Minton, Characterization of heterologous protein-protein interactions via analytical ultracentrifugation, Methods: A Companion to Methods in Enzymology 19 (1999) 194-212.
9. J. Liu, S. J. Shire, Analytical ultracentrifugation in the pharmaceutical industry, J Pharm Sci 88 (1999) 1237-41.
10. F. Arisaka, Applications and future perspectives of analytical ultracentrifugation, Tanpakushitsu Kakusan Koso 44 (1999) 82-91.
11. P. Schuck, E. H. Braswell, in: J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober (Ed.)^(Eds.), Current Protocols in Immunology, John Wiley & Sons, New York, 2000, pp. 18.8.1-18.8.22.
12. J. Lebowitz, M. S. Lewis, P. Schuck, Modern analytical ultracentrifugation in protein science: a tutorial review, Protein Sci 11 (2002) 2067-79.

13. D. E. Roark, Sedimentation equilibrium techniques: multiple speed analyses and an overspeed procedure, Biophys Chem 5 (1976) 185-196.
14. J. M. Varah, On fitting exponentials by nonlinear least squares, SIAM J. Sci. Stat. Comput. 6 (1985) 30-44.
15. S. P. Becerra, A. Kumar, M. S. Lewis, S. G. Widen, J. Abbotts, E. M. Karawya, S. H. Hughes, J. Shiloach, S. H. Wilson, Protein-protein interactions of HIV-1 reverse transcriptase: implication of central and C-terminal regions in subunit binding, Biochemistry 30 (1991) 11707-19.
16. J. S. Philo (1997) *in* Proceedings of Ultrasensitive Biochemical Diagnostics II, Vol. 2985, pp. 170-177, SPIE, San Jose, CA.
17. J. S. Philo, Sedimentation equilibrium analysis of mixed associations using numerical constraints to impose mass or signal conservation, Methods in Enzymology 321 (2000) 100-120.
18. A. P. Minton, Alternative strategies for the characterization of associations in multicomponent solutions via measurement of sedimentation equilibrium., Progr. Colloid Polym. Sci. 107 (1997) 11-19.
19. H. Fujita, Foundations of ultracentrifugal analysis, John Wiley & Sons, New York,1975.
20. L. W. Nichol, P. D. Jeffrey, B. K. Milthorpe, The sedimentation equilibrium of heterogeneously associating systems and mixtures of non-interacting solutes: analysis without determination of molecular weight averages, Biophys. Chem. 4 (1976) 259-267.
21. D. Pörschke, D. Labuda, Codon-induced transfer ribonucleic acid association: Quantitative analysis by sedimentation equilibrium, Biochemistry 21 (1982) 53-56.
22. W. Bujalowski, M. Jung, L. W. McLaughlin, D. Pörschke, Codon-induced association of the isolated anticodon loop of tRNA, Biochemistry 25 (1986) 6372-6378.
23. M. S. Lewis, R. J. Youle, Ricin subunit association, J Biol Chem 261 (1986) 11572-11577.
24. M. R. Bubb, M. S. Lewis, E. D. Korn, The interaction of monomeric actin with two binding sites on acanthamoeba actobindin, J Biol Chem 266 (1991) 3820-3826.

25. E. G. Richards, D. C. Teller, H. K. Schachman, Ultracentrifuge studies with Rayleigh interference optics. II. Low-speed sedimentation equilibrium of homogeneous systems., Biochemistry 7 (1968) 1054-1076.

26. D. A. Yphantis, Equilibrium ultracentrifugation of dilute solutions, Biochemistry 3 (1964) 297-317.

27. P. C. Hansen, Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion, SIAM, Philadelphia,1998.

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

29. T. Arakawa, M. Haniu, L. O. Narhi, J. A. Miller, J. Talvenheimo, J. S. Philo, H. T. Chute, C. Matheson, J. Carnahan, J. C. Louis, et al., Formation of heterodimers from three neurotrophins, nerve growth factor, neurotrophin-3, and brain-derived neurotrophic factor, J Biol Chem 269 (1994) 27833-9.

30. J. S. Philo, K. H. Aoki, T. Arakawa, L. O. Narhi, J. Wen, Dimerization of the extracellular domain of the erythropoietin (EPO) receptor by EPO: One high-affinity and one low-affinity interaction, Biochemistry 35 (1996) 1681-1691.

31. T. P. Horan, F. Martin, L. Simonet, T. Arakawa, J. S. Philo, Dimerization of granulocyte-colony stimulating factor receptor: The Ig plus CRH construct of granulocyte-colony stimulating factor receptor forms a 2:2 complex with a ligand, J. Biochem. 121 (1997) 370-375.

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

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

34. S. R. Kar, J. S. Kinsbury, M. S. Lewis, T. M. Laue, P. Schuck, Analysis of transport experiment using pseudo-absorbance data, Anal. Biochem. 285 (2000) 135-142.

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

36. D. R. Wiff, M. T. Gehatia, Inferring a molecular weight distribution, an ill-posed problem; and establishing the molecular weight scale using magnetic float techniques., Biophys. Chem. 5 (1976) 199-206.

37. D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, Assoc. Comput. Mach. 9 (1962) 84-97.

38. S. W. Provencher, A constrained regularization method for inverting data represented by linear algebraic or integral equations, Comp. Phys. Comm. 27 (1982) 213-227.

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

40. J. Svitel, A. Balbo, R. A. Mariuzza, N. R. Gonzales, P. Schuck, Combined affinity and rate constant distributions of ligand populations from experimental surface-binding kinetics and equilibria., Biophys. J. 84 (2003) 4062-4077.

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

42. J. J. Correia, D. A. Yphantis, in: S. E. Harding, A. J. Rowe, J. C. Horton (Ed.)^(Eds.), Analytical ultracentrifugation in biochemistry and polymer science, The Royal Society of Chemistry, Cambridge, U.K., 1992, pp. 231-252.

43. S. Darawshe, G. Rivas, A. P. Minton, Rapid and accurate microfractionation of the contents of small centrifuge tubes: application in the measurement of molecular weights of proteins via sedimentation equilibrium., Anal. Biochem. 209 (1993) 130-135.

44. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical Recipes in C, University Press, Cambridge,1992.

45. P. Schuck, A model for sedimentation in inhomogeneous media. II. Compressibility of aqueous and organic solvens, Biophys. Chem. in press (2003).

46. A. T. Ansevin, D. E. Roark, D. A. Yphantis, Improved ultracentrifuge cells for high-speed sedimentation equilibrium studies with interference optics., Anal. Biochem. 34 (1970) 237-261.

