

**Chapter for “Modern Analytical Ultracentrifugation: Techniques and Methods” (D.J. Scott, S.E. Harding, A.J. Rowe, Editors) The Royal Society of Chemistry, Cambridge (in press)**

## **Diffusion-deconvoluted sedimentation coefficient distributions for the analysis of interacting and non-interacting protein mixtures**

Peter Schuck

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

### **\*Address for Correspondence and Proofs:**

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

phone: (301) 4351950  
fax: (301) 4966608  
email: pschuck@helix.nih.gov

## Summary

By modeling the evolution of concentration profiles measured in a sedimentation velocity experiment with a distribution of Lamm equation solutions it is possible to calculate diffusion-free sedimentation coefficient distributions,  $c(s)$ . For non-interacting or slowly interacting mixtures of sedimenting macromolecules, these report on the number and  $s$ -value of the species present in solution. For rapidly interacting mixtures, the sedimentation boundaries and peaks in  $c(s)$  can be viewed as approximations of the asymptotic reaction boundary from Gilbert-Jenkins theory, which describe characteristic features of the sedimenting system, and provide a rigorous approach to the weight-average sedimentation coefficient. The extension to sedimentation coefficient distributions of the individual components in a mixture,  $c_k(s)$ , calculated by global modeling of multiple signals from the sedimentation process, can reveal the stoichiometry of protein complexes. This chapter describes the theory, some aspects related to the experimental setup, and illustrates the practical use of  $c(s)$  and  $c_k(s)$  in several examples.

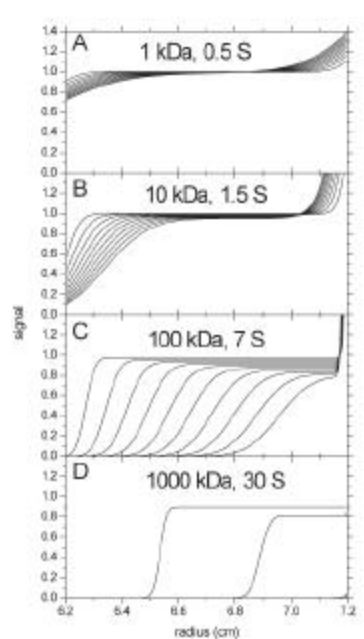
## Introduction

The analysis of the macromolecular redistribution following the application of a high gravitational field provides very rich information on macromolecules and macromolecular mixtures<sup>1,2</sup>. In the past, sedimentation velocity (SV) analytical ultracentrifugation has played an important role in several fields, including physical chemistry, polymer chemistry, biochemistry, and molecular biology, and it is continuing to provide unique information for the study of both synthetic and biological macromolecules in solution. It can permit the determination of homogeneity, molar mass, and gross shape of macromolecules, the thermodynamic and kinetic characterization of protein-protein interactions, as well as the detection of ligand-induced conformational changes. For both self-association and heterogeneous associations of proteins, frequently the number and type of complexes formed and the time-scale of the interaction can be discerned and the equilibrium constants can be determined.

The theoretical description of the sedimentation process has long been established, starting in Svedberg's laboratory in 1929 with Lamm's derivation of the partial differential equation for macromolecular sedimentation and diffusion in the centrifugal field<sup>3</sup>. Several generations of physical chemists have continued this work and contributed to the theoretical and experimental development of this technique. SV is still a modern tool and active area of research, which has rapidly developed in the last decade as the Lamm equation has become tractable in routine data analysis<sup>4-11</sup>, and increasingly more comprehensive models for the sedimentation process can be used. In conjunction with modern data acquisition systems of the commercial analytical ultracentrifuges, this has permitted the analysis of significantly increased data sets comprising one or more signals from the complete sedimentation process and across the whole solution column.

In the present communication, we review how Lamm equation solutions can be used in an approach of direct least-squares modeling of the complete sedimentation data in order to describe distributions of sedimenting particles. For samples that exhibit a continuous, more or less broad distribution of sizes, such as carbohydrates, synthetic polymers, or even populations of single-cell organisms, this can lead to a characterization of the shape and polydispersity of the distribution of sedimentation parameters. For samples that consist of a few discrete macromolecular species this description can help to identify the number, size, and  $s$ -value of the sedimenting species, as well as molecular interactions. The latter is particularly important as analytical ultracentrifugation can provide unique information on protein-protein interactions and the formation of multi-protein complexes. Key features of the modeling with distributions of Lamm equation solutions, termed  $c(s)$ , are that both diffusion and the characteristic noise structure of the sedimentation data can be deconvoluted, leading

to high resolution and sensitivity. It also lends itself to the incorporation of multiple signals from the sedimentation process into a global analysis, resulting in component sedimentation coefficient distributions, for example, revealing stoichiometry of mixed protein complexes. The present chapter gives an overview of the theory, practice, and limitations of this approach.



**Figure 1:** Characteristic shapes of sedimentation profiles simulated for globular species of 1 kDa (A), 10 kDa (B), 100 kDa (C), and 1 GDa (D) at 50,000 rpm and unit loading concentration. Scans are depicted in time-intervals of 10 min after start of centrifugation. Larger size-range can be accommodated easily by adjusting the rotor speed and the time-interval of data acquisition.

## Theory

An unknown macromolecular mixture in a sedimentation velocity experiment can be characterized via a differential distribution of sedimentation coefficients,  $c(s)$ , where the concentration of species with sedimentation coefficients between  $s_1$  and  $s_2$  is given by the integral  $\int_{c_1}^{c_2} c(s)ds$ . In order to calculate the distribution  $c(s)$  consistent with least-squares data analysis, we require that if we sum over the sedimentation signals of all species the resulting total signal matches the experimental data as close as possible. This can be expressed as

$$a(r,t) \cong \int_{s_{\min}}^{s_{\max}} c(s)\chi_1(s,D(s),r,t)ds \quad (1)$$

<sup>12</sup> with  $a(r,t)$  denoting the experimental signal at radius  $r$  and time  $t$ , and  $s_{\min}$  and  $s_{\max}$  the minimum and maximum  $s$ -value of the distribution.  $\chi_1(s,D,r,t)$  represents the characteristic signal pattern that a single macromolecular species at unit concentration would produce (Figure 1). The latter was given by Lamm in the form of a differential equation for sedimentation and diffusion

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

<sup>3</sup> where  $r$  denotes the distance from the center of rotation,  $\omega$  the rotor angular velocity, and  $s$  and  $D$  the