47. Y. Li, H. Li, N. Dimasi, J. K. McCormick, R. Martin, P. Schuck, P. M. Schlievert, R. A. Mariuzza, Crystal structure of a superantigen bound to the high-affinity, zinc-dependent site on MHC class II, Immunity 14 (2001) 93-104.

48. D. A. Yphantis, T. Arakawa, Sedimentation equilibrium measurements of recombinant DNA derived human interferon gamma, Biochemistry 26 (1987) 5422-7.

49. S. J. Advant, E. H. Braswell, C. V. Kumar, D. S. Kalonia, The effect of pH adn temperature on the self-association of recombinant human interleukin-2 as studied by equilibrium sedimentation, Pharm. Res. 12 (1995) 637-641.

50. Z. F. Taraporewala, P. Schuck, R. F. Ramig, L. Silvestri, J. T. Patton, Analysis of a temperature-sensitive mutant rotavirus indicates that NSP2 octamers are the functional form of the protein, J Virol 76 (2002) 7082-93.

51. S. A. Ali, N. Iwabuchi, T. Matsui, K. Hirota, S. Kidokoro, M. Arai, K. Kuwajima, P. Schuck, F. Arisaka, Rapid and dynamic association equilibrium of a molecular chaperone, gp57A, of bacteriophage T4, Biophys J (in press).

52. J. Dam, R. Guan, K. Natarajan, N. Dimasi, L. K. Chlewicki, D. M. Kranz, P. Schuck, D. H. Margulies, R. A. Mariuzza, Variable MHC class I engagement by Ly49 NK cell receptors revealed by the crystal structure of Ly49C bound to H-2Kb, Nature Immunology Published online: 2 November 2003, doi:10.1038/ni1006 (2003).

53. T. G. Scholte, Molecular weights and molecular weight distribution of polymers by equilibrium ultracentrifugation. Part II. Molecular weight distribution., J. Polym. Sci. 6 (1968) 111-127.
54. P. J. Wan, E. T. Adams, Molecular weights adn molecular-weight distributions from ultracentrifugation of nonideal solutions, Biophys Chem 5 (1976) 207-241.
55. A. P. Minton, in: T. M. Schuster, T. M. Laue (Ed.)^(Eds.), Modern Analytical Ultracentrifugation, Birkhäuser, Boston, 1994, pp. 81-93.
56. P. R. Bevington, D. K. Robinson, Data Reduction and Error Analysis for the Physical Sciences, Mc-Graw-Hill, New York,1992.
57. P. Schuck, Sedimentation equilibrium analysis of interference optical data by systematic noise decomposition, Anal. Biochem. 272 (1999) 199-208.
58. E. G. Richards, H. K. Schachman, Ultracentrifuge studies with Rayleigh interference optics. I. General applications, J. Phys. Chem. 63 (1959) 1578-1591.
59. J. Edelstein, H. Schachman, Simultaneous determination of partial specific volumes and molecular weights wtih microgram quantities, J. Biol. Chem. 242 (1967) 306-311.
60. D. Noy, J. R. Calhoun, J. D. Lear, Direct analysis of protein sedimentation equilibrium in detergent solutions without density matching, Anal. Biochem. in press (2003).
61. C. L. Lawson, R. J. Hanson, Solving least squares problems, Prentice-Hall, Englewood Cliffs, New Jersey,1974.

Legends for the Figures

Figure 1: Experimental SE data and modeling with mass conservation. Shown are absorbance data of the protein SPE-C (47) at rotor speeds of 15,000 rpm (crosses), 20,000 rpm (circles), 25,000 rpm (triangles), and 30,000 rpm (squares) (for clarity, only every second data point is shown). For experimental details, see (47). The wavelength of the data acquisition was 280 nm (A,B) and 231 nm (C,D), respectively, and the data were loaded in SEDPHAT as two separate multi-speed SE channels. (Also included in the global fit were the data acquired with the interference optical system shown in Figure 3.) The solid lines are the best-fit models for a single species with molar mass of 23.6 kDa, at loading absorbencies of 0.11 OD_{280} (A,B) and 0.80 OD_{231} (C,D), respectively, redistributing at the different rotor speeds with mass conservation in the outer channel of a six-channel centerpiece. Panels B and D are the residuals of the fit, which have a local rms error for each channel of 0.0060 OD_{280} (B) and 0.012 OD_{231} (D). The meniscus and bottom position of the solution column were optimized in a global fit, but constrained to be the same in both data channels. At each wavelength (i.e. each channel), a single baseline parameter was included as a floating parameter common to all rotor speeds (best-fit estimates were < 0.001 OD).

Figure 2: Analysis of simulated sedimentation data examining the discrimination of impostor models. Panel A: Simulated SE data of a non-interacting mixture of species with molar masses of 25 kDa (20%), 50 kDa (30%) and 100 kDa (50%). The total loading signal was 0.5 for each data set, and the sedimentation profiles were calculated at rotor speeds of 10,000 rpm (squares), 12,000 rpm (circles), 15,000 rpm (triangles), and 20,000 rpm (crosses) and a temperature of 8 °C. Only every second data point is shown. Random noise of magnitude 0.005 was added. The solid lines show the best-fit mass conservation model assuming the presence of the 50 kDa and 100 kDa species only. χ^2_r of the impostor fit was 3.1, as compared with the χ^2_r of 1.9 obtained without mass conservation. In the global

fit, a single baseline for all rotor speeds was assumed, and the bottom position, as well as the loading concentration, were treated as floating parameters. The best-fit loading signal was 0.468, the baseline 0.024, and the rms error of the fit was 0.0087. The bottom converged to close to the correct value of 7.2 cm. Panel B shows the residuals of the fit. The inset shows the best-fit molar mass distribution c(M) from a global analysis with floating bottom position (bold solid line), and from the individual analysis (with fixed bottom position) of the data at 10,000 rpm (dashed line), 12,000 rpm (dash-dotted line), 15,000 rpm (dash-dot-dotted line), and 20,000 rpm (thin solid line), respectively. Panel C: Simulated SE profiles of a 75 kDa species not obeying mass conservation. Loading concentrations were 0.5 at 10,000 rpm (squares), 0.45 at 12,000 rpm (circles), 0.4 at 15,000 rpm (triangles), and 0.35 at 20,000 rpm (crosses), respectively. Other conditions were identical to the data in A. Solid lines are the best-fit distributions under the incorrect assumption of mass conservation, using the molar mass as an unknown parameter (converging to 77.9 kDa), a floating bottom position (best-fit is 7.214 cm), and a single unknown baseline shared between the scans (0.023). Residuals of this fit are shown in D, with a χ^2_r of 4.4. If a single radial-dependent baseline is permitted, the fit improves to χ^2_r of 3.1 and residuals as shown in E.

Figure 3: Interference optical data of the same experiment as shown in Figure 1, in SE at 15,000 rpm (dash-dot-dotted line, trace of optical artifacts in the bottom region continued as dotted line), 20,000 rpm (dash-dotted line), 25,000 rpm (dashed line), and 30,000 rpm (solid line). The data were included in the global fit of Figure 1, as a third data channel with multi-speed SE. The model of this channel included radial-invariant offsets for each rotor speed, and a rotor-speed invariant but radial-dependent baseline profile. Panel A shows the raw data, with the best-fit radial-invariant baselines at each rotor speed subtracted to clarify the presentation. The bold solid line is the best-fit rotor-speed invariant baseline profile. Panel B shows the calculated net signal from macromolecular redistribution at the

different rotor speeds. Panel C are the residuals of the global fit, which have a local rms error of 0.0056 fringes. The meniscus and bottom position for the IF data channel were also constrained to be the same as in Figure 1. The bottom position is indicated in Panel B as dashed vertical line at 7.164 cm. The best-fit molar mass distribution c(M) (based on the IF data alone, with floating bottom position) is shown in the inset of Panel B. Best-fit rms error of the c(M) model is 0.0053 fringes, with a best-fit bottom position of 7.162 cm.

Figure 4: Interference optical data of the octameric rotavirus protein tsE in SE at rotor speeds of 3,000 rpm (dash-dot-dotted line), 5,000 rpm (dash-dotted line), 8,000 rpm (dashed line), 12,000 rpm (dotted line), and 16,000 rpm (solid line). For experimental details, see (50). The data were modeled in SEDPHAT as a single species with mass conservation governing the redistribution at different rotor speeds. The bottom position of the solution column, the loading concentration, the radial-invariant offsets at each rotor speed and a rotor-speed invariant baseline profile common to each scan were treated as a fitting parameter. Shown are the raw interference data (A, solid lines) together with the best-fit rotor-speed invariant baseline profile (A, bold solid line), and the calculated net signals from macromolecular redistribution (B). The best-fit position of the bottom is indicated as the dashed line at 7.195 cm, for comparison the raw data in the bottom region for 3,000 rpm and 5,000 rpm scans are shown as dotted line in A. The residuals of the fit are shown in C for the rotor speeds of 3,000 rpm (triangles), 5,000 rpm (plus), 8,000 rpm (squares), 12,000 rpm (crosses), and 16,000 rpm (circles) (every 2nd data point shown). Rms error of the fit is 0.0066 fringes.

Figure 5: Panel A: A comparison between the calculated rotor-speed invariant noise (solid line) from the analysis of Figure 4 and the measured baseline profiles from water blanks before (+) and after (o) the SE experiment (only every 3rd data point shown). The rms error between all data shown after correction for the radial-invariant offset is 0.0031 fringes. For comparison, the rotor-speed invariant

noise is also shown (squares, every 5th data point) when the analysis was based only on the equilibrium data at the rotor speeds of 8,000 rpm, 12,000 rpm, and 16,000 rpm. Panel B: Experimental baseline profiles from water blanks measured before the SE experiment at rotor speeds of 3,000 rpm (solid line), 5,000 rpm (dotted line), 8,000 rpm (bold solid line), 12,000 rpm (dash-dotted line), and 16,000 rpm (dashed line). The rms deviation among these profiles (except for a constant signal offset) is 0.0041 fringes.

Figure 6: Example for the analysis of a protein in a monomer-dimer self-association equilibrium, measured by absorbance optics. Monomer molar mass is 100 kDa, and SE scans at rotor speeds of 8,000 rpm (circles), 12,000 rpm (crosses), and 15,000 rpm (triangles) were obtained at three different loading concentrations (A, C, and E, respectively), and absorbance scans were acquired at wavelengths of 280 nm (A, C), 250 nm (A, scaled by factor 2 and offset by 0.3 vertically, 0.05 horizontally), and/or 230 nm (E). For clarity, only every second data point is shown. For details see (Yikilmaz et al, manuscript in preparation). Solid lines are the best-fit distributions from global analysis including soft mass balance constraints using loading concentrations, bottom positions, and a single baseline as parameters local to each cell, but common to all absorbance traces for each cell (except for 230 nm data E, which required different baseline offsets for each trace). Rms errors of the fit are 0.0119 OD for the 280 nm data in A (residuals in B, symbols), 0.0074 OD for the 250 nm data in Panel A (residuals in B, lines), 0.0054 OD for the data in C (residuals in D), and 0.0081 for the data in E (residuals in F). Best-fit effective loading concentrations are 3.71 μM (A), 1.23 μM (C), and 0.46 μM (E), and best-fit dimerization constant is 7.7 μM (with an error interval from 6.9 – 8.9 μM, calculated by projections of the error surface and F-statistics).

Figure 7: Analysis of the same data as shown in Figure 6 including a radial-dependent baseline for

each cell. The calculated best-fit distributions of protein absorbance are indicated as solid lines, the estimated radial-dependent baseline as bold solid lines. Rms errors of the fit are 0.0077 OD for the 280 nm data in A (residuals in B, symbols), 0.0047 OD for the 250 nm data in A (residuals in B, lines), 0.0023 OD for the data in C (residuals in D), and 0.0063 for the data in E (residuals in F). Best-fit effective loading concentrations are 3.60 μM (A), 1.16 μM (C), and 0.45 μM (E), and best-fit dimerization constant is 8.1 μM (with an error interval from 7.1 – 9.3 μM, calculated by projections of the error surface and F-statistics).

Figure 8: Modeling the data shown in Figure 1A with systematic radial-dependent offsets incorporating corrections for rotor stretching at higher speeds. Panel A: Absorbance optical data at rotor speeds of 15,000 rpm (crosses), 20,000 rpm (circles), 25,000 rpm (triangles), and 30,000 rpm (squares) and best-fit models (solid lines). The displacement of the rotor at 30,000 rpm relative to 15,000 rpm is 0.0065 cm and can be observed, for example, in the features of the fit at ~ 6.88 cm. The calculated radial-dependent offset profile is shown on top (solid line, in the radial frame of reference at 15,000 rpm), and in comparison with the offset profile calculated without corrections for rotor stretching (dotted line). Residuals of the fit are shown in Panel B.