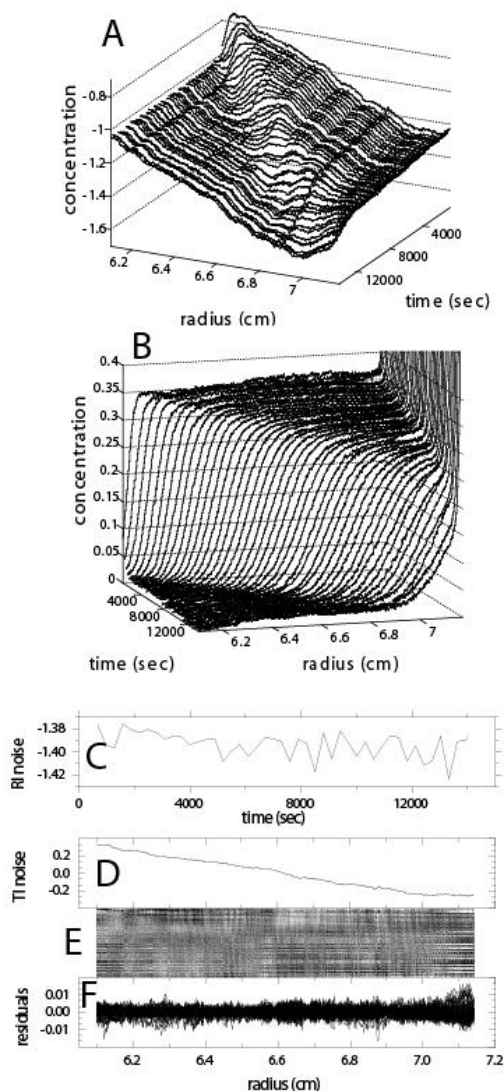
macromolecular sedimentation and diffusion coefficient, respectively. Adjustments can be made to account for compressible solvents, or sedimentation in dynamic density gradients<sup>13,14</sup>. It is apparent from Figure 1 that the sedimentation profiles  $c_1(s,D,r,t)$  exhibit large qualitative differences and are strongly size-dependent. (An exception to the latter is very elongated structures, which can reach a limiting  $s$ -value<sup>15</sup>). Accordingly, the modeling problem in Eq. 1 is relatively well-posed – it consists essentially in unraveling which combination of sedimentation patterns will fit the observed data best<sup>12</sup>. Numerical details how Eqs. 1 and 2 can be solved are described in<sup>16</sup>. This method and those described in the following are implemented in the programs SEDFIT and SEDPHAT. It can be obtained from the website [www.analyticalultracentrifugation.com](http://www.analyticalultracentrifugation.com), where tutorials and detailed information on the use of the software can also be found.

It should be noted that sedimentation velocity is conceptually a comparatively simple experiment, consisting only of the application of a gravitational force to the macromolecules in free solution, and, as a consequence, the sedimentation/diffusion process is very accurately described by Eq. 1. This can make the analysis very quantitative. It should comprise the entire sedimentation process, starting from the first signs of non-uniform distributions close to the meniscus and bottom, to the complete depletion of all material within the observable solution column, or the attainment of sedimentation equilibrium (Figure 2A). Systematic noise components in the data acquisition can be taken into account easily using algebraic methods. They lead to additional terms in Eq. 1 that explicitly express the systematic noise as radial-dependent baseline profiles and time-dependent but radially uniform offsets<sup>17</sup>. Least-squares optimal estimates of these noise components can be subtracted from the data without introducing bias (as long as the noise parameters are maintained as unknowns in the following analyses), which permits a better visual inspection of the sedimentation profiles and the details of the fit. Also, a best-fit meniscus position can be calculated and is usually well-determined by the data. This can lead to a higher degree of precision than permitted by the visual inspection of the optical artifact, with the latter being used to only for estimating limits for minimum and maximum values of the meniscus. In general, the distribution model Eq. 1 leads to fits within the noise of the data acquisition, as judged by the root-mean-square (rms) deviation of the fit and the absence of systematic misfit of the sedimentation boundary in the residuals bitmap, and this is a pre-requisite for the detailed interpretation of the results.

The numerical solution of the Fredholm integral equation Eq. 1 requires discretization of the range of  $s$ -values of the distribution. With a spacing of  $s$ -values on the order of 0.1 S, the inversion of Eq. 1 is generally well-conditioned and numerically stable. This is due to the characteristic shapes of  $c_1(s,D,r,t)$ , but also aided by the positivity requirement for the distribution  $c(s)$ . However, it can be shown mathematically that in the limit of very fine discretization of  $s$  the calculated distribution  $c(s)$  can become very noisy and many different curves  $c(s)$  may fit the data equally well within the noise of the data acquisition<sup>12</sup>. A well-established technique to suppress noise amplification in  $c(s)$  and to avoid detail in the distribution not warranted by the data is regularization. It is based on predefined measures of likelihood or parsimony of the distributions, such as maximum entropy

–  $\int c(s)\log(c(s))ds$  for discrete mixtures, or minimal total curvature  $\int |c''(s)|^2 ds$  for broader distributions. Regularization ensures that if there is any statistical ambiguity in the resulting distribution, the most likely distributions will be selected among all that fit the data within a pre-defined confidence limit<sup>18,19</sup>. The rationale is that, following Occam's razor, the solution with the highest parsimony is to be preferred. While this can introduce a small bias, by design it is certain that the selected result produces fits of the raw data that are of a quality statistically indistinguishable from the overall best fit. Regularization also provides a tool to explore alternate interpretations of the data, both by switching the regularization method and by changing the pre-defined confidence limit to

achieve weaker or stronger regularization (for example, going from a P-value of 0.7 to 0.95). If detailed features of the calculated distribution disappear on a higher confidence limit, they should be interpreted with care, while features that appear independent over a large range of P-values are details that are an integral part of the observed data.



*Figure 2:* Illustration of the data analysis of an IgG sample at 40,000 rpm. Panel A shows the raw data from interference optical data acquisition. Panel B shows the same data after subtraction of the systematic noise contributions, which are the radial-invariant time-dependent offsets (Panel C), and the time-invariant radial-dependent baseline profile (Panel C). The systematic noise was calculated as part of the  $c(s)$  analysis shown in Figure 3A. The rmsd of the fit was 0.0027 fringes, with the residuals in Panel F. The residual bitmap in Panel E is scaled to  $\pm 0.01$  fringes from black to white. It shows that the remaining noise is dominated by vertical and horizontal lines, which are likely caused by higher-order vibrations of the optical system. Importantly, no diagonal pattern is visible which would indicate the presence of a systematic misfit of the sedimentation boundary.

Complicating the analysis outlined above is the fact that two parameters,  $s$  and  $D$ , are required to specify the shape of the single macromolecular species sedimentation boundaries, but that we calculate only a one-dimensional sedimentation coefficient distribution. Several different options are available to arrive at an estimate for  $D$ . A very straightforward approach can be taken if all species in solution have the same diffusion coefficient, for example, distributions of apo-ferritin and ferritin,

which can be measured separately. If  $D$  is known, the Svedberg relationship

$$M(1 - \bar{v}\rho) = \frac{sRT}{D} \quad (3)$$

<sup>1</sup> (with the protein partial-specific volume  $\bar{v}$ , the solution density  $\rho$ , the gas constant  $R$  and the absolute temperature  $T$ ) can be used to transform the measured sedimentation coefficient distribution  $c(s)$  directly into a molar mass distribution  $c(M)$ <sup>12,20</sup>.