Figure 9: Heterogeneous interaction between an NK receptor dimer (30 kDa) binding up to two MHC molecules (44 kDa) at symmetric sites (52). Shown are three cells loaded in a dilution series of a mixture of receptor and MHC, sedimented to attain equilibrium at rotor speeds of 15,000 rpm (crosses), 20,000 rpm (triangles), and 25,000 rpm (circles), each scanned at wavelengths of 280 nm, 250 nm, and 230 nm. Experimental details are described in (52). Shown are the absorbance data of the cell with ~ 2 μM of each protein at 280 nm (A/B), ~ 7 μM at 250 nm (C/D), and 0.5 μM at 230 nm (E/F). For clarity, only every 2nd data point is shown. Solid lines show the best-fit distributions with

soft mass conservation, a common molar ratio between the effective loading concentration of receptor and MHC, assuming no cooperativity between the sites, using predetermined buoyant molar masses and extinction coefficients for each protein, with optimized bottom position of the cell, and radial-dependent baseline. Because the absorbance traces at 20,000 rpm and 25,000 rpm in C intersect, demonstrating a baseline shift in this set, a rotor speed-dependent but radially constant offset was included as parameters local to this set. The best-fit binding constant was K_D ~ 80 [40 – 200] μM, with a best-fit molar ratio of effective loading concentrations of ~ 1.7.

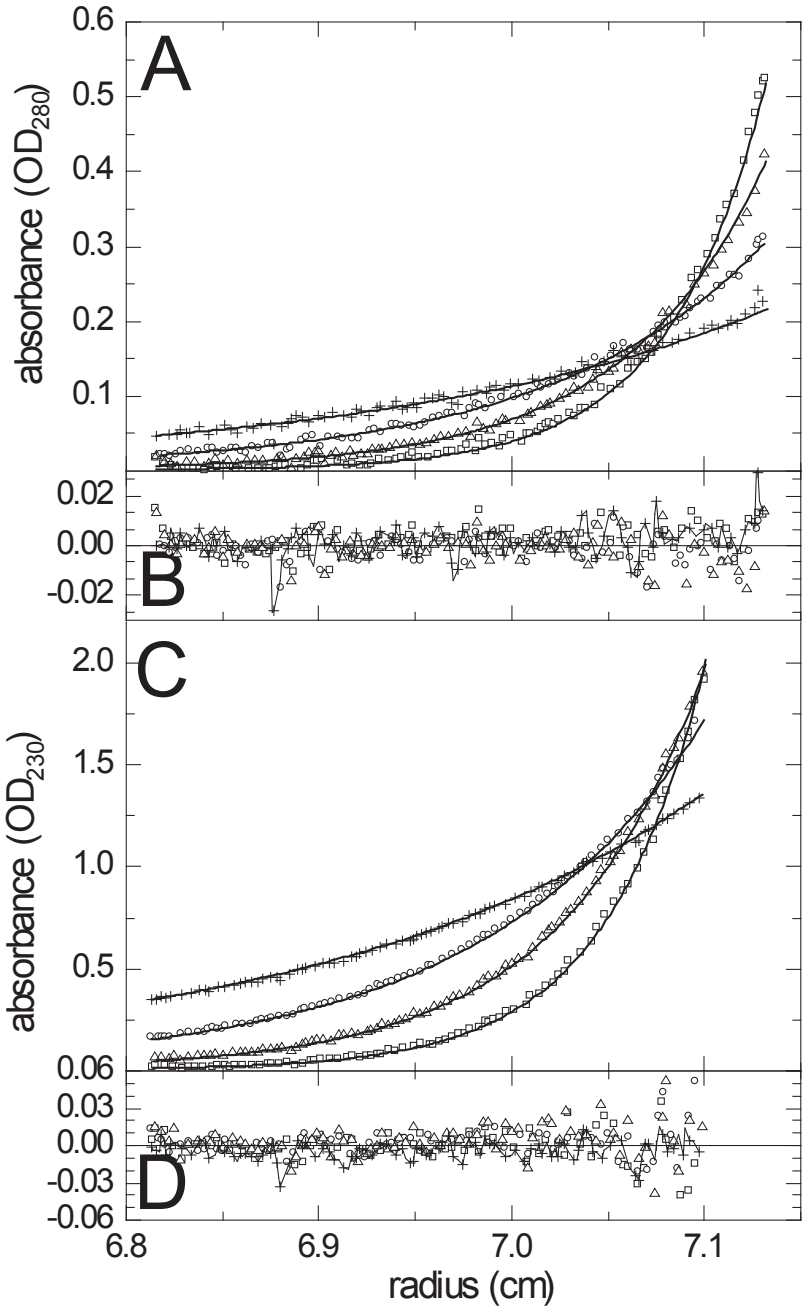
A
B
C
D
absorbance (OD280)
absorbance (OD230)
radius (cm)
0.6
0.5
0.4
0.3
0.2
0.1
0.02
0.00
-0.02
2.0
1.5
1.0
0.5
0.06
0.03
0.00
-0.03
-0.06
6.8
6.9
7.0
7.1