Another special case is that of negligible diffusion during the time of the sedimentation experiment, for example with very large particles. Here, the approximation  $D = 0$  can be used and the resulting sedimentation coefficient distribution is termed  $ls-g^*(s)$ <sup>21</sup>. The Lamm equation solutions simplify to step-functions for incompressible solvents, and to step-functions with sloping plateaus if solvent compressibility is taken into account<sup>14</sup>. Conceptually, the  $ls-g^*(s)$  distribution is related to  $g(s^*)$  by  $dc/dt$  or  $dc/dr$ , although there are large differences in the practical application, among them the applicability to data with steep boundaries and large time intervals<sup>21</sup>. If the  $ls-g^*(s)$  analysis is applied to the sedimentation analysis of particles that do exhibit significant diffusion, the distribution results in an apparent sedimentation coefficient distribution, which are convoluted by functions that are in first approximation Gaussians<sup>21</sup> from which for single species  $D$  and  $M$  may be estimated<sup>22</sup>. However, more direct approaches to combine single species analyses with distributions will be described below.

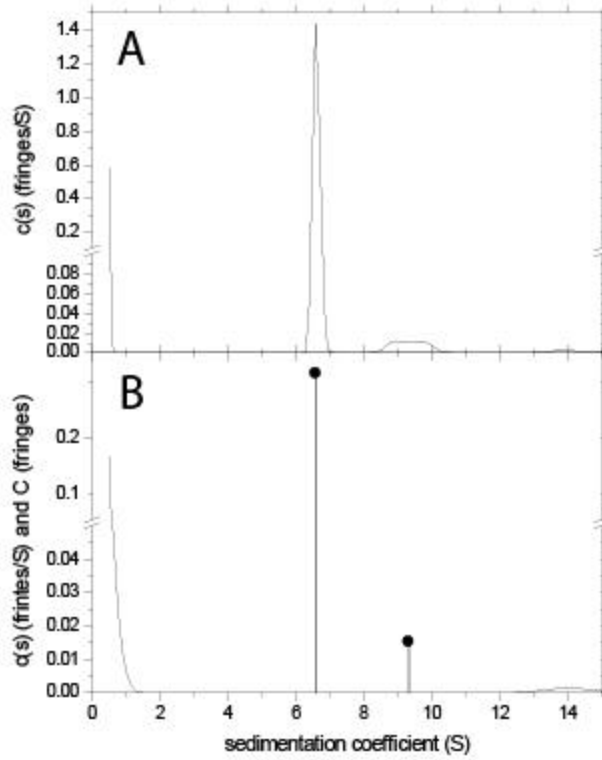
A general approach to arrive at realistic estimates for  $D$  in unknown distributions is based on the description of  $D$  as a monotonous single-valued function of  $s$ . The second assumption is made that, in a first approximation, all sedimenting species have similar frictional ratios  $f/f_0$ , which can be described with the weight-average frictional ratio  $(f/f_0)_w$ . For this case, we can derive the hydrodynamic scaling law

$$D(s) = \frac{\sqrt{2}}{18\pi} kT s^{-1/2} \left( \eta(f/f_0)_w \right)^{-3/2} \left( (1 - \bar{v}\rho) / \bar{v} \right)^{1/2} \quad (4)$$

(with  $k$  denoting the Boltzmann constant and  $\eta$  the solvent viscosity)<sup>16</sup>. Several observations are in favor of this approach: 1) It is well-known that the values of  $f/f_0$  are only very weakly dependent on the macromolecular shape<sup>23</sup> (examples for commonly observed values of the hydrated frictional ratio are 1.2 – 1.3 for relatively globular proteins, 1.5-1.8 for asymmetric or glycosylated proteins, and larger values for very asymmetric or unfolded proteins or linear chains). 2) Diffusion is a much less size-dependent property than the sedimentation coefficient (for a particle of radius  $R$ , it is  $D \sim R$ , while  $s \sim R^2$ ). 3) the peak positions of  $c(s)$  are largely independent of the precise value of  $(f/f_0)_w$ , which mainly affects the peak shape and the resolution of species with different sedimentation coefficients. In fact, a family of  $c(s)$  with different values of  $(f/f_0)_w$  – going continuously from  $c(s)$  with  $(f/f_0)_w = \infty$  (which is identical to the case of  $D(s) = 0$  in the  $ls-g^*(s)$  method and shows Gaussian broadened peaks) to  $c(s)$  with correct weight-average  $(f/f_0)_w$  (which shows sharp diffusion deconvoluted peaks) – will exhibit a constant peak position. 4) Since the boundary spreading from differential migration and from diffusion is different,  $(f/f_0)_w$  can be extracted by non-linear regression from the experimental data, in addition to the  $c(s)$  distribution<sup>20</sup>. 5) The extent of diffusion relative to sedimentation decreases with higher rotor speed. As a consequence, any errors in the approximations involved in the description of diffusion can be minimized at the highest rotor speeds, which optimize the hydrodynamic resolution. (Although  $c(s)$  is very well suited for sedimentation at high rotor speeds, it can also be used very effectively for data that exhibit a large boundary spread, such as from small molecules or from macromolecules at low rotor speed.) This is the standard  $c(s)$  approach used in most of the applications so far<sup>24</sup>. The results of the analysis for the data of Figure 2 is shown in Figure 3A.

If the weight-average frictional ratio  $(f/f_0)_w$  is well-determined by the data (which can be tested by

comparing the rms deviation of  $c(s)$  models with non-optimal  $f/f_0$  values), and if the distribution shows a single major peak,  $c(s)$  can be transformed using Eq. 3 to a molar mass distribution  $c(M)$ . Ordinarily, this will give estimates within 10% of the true molar mass (provided the correct partial-specific volume is known), which can provide a reliable answer, for example, if the protein represented in a major  $c(s)$  peak is monomeric or dimeric. Caution should be applied, however, since, in contrast to  $c(s)$ , the peak locations of  $c(M)$  will be strongly dependent on that the  $(f/f_0)_w$  is correct for the individual species, which may not be true, in particular, for trace components.



*Figure 3: Panel A:  $c(s)$  distribution from the analysis of the data shown in Figure 2. Panel B: hybrid continuous/discrete distribution (Eq. 5) with two discrete species of best-fit apparent molar mass of 159 kDa and 6.59 S, and a second species constrained to be of twice the molar mass and with best-fit  $s$ -value of 9.32 S. The continuous sections cover the range from 0.5 to 5 S, and 12 to 20 S, with a fix  $f/f_0$  of 1.3. It should be noted that the molar mass values of the discrete species was calculated independently of the  $f/f_0$  of the continuous sections.*

A more precise approach to extract molar mass information from the sedimentation profiles is the combination of the continuous distribution  $c(s)$  with discrete species, as in the hybrid discrete/continuous distribution model of SEDPHAT. Here, several continuous segments can surround intervals of  $s$ -values that contain only one or more discrete species:

$$a(r, t) \cong \sum_i C^{(i)} c_1(s_i, M_i, r, t) + \sum_j \int_{s_{\min, j}}^{s_{\max, j}} c^{(j)}(s) c_1(s, (f/f_0)_{w, j}, r, t) ds \quad (5)$$

with  $c^{(j)}(s)$  denoting the population of species with sedimentation coefficients in different non-overlapping intervals  $I_j = [s_{\min, j}, s_{\max, j}]$  which can be characterized by separate frictional ratios, in combination with discrete species at loading concentrations  $C^{(i)}$  (in signal units) with  $s$ -values  $s_i$  outside the intervals  $I_j$ <sup>25</sup>. The discrete species are described by Lamm equation solutions, each with the two parameters  $s$  and  $M$ , and are not connected to the determination of  $(f/f_0)_{w, j}$  for the continuous segments. In the implementation of SEDPHAT, up to 10 discrete species can be fitted, and their molar

mass values can be constrained to be multiples, for example, of a monomer molar mass. Such an analysis can make use of the prior knowledge of the molar mass relationship and discrete nature of the sedimenting species, while at the same time providing enough flexibility to describe in the continuous  $c(s)$  segments possibly present larger species, such as aggregates, as well as contaminations with smaller molar mass or degradation products. An example is shown in Figure 3B for an antibody sample. If multiple sedimentation velocity data sets are available from the same sample, they can be combined in a global analysis with the sedimentation coefficient and molar mass values of the discrete species treated as global parameters.

For the practical application of the  $c(s)$  analysis, it is important to select the limits  $s_{\min}$  and  $s_{\max}$  of the distribution wide enough to allow for the description of all sedimenting species. However, a range much wider than necessary will significantly increase the computational time. A good range can be ensured by observing the calculated distribution at the extreme values. A non-zero value of  $c(s_{\max})$  indicates the possible presence of material sedimenting faster than  $s_{\max}$ , and suggests that this value should be increased. A nonzero value of the distribution at the minimum value  $s_{\min}$ ,  $c(s_{\min})$ , can indicate either the presence of material sedimenting slower than  $s_{\min}$ , in which case  $s_{\min}$  should be reduced, or it can be caused by a correlation of the theoretical sedimentation pattern at  $s_{\min}$  with the baseline parameters (this can be the case, in particular, for experiments at low rotor speed and with short experiment times). The two cases can be distinguished by comparing the distribution  $c(s)$  with the result after lowering the value of  $s_{\min}$ : if  $c(s_{\min})$  increases,  $s_{\min}$  is in the regime of baseline correlation; alternatively, if  $c(s_{\min})$  decreases, a smaller species was likely present and of  $s_{\min}$  should be further reduced until  $c(s_{\min}) = 0$ , or until  $s_{\min}$  is in the regime of the baseline correlation. In case of a baseline correlation,  $s_{\min}$  can be kept if it is well below the species of interest, but  $c(s_{\min})$  should be ignored. In this case, the correlation at  $s_{\min}$  only indicates the over-parameterization of the description of small molecule sedimentation, which is without consequence and does not introduce a bias into the results of the remaining  $c(s)$  distribution. Only the reverse situation – a constraint in the description of sedimentation – would create a bias and must be avoided.

A related practical question is that of the discretization of the  $s$ -range and the spacing of the values. Frequently, for studies on medium sized proteins, a density of grid points of 10 per S unit can be a good compromise between ‘smoothness’ of the calculated distribution and excessive computational time. It should be verified that the numerical resolution of  $s$ -values is sufficiently high not to constrain the quality of the fit.

For the characterization of sedimentation coefficient distributions of multi-component mixtures of spectrally dissimilar macromolecules, it is useful to follow the sedimentation process simultaneously at different characteristic signals and to acquire multiple data sets  $a_{\lambda}(r,t)$ <sup>26</sup>. For the analysis, it is possible to generalize Eq. 1 and formulate component sedimentation coefficient distributions  $c_k(s)$  as

$$a_I(r,t) \cong \sum_{k=1}^K \mathbf{e}_{kI} \int_{s_{\min}}^{s_{\max}} c_k(s) \mathbf{c}_1(s, (f/f_0)_{k,w}, r, t) ds \quad \mathbf{I} = 1 \dots \Lambda, K \leq \Lambda, \det(\mathbf{e}_{kI}) \neq 0 \quad (6)$$

where each macromolecular component  $k$  is described as contributing in a characteristic way to the signal  $\lambda$  according to a molar extinction coefficient (or molar signal increment) matrix  $\mathbf{e}_{kI}$ . This matrix can be determined from separate experiments of the individual sedimenting components. An example is the analysis of protein-protein interactions between proteins of different weight-based extinction coefficients: Here, the acquisition of absorbance data simultaneous to the interference data allows distinguishing the sedimentation behavior of each protein component in the mixture. The component sedimentation coefficient distributions  $c_k(s)$  are in molar units, and their ratio allows unraveling protein stoichiometries in complexes.



Refinements are possible, for example, by constraining different intervals of sedimentation coefficients to certain ranges of possible complex stoichiometries<sup>26</sup>. Another extension is the combination with discrete species analogous to the hybrid continuous/discrete model of Eq. 5. This can take the form

$$a_1(r, t) \cong \sum_i \sum_{k=1}^K e_{kI} C_k^{(i)} c_1(s_i, M_i, r, t) + \sum_j \sum_{k=1}^K e_{kI} \int_{s_{\min,j}}^{s_{\max,j}} c_k^{(j)}(s) c_1(s, (f/f_0)_{k,j,w}, r, t) ds \quad (8)$$

with  $C_k^{(i)}$  representing the concentration of protein component  $k$  sedimenting as a discrete species  $i$  (characterized by the sedimentation coefficient  $s_i$  and the molar mass  $M_i$ ), and  $c_k^{(j)}(s)$  the sedimentation coefficient distribution of component  $k$  in the continuous segment between  $s_{\min,j}$  and  $s_{\max,j}$ .

## Experimental

The practical techniques for conducting sedimentation velocity experiments have long been worked out in great detail at the highest level of sophistication<sup>1,2</sup>. The direct least-squares modeling of the sedimentation profiles comprising the entire sedimentation process to within the noise of the data acquisition requires a certain experimental design, and introduces particular sensitivity towards certain experimental aspects which I will comment on in the following.

As is well-known, macroscopic convection can be caused by temperature gradients, and therefore thorough temperature equilibration of the rotor is necessary. In our experience with current instrumentation, after the temperature set point and the temperature reading are within 0.2°C, extending the equilibration time to additional 60 minutes has given best results. During this equilibration period, the rotor should be at rest, since rotation – even at a low rotor speed – will not only cause some macromolecular sedimentation but also lead to misleading entries in the scan files for the elapsed time after start of centrifugation. (This is in contrast to some other sedimentation and analysis techniques, which may be less sensitive towards initial rotation.) Rotor acceleration can be accounted for in SEDFIT and SEDPHAT in the finite element solution of the Lamm equation. If initial convection does occur, it may result in larger residuals close to the meniscus predominantly at the beginning of sedimentation. Unfortunately, because the further sedimentation will depend on the entire history of the experiment, later scans will also be affected in their information and the detailed analysis is problematic even if the initial scans are removed from the analysis and an apparently high quality of fit is obtained.