Figure 1

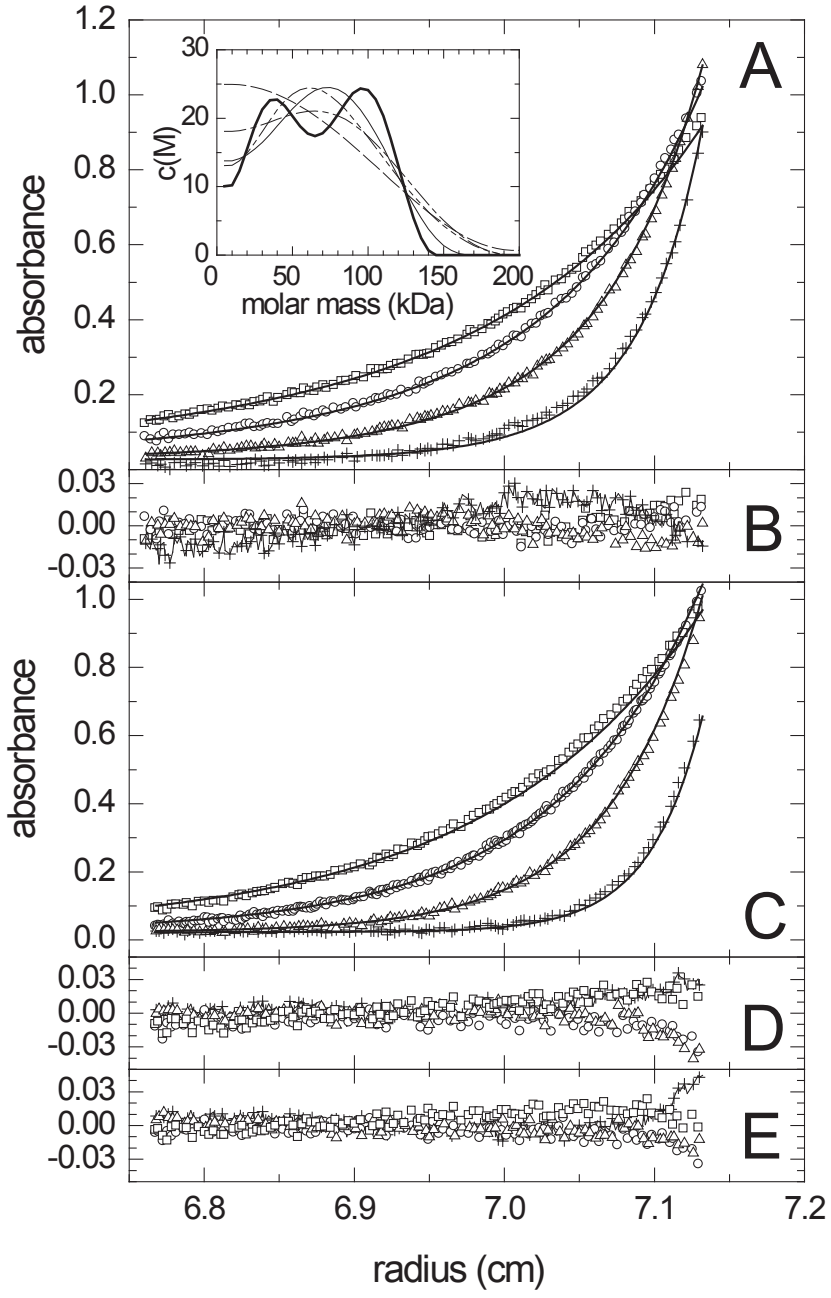
A
B
C
D
E
absorbance
absorbance
c(M)
molar mass (kDa)
radius (cm)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.03
0.00
-0.03
30
20
10
0
0
50
100
150
200
6.8
6.9
7.0
7.1
7.2

Figure 2

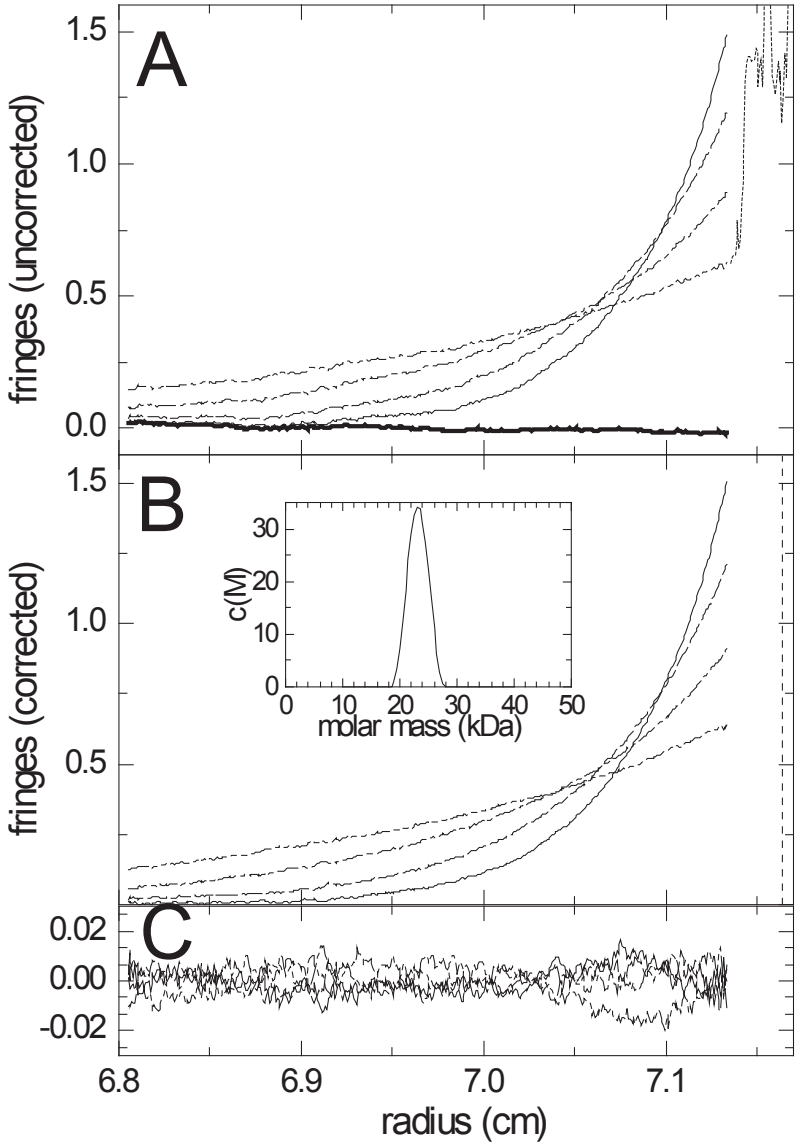
A
B
C
fringes (uncorrected)
fringes (corrected)
1.5
1.0
0.5
0.0
0.02
0.00
-0.02
c(M)
30
20
10
0
0
10
20
30
40
50
molar mass (kDa)
6.8
6.9
7.0
7.1
radius (cm)