Regarding the loading composition of the sample, the  $c(s)$  method requires the absence of repulsive non-ideal sedimentation, and accordingly, low macromolecular concentrations are required. Fortunately, the optical systems are highly sensitive, and concentration below 0.05 mg/ml can routinely be studied with both the absorbance and the interference optical detection system. For the interference optical system, the solution in the sample and reference sector should be precisely matched in volume and chemical composition, in order to cancel the potentially large signal originating from buffer salts. It is possible to analyze unmatched samples by explicitly modeling the redistribution of the buffer salts (see below), but frequently at the cost of some details that can be reliably extracted from the analysis.

The highest possible rotor speed will give the best hydrodynamic resolution, and in our laboratory, we routinely use Epon centerpieces at rotor speeds of 50,000 rpm and 60,000 rpm after verifying the absence of leaks<sup>27</sup>. Intermediate rotor speeds contain more molar mass information, and the method can also be applied – at lower resolution – to data at low rotor speeds and the approach to equilibrium. Regarding the period of time for data acquisition, if possible it should extend from the beginning of the sedimentation to the complete depletion or the absence of further observable migration. The Lamm equation modeling approach is not restricted in the consideration of early scans or large time intervals, and these can carry valuable information on the absence or presence of very small and very large particles, as well as the shape of the baseline profiles.

As indicated above, the meniscus position should be included as a fitting parameter to be optimized by non-linear regression, because it generally cannot be determined experimentally with sufficient precision. It may not coincide with the peak maximum of the characteristic optical artifact, but the latter can be used as a guide for minimal and maximal values for this parameter. Similarly, if the sedimentation data to be analyzed include regions of back-diffusion from the bottom of the solution column, or if a small molar mass species is included in the model, the bottom position of the solution column should also be determined by non-linear regression.

A step-by-step experimental protocol for novices in analytical ultracentrifugation which also includes the basic steps for data analysis in SEDFIT and SEDPHAT for sedimentation velocity and sedimentation equilibrium can be found in<sup>27</sup> and<sup>28</sup>.

## Results

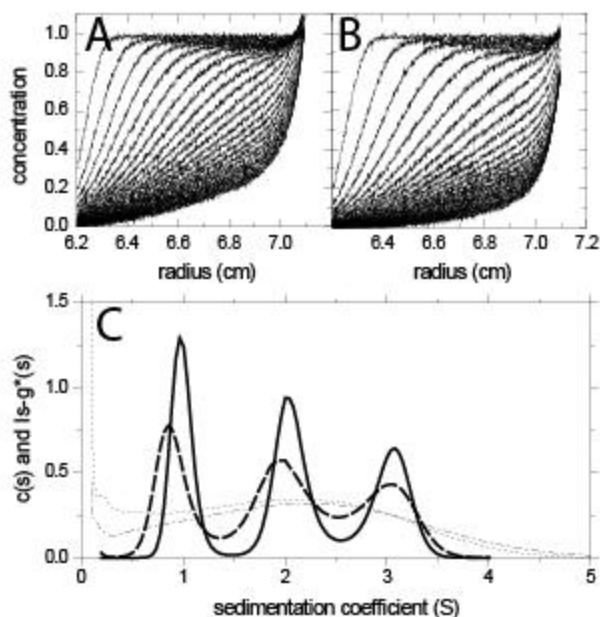
Figure 2 illustrates the results of the  $c(s)$  model in the original data space. The interference optical raw data, shown in Figure 2A, are clearly superimposed by well-known systematic radial-dependent baseline offsets ('TI noise'). Additionally superimposed are time-dependent offsets constant in radius which ('RI noise') are less well visible in the presentation of Figure 2A. As a consequence, the  $c(s)$  boundary model included parameters for TI and RI offsets. It was fitted to the raw data, including meniscus position and the weight-average frictional value as unknown parameters. After the  $c(s)$  fit has converged at an rms deviation of 0.0027 fringes, the raw data was decomposed into the different contributions. Figure 2B shows the calculated signal after subtraction of the systematic signal offsets. The time-dependent offsets are shown in Figure 2C, and the radial-dependent baseline is shown in Figure 2D. The residuals of the fit are encoded in a grayscale and shown as a bitmap in Figure 2E, with lines of pixels corresponding to different scans (first scan at the top, last at the bottom), and rows of pixels corresponding to different radial positions. In this representation, systematic misfits of the boundary would be visualized as diagonal patterns. These appear to be absent in Figure 2E, and the horizontal and vertical stripes suggests that part of the residuals likely originate from vibrations or thermally induced noise. The magnitude of the maximal residuals is shown in Figure 2F. In general, we found a similar procedure useful in the analysis of absorbance optical data, which appeared to differ mainly in the absence of time-dependent offsets ('RI noise') and in the radial-dependent baseline patterns exhibiting smaller slopes (data not shown). The high quality of the fit shows that the calculated  $c(s)$  distribution represents the data well, which justifies a detailed interpretation.

The calculated  $c(s)$  distribution is shown in Figure 3A. It exhibits a major peak for the monomeric species at ~ 6.5 S (without buffer correction), a baseline-separated a broader peak at 9 – 10 S, comprising of 4.8% of the material, and traces of higher oligomers at 14 S. Next, we made the assumption for the 9 S species to be dimers, enabling a more refined analysis with the hybrid discrete/continuous model of Eq. 5. We inserted two discrete species at 6.5 S and 9 S, constraining their molar mass to be in the ratio 2:1, and treated the monomer molar mass and the  $s$ -values of the

monomer and dimer as unknowns. In order to avoid trace contamination of other species to bias these estimates for the discrete species, they were combined with two sections of continuous sedimentation coefficient distributions, from 0.5 – 5 S and 12 – 20 S. The resulting distribution is shown in Figure 3B.

This example indicates the sensitivity of the  $c(s)$  analysis. Simulations show that from a set of profiles covering the complete sedimentation process at a signal-to-noise ratio of 200:1 (which can be readily achieved, e.g. at a loading concentration of 0.3 to 0.4 mg/ml of protein in the interference optics), minor peaks consisting of 0.2% of the total protein concentration should in theory be detectable with statistical confidence. The statistical accuracy of the calculated  $c(s)$  distribution can be assessed by Monte-Carlo simulations<sup>29</sup>. For some applications, it may be advantageous to apply the statistical tests to the integral of  $c(s)$  over a certain  $s$ -range, rather than calculating the error contours of the  $c(s)$  distribution.