Figure 3

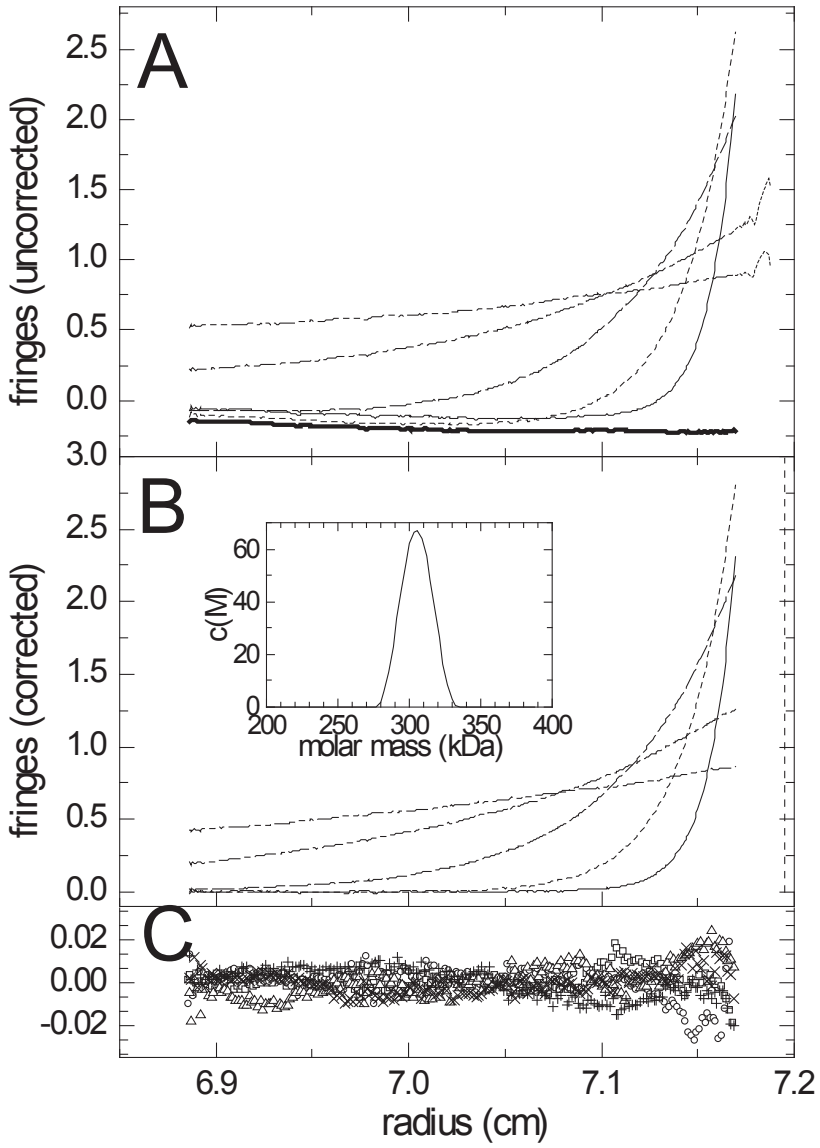
A
B
C
fringes (uncorrected)
fringes (corrected)
c(M)
molar mass (kDa)
radius (cm)
2.5
2.0
1.5
1.0
0.5
0.0
3.0
60
40
20
0
200
250
300
350
400
0.02
0.00
-0.02
6.9
7.0
7.1
7.2

Figure 4

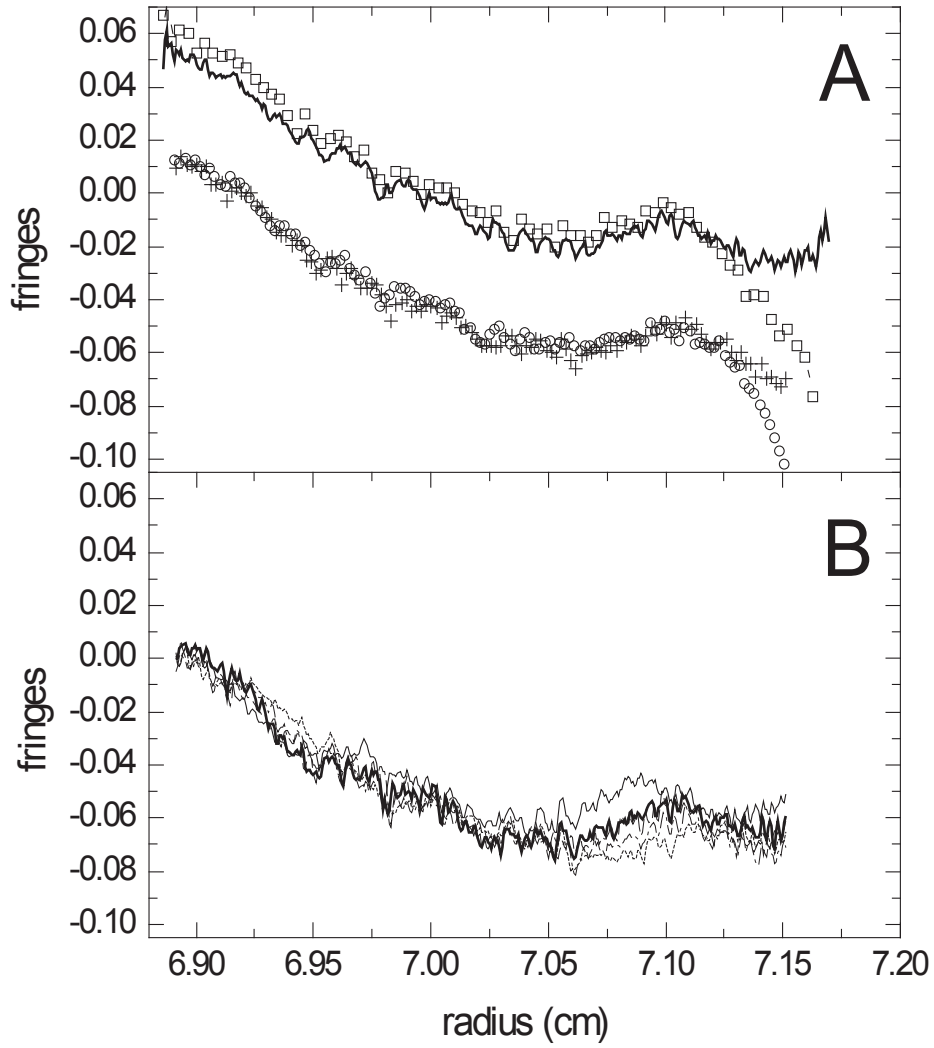
A
B
fringes
fringes
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.08
-0.10
6.90
6.95
7.00
7.05
7.10
7.15
7.20
radius (cm)