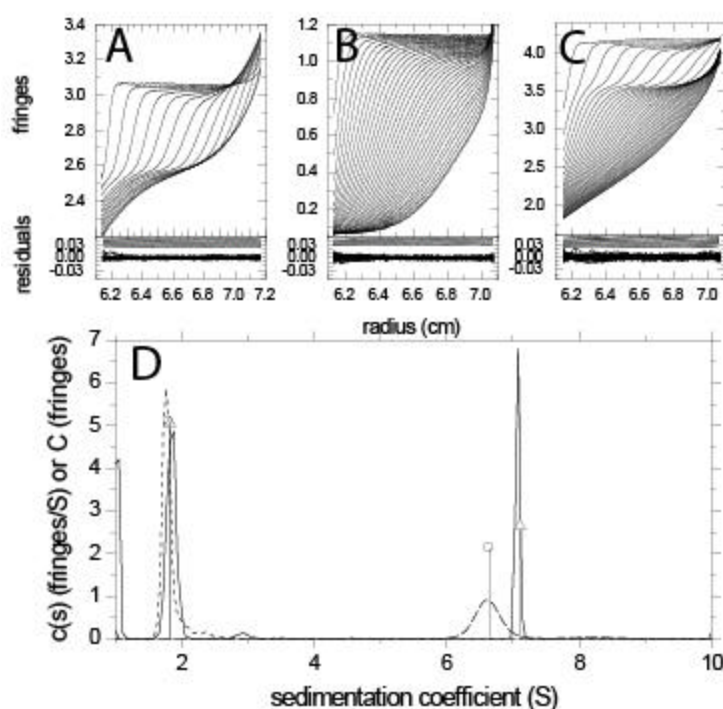
In order to test the resolution of  $c(s)$  for sedimentation profiles that are governed by diffusional spread, sedimentation velocity experiments were simulated for a mixture of species with sedimentation coefficients of 1 S, 2 S, and 3 S at equal concentration. Even at rotor speeds of 50,000 and 60,000 rpm diffusion broadening does not allow clearly sedimentation boundaries to establish (Figure 4). The resulting  $c(s)$  distributions show clearly defined peaks. Limitations in the deconvolution of diffusion are encountered at the lower rotor speed, which shows a degrading resolution and a small shift in the 1 S peak.



**Figure 4:** Resolution of the  $c(s)$  distribution for mixtures of species with low molar mass. Sedimentation profiles were simulated for 3 species with 6 kDa and 1 S, 10 kDa at 2 S, and 20 kDa and 3 S, respectively, at a loading concentration 0.33 each and with 0.01 simulated noise. Sedimentation was simulated for data acquisition in 10 min intervals (every 2<sup>nd</sup> scan shown) for rotor speeds of 50,000 rpm (Panel A) and 60,000 rpm (Panel B). Panel C shows the  $c(s)$  distributions calculated assuming the presence of TI noise, an unknown meniscus position and optimizing the weight-average frictional ratio.  $c(s)$  from the 50,000 rpm data (dotted line) and from the 60,000 rpm data (solid line).

A second example is the interaction between D1.3 monoclonal antibody and hen egg lysozyme (Figure 5). Panel A to C show the fringe displacement data of D1.3 (A), lysozyme (B), and a mixture (C) after subtraction of the systematic noise components. A sloping signal offset can be discerned in the data from the sedimentation of the antibody (A), which are signals from the redistribution of unmatched buffer salts superimposed to the macromolecular sedimentation. These are modeled as an

extra discrete species at 0.12 S and buoyant molar mass of 55.4 Da (best-fit estimates from the  $c(s)$  analysis). Compared to the  $c(s)$  peak of the antibody alone (dashed line), the mixture (solid line) shows a peak shifted by 8% (0.5 S) towards a larger sedimentation coefficient, indicating binding of lysozyme to the antibody. Because the concentrations used are > 100fold the equilibrium dissociation constant and lysozyme is in large excess, it can be assumed that all available sites on D1.3 in this antibody preparation are saturated and the 7.1 S peak in the mixture represents the antibody-antigen complex, possibly in addition to contaminations with unreacted antibody.



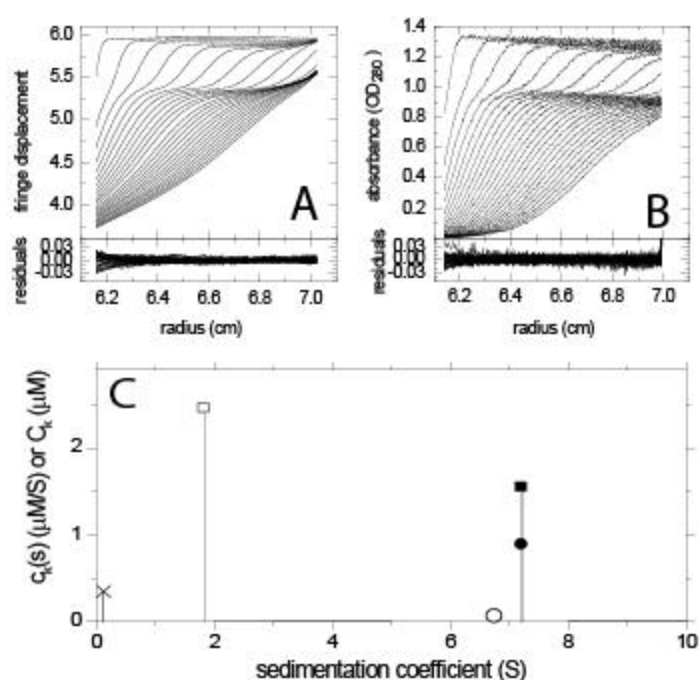
**Figure 5:** Analysis of the interaction between D1.3 monoclonal antibody and hen egg lysozyme, using interference optical fringe displacement data acquired in sedimentation velocity experiments at 50,000 rpm, 21 °C. Panels A – C: Fringe displacement profiles (every third scan shown) of D1.3 (A), hen egg lysozyme (B), and a mixture of both (C). The residuals are from the  $c(s)$  analysis shown in Panel D. It leads to an rms deviation of 0.0028 (A), 0.0028 (B), and 0.0043 (c) fringes, respectively. The residuals bitmap was scaled to a maximum dynamic range of 0.02. Panel D:  $c(s)$  distribution of D1.3 (dashed line), hen egg lysozyme (dotted line), and the mixture (solid line). For comparison, the results of the hybrid discrete continuous model are indicated as open symbols, showing only the discrete species for the hen egg lysozyme (open circles,  $M^* = 15.6$  kDa), the D1.3 antibody (open squares,  $M^* = 150$  kDa), and the two macromolecular species observed in the mixture (open triangles,  $M^* = 15.1$  kDa and 175 kDa). The discrete species are in units of fringes, scaled 5fold for clarity.

It is interesting to examine how information on the stoichiometry of the complex could be extracted from the sedimentation data. Three different approaches are possible. First, the relative shift in the  $s$ -value may be taken to estimate the increase in mass with the 2/3-power rule, assuming the constancy of shapes. This would suggest an increase in mass by 12%, or ~ 17 kDa, which is 1.2fold the molar mass of the lysozyme.

A second source of information can be the measurement of the species molar mass from boundary spreading. Using the best-fit frictional ratios, the  $c(s)$  distributions can be transformed into molar mass distributions, which yields a peak of 163 kDa (apparent molar mass assuming  $\bar{v}\rho = 0.73$ ) for the antibody ( $(f/f_0)_w = 1.6$ ), and 15.8 kDa ( $(f/f_0)_w = 1.3$ ) for lysozyme, respectively. For the individual proteins, these values are within the typical precision of  $c(M)$  from the true molar mass

values. For the mixture, however, unrealistic values of 19.4 kDa for the free lysozyme and 141 kDa for the complex are obtained, with a best-fit  $(f/f_0)_w$  value of 1.4. The reason for the unsuccessful  $c(M)$  characterization of the mixture lies in the fact that the weight-average frictional ratio of 1.4 does neither reflect the frictional ratio of lysozyme nor the antibody well. It is too high for the lysozyme and leads to an overestimation of its mass, and *vice versa* for the antibody. Such a situation can be encountered if the  $c(s)$  distribution does not exhibit a single major peak. This problem may be addressed by calculating  $c(s)$  with two  $(f/f_0)_w$  values for different ranges of  $s$ -values. The prerequisite for this technique is that the sedimenting species are well separated in  $s$ -value, which is fulfilled in the present case.

A better approach for extracting the molar mass information from the data in Figure 5A-C is the hybrid discrete/continuous model (Eq. 5), if the peak regions of  $c(s)$  are replaced by discrete species for which  $s$ , and  $M$  are treated directly as unknown parameters. At the same time, continuous segments of  $c(s)$  can be used to model trace components at higher and lower  $s$ -values. These will use the scaling method Eq. 4, but the frictional ratio of the continuous segments are independent of the molar mass values for the discrete species. This method gives apparent molar mass values of 15.6 kDa for the lysozyme, 150 kDa for the antibody, and two species of 15.1 and 175 kDa for the mixture (open symbols in Figure 5D).

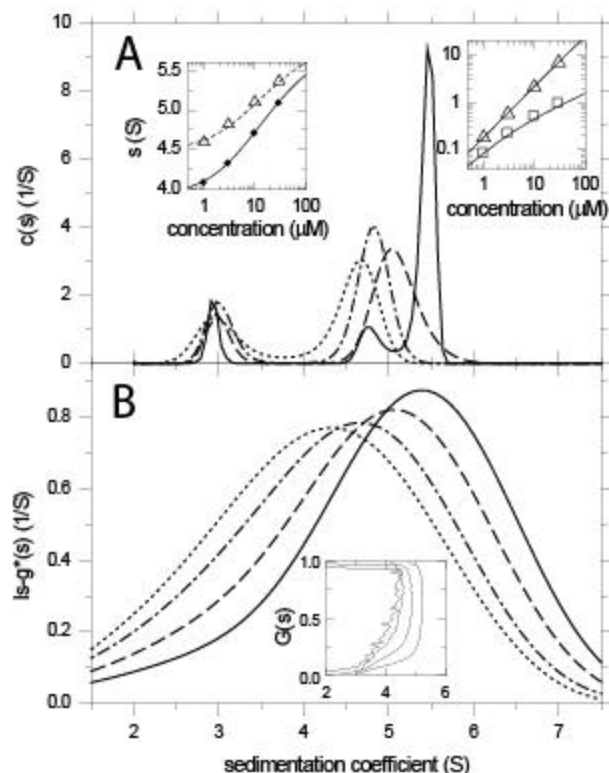


**Figure 6:** Multi-signal analysis of the antibody-antigen mixture shown in Figure 5C. Shown are the interference optical fringe displacement data (A) and absorbance optical data at 280 nm (B) acquired simultaneously during sedimentation. For clarity, only every third scan is shown. In a preliminary analysis, similar multi-signal data pairs from the antibody and the antigen were modeled, based on a theoretical molar interference optical signal increment ( $2.75 \times M_w$  fringes/Mcm), and the value for the protein molar extinction coefficient at 280 nm was determined. Values of  $2.22 \times 10^5$   $OD_{280}/Mcm$  and  $3.42 \times 10^4$   $OD_{280}/Mcm$  were obtained for the D1.3 antibody and lysozyme, respectively. This was used as a constraint in the analysis of the mixture. The model consisted of the following discrete species: of a small buffer component at 0.12 S contributing only to the interference data (cross), the free lysozyme (open square), free antibody (open circle), and the bound lysozyme (solid square) and antibody (solid circle). The bound species were constrained to be co-sedimenting by establishing a link in molar mass and sedimentation coefficient. This was combined with two continuous sections for both the lysozyme and the antigen component (not visible on this scale).

The third method to determine the complex stoichiometry consists in the global analysis of sedimentation data acquired with the refractive index sensitive interference system and the absorbance optical system (Figure 6A and B). With the wavelength set to 280 nm, the global analysis can exploit differences in the fraction of aromatic amino acids of the antibody and lysozyme. In general, this requires differences in the mg/ml based extinction coefficient of the two proteins. In the present case, although the mg/ml-molar extinction coefficients are significantly different (1.58 OD<sub>280</sub>/(mg/ml×cm) for IgG and 2.57 OD<sub>280</sub>/(mg/ml×cm) for lysozyme, respectively), a 2:1 complex and free antibody differ only by 10% due to the large difference in molar mass. The component sedimentation coefficient distributions  $c_k(s)$  according to Eq. 6 are shown in <sup>26</sup>, indicating a molar ratio of 1.8:1 and the presence of unreacted antibody. The result of the multi-signal hybrid discrete/continuous analysis following Eq. 8 is shown in Figure 6C. The molar ratio of the content of antibody and antigen in the 7.1 S component is 1.74:1, consistent with the previously reported results for the continuous multi-signal  $c_k(s)$  analysis <sup>26</sup>. The best-fit estimate for the molar mass of the complex species is 171 kDa.

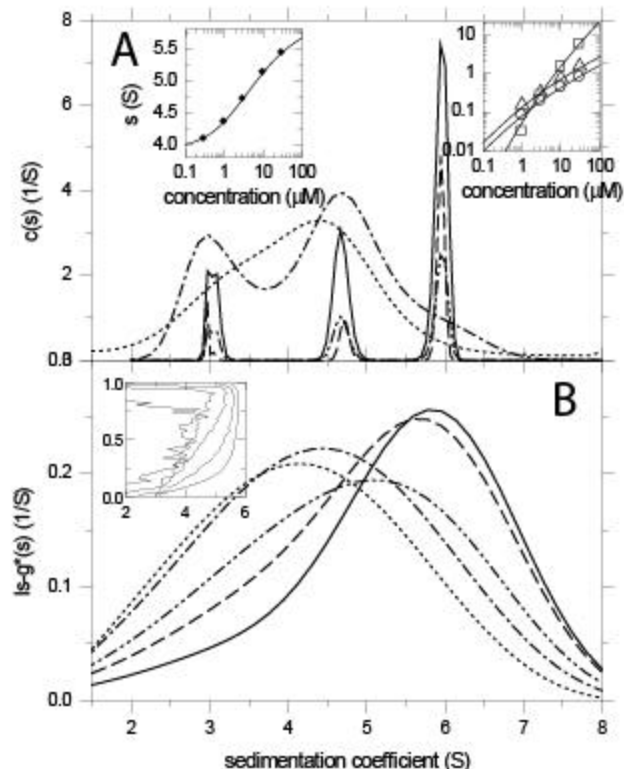
$c(s)$  is based on equations describing superpositions of non-interacting species. This is a good approximation for the application to mixtures of reversibly interacting proteins if the complexes formed in the loading mixture are maintained during the sedimentation experiment. Sufficient stability of complexes can be achieved either due to a low off-rate constant or due to loading concentrations far above or below the equilibrium dissociation constant. If these conditions are not fulfilled, the sedimentation coefficient distribution will be governed by the features of the sedimentation boundary. It is of interest to study how these features will be represented in the  $c(s)$  curves. For this purpose, using the finite element Lamm equation solution with reaction terms described in <sup>30</sup>, the sedimentation of an interacting system of the type  $A + B \leftrightarrow AB$  was simulated for interacting proteins of 30 kDa and 50 kDa, at different reaction rate constants, and at different concentrations of the components.