Figure 5

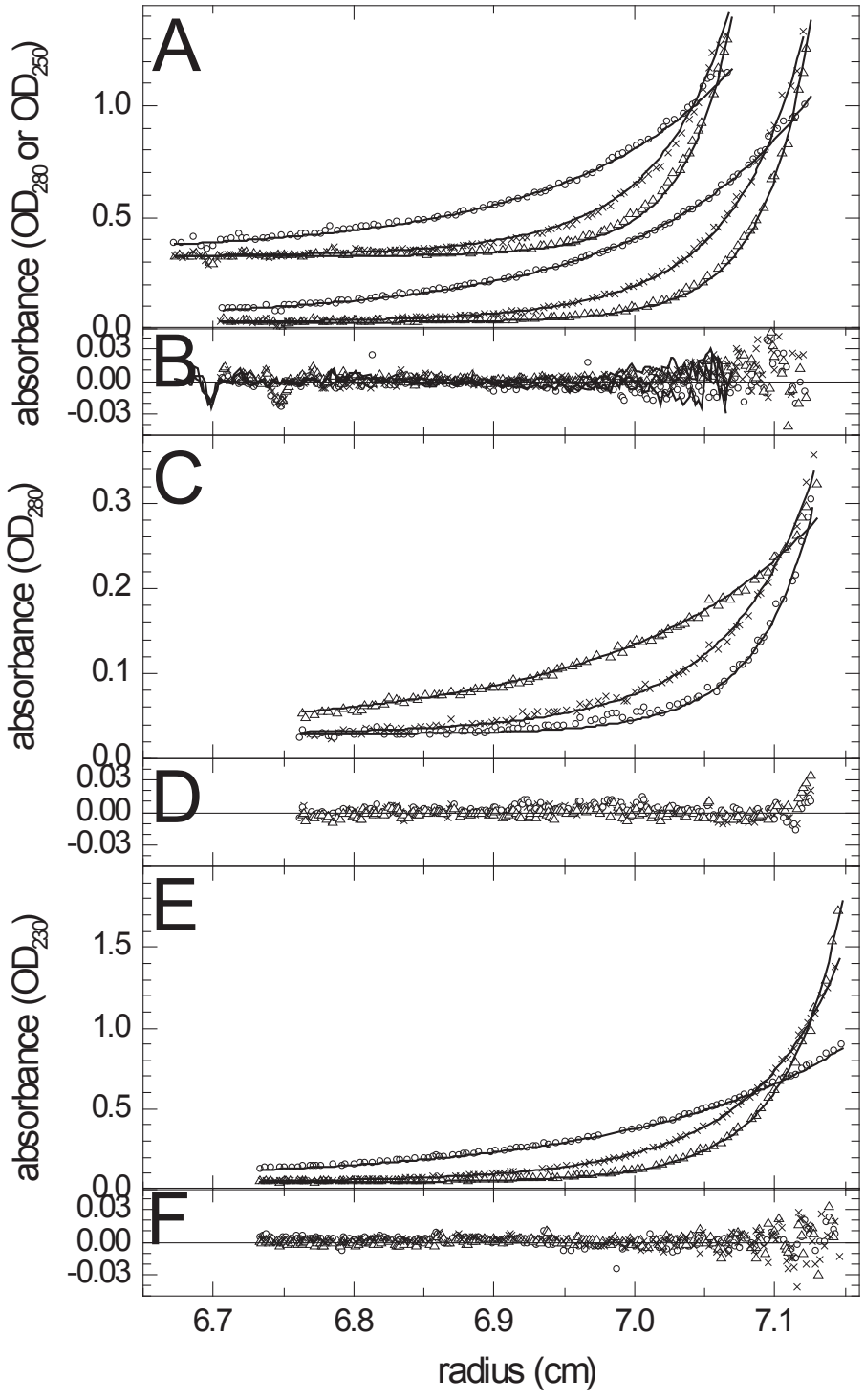
A
B
C
D
E
F
absorbance (OD280 or OD250)
absorbance (OD280)
absorbance (OD230)
1.0
0.5
0.0
0.03
0.00
-0.03
0.3
0.2
0.1
1.5
6.7
6.8
6.9
7.0
7.1
radius (cm)

Figure 6

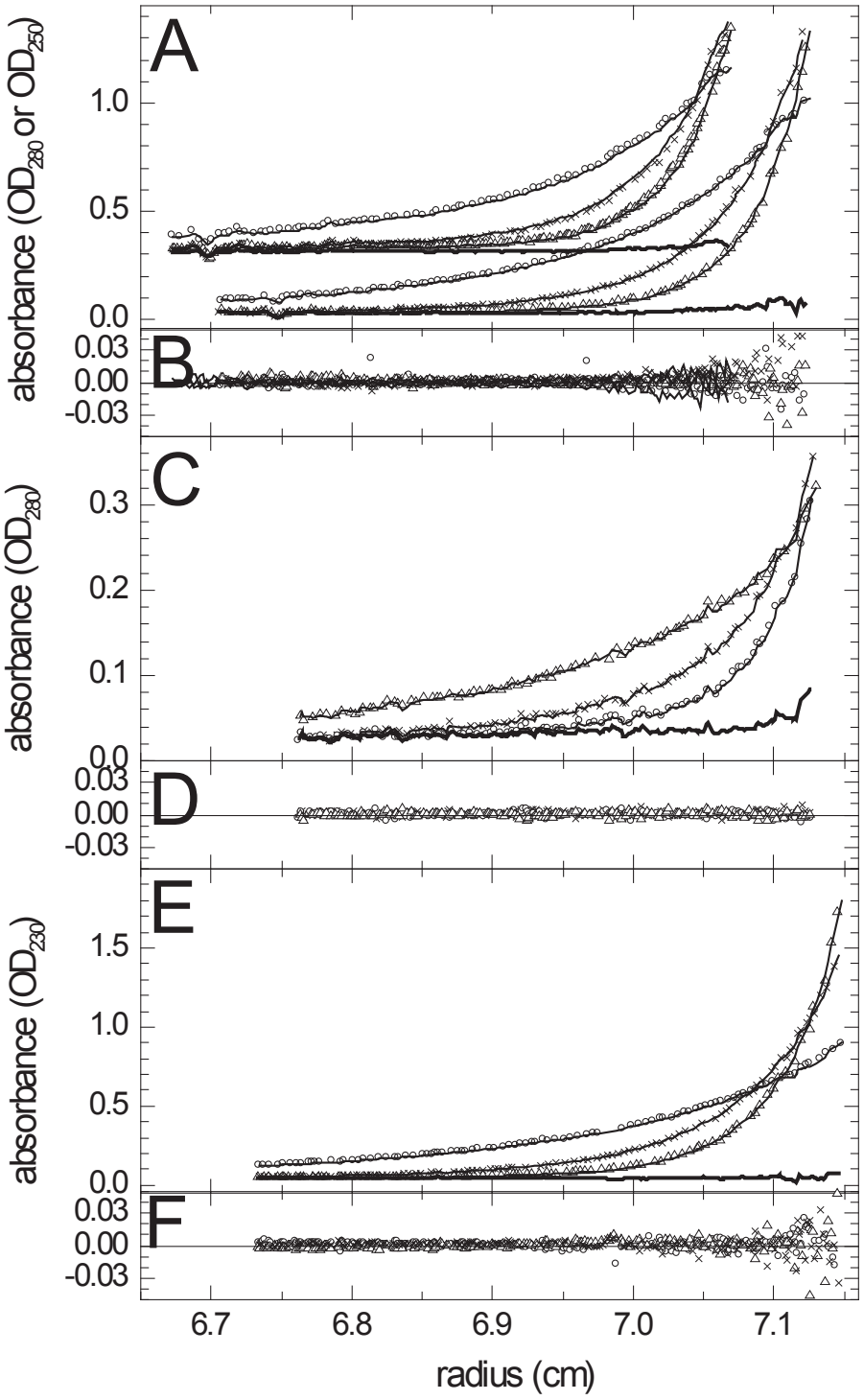
A
B
C
D
E
F
absorbance (OD280 or OD250)
absorbance (OD280)
absorbance (OD230)
1.0
0.5
0.0
0.03
0.00
-0.03
0.3
0.2
0.1
1.5
6.7
6.8
6.9
7.0
7.1
radius (cm)

Figure 7

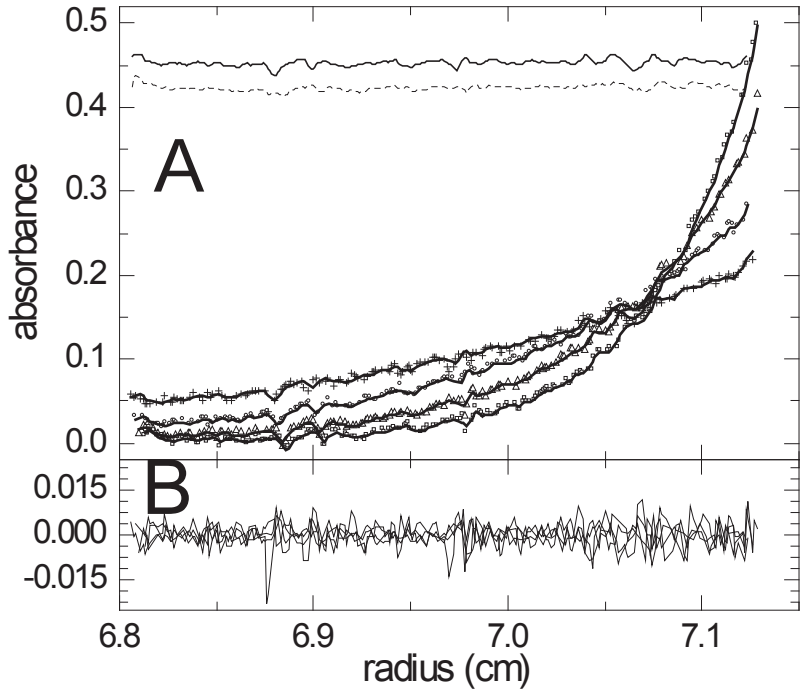
0.5
0.4
0.3
0.2
0.1
0.0
absorbance
A
B
0.015
0.000
-0.015
6.8
6.9
7.0
7.1
radius (cm)

Figure 8

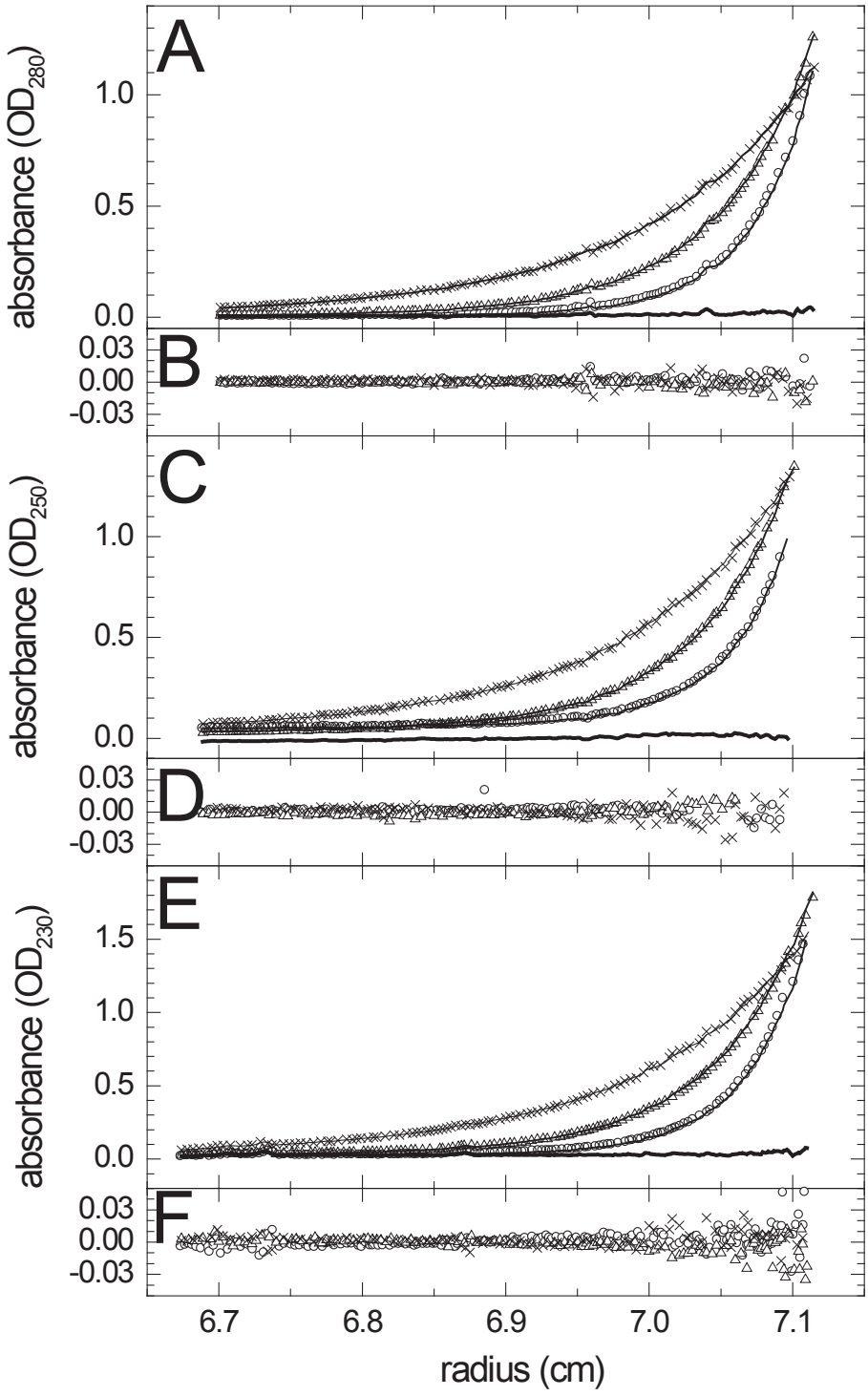
A
B
C
D
E
F
absorbance (OD280)
absorbance (OD250)
absorbance (OD230)
1.0
0.5
0.0
1.5
0.03
0.00
-0.03
6.7
6.8
6.9
7.0
7.1
radius (cm)

Figure 9