For fast reactions, Gilbert and Jenkins have developed a theoretical description for the asymptotic shapes of the sedimentation boundary at infinite time <sup>31</sup>, which predicts the sedimentation to proceed with an undisturbed boundary with the sedimentation rate of one of the free species, in combination with a reaction boundary that is composed of a mixture of both free species and the complex and sediments at a range of  $s$ -values in between the  $s$ -values of the larger free species and the complex. Figure 7A shows typical results obtained for a fast reaction (dissociation rate constant  $k_{\text{off}} = 0.01/\text{sec}$ , which can usually be considered instantaneous on the time-scale of sedimentation) at equimolar loading concentrations from 0.1 to 3fold  $K_d$ . Typical features of the  $c(s)$  distribution are a bimodal boundary shape, with a peak appearing at a concentration-independent  $s$ -value close to that of the smaller protein component, and a second peak at a concentration-dependent position. This is consistent with the predictions by Gilbert for the migration of the reaction boundary of such a system in the limit of no diffusion <sup>31</sup>. A quantitative analysis of the concentration dependence of the faster boundary component (calculated by integration over the faster  $c(s)$  peak) can yield valuable information on the binding constant and the  $s$ -value of the complex <sup>32</sup> (insets in Figure 7A). Similarly, isotherms of the relative signals of the two boundary components can be analyzed on the basis of Gilbert-Jenkins theory <sup>32</sup>. This can significantly improve the traditional analysis of the concentration dependence of the overall weigh-average sedimentation coefficient <sup>32-34</sup>. For comparison, the  $g^*(s)$  distributions calculated as  $ls-g^*(s)$  in Figure 7B show a single broad peak, from which no finer structure can be discerned and only the overall weight-average  $s$ -value be determined. Similarly, the integral sedimentation coefficient distributions  $G(s)$  from the van Holde-Weischet extrapolation method <sup>35</sup> does not allow to discern information on the time-scale of reaction, or any further quantitative analysis <sup>32</sup>.



**Figure 7:** Results of sedimentation coefficient distribution analysis applied to the rapidly interacting proteins. Calculations are based on a species with 30kDa and 3 S binding to a second species with 50 kDa and 4.5 S to form a 6 S complex with an equilibrium dissociation constant of  $K_d = 10 \text{ nM}$ , and an off-rate constant  $k_{off} = 1 \times 10^{-2}/\text{sec}$ .

Sedimentation was simulated for a rotor speed of 50,000 rpm, mimicking signal and noise of interference optical data acquisition, with 50 scans in intervals of 5 min subjected to the  $c(s)$  analysis. Component concentrations were equimolar at 1 (dotted line), 3 (dash-dotted line), 10 (dashed line), and 30  $\mu\text{M}$  (solid line). Panel A shows the result of the  $c(s)$  analysis, using meniscus, time-invariant noise, and weight-average frictional ratio as unknown parameters. The left inset shows the isotherms of the weight-average  $s$ -value (solid circles) and the  $s$ -value of the fast boundary component (open triangles), both determined by integration of the  $c(s)$  distribution, and the theoretical isotherms predicted from mass action law (solid line) and Gilbert-Jenkins theory (dotted line), respectively. The right insets shows the amplitudes in signal units of the undisturbed (squares) and reaction boundary component (triangles), and the theoretical isotherms predicted from Gilbert-Jenkins theory (solid lines). Panel B shows the apparent sedimentation coefficient distribution  $ls-g^*(s)$  without deconvolution of diffusion, applied to the scans from 50–100 min only to ensure a reasonable fit of the data. All distributions are scaled to unit area. The inset shows the integral sedimentation coefficient distributions from the van Holde-Weischet method, calculated with the least-squares method described in<sup>20</sup>.



**Figure 8:** Results of the sedimentation coefficient distribution analysis applied to slowly interacting proteins. Calculations are based on the same system as in Figure 7, but with  $K_d = 3.2 \text{ nM}$  and  $k_{off} = 3.2 \times 10^{-5}/\text{sec}$ . Component concentrations were equimolar at 0.3 (dotted line), 1 (dash-dotted line), 3 (dash-dot-dotted line), 10 (dashed line), and 30  $\mu\text{M}$  (solid line). All other parameters of the simulated sedimentation and the analysis were identical to those in Figure 7. Panel A shows the  $c(s)$  distributions, with the isotherm of the weight-average  $s$ -value (left inset, solid circles), and the isotherms of the concentrations of the individual species (right inset, in signal units) with free A (circles), free B (triangles), and complex (squares), and the corresponding theoretical isotherms based on mass action law (solid lines). Panel B shows the  $ls-g^*(s)$  distributions. All distributions are scaled to unit area, except curves for 1 and 3  $\mu\text{M}$  in Panel A, which are scaled 10fold for clarity. The inset shows the integral sedimentation coefficient distributions from the van Holde-Weischet method.

Figure 8A shows the  $c(s)$  distributions for a system with slower complex dissociation,  $k_{\text{off}} = 3.2 \times 10^{-5}/\text{sec}$  (for equimolar loading concentrations from 0.1 to 10fold  $K_d$ ). Peaks can be discerned at constant positions (the  $c(s)$  curves at the lowest concentrations are broadened due to a limited signal-to-noise ratio), resolving the sedimenting species. The reaction is reflected here in the changes of the relative heights of the peaks, reflecting redistribution according to mass action law. The isotherms of the weight-average  $s$ -value and the signal amplitudes of the three boundary components, which can be modeled jointly in a global analysis, are shown in the insets. Interestingly, the  $ls-g^*(s)$  distribution in Figure 8B again show a single broad peak, qualitatively similar to those in Figure 7B. Similarly, also the integral sedimentation coefficient distributions  $G(s)$  do not resolve the species and resemble those from Figure 7B.

In comparison, Figure 7A and 8A illustrate the qualitatively different behavior of the  $c(s)$  distribution for slow ( $k_{\text{off}} < 10^{-5}/\text{sec}$ ) and fast ( $k_{\text{off}} > 10^{-3}/\text{sec}$ ) reactions, which can be diagnosed from the concentration dependence of the  $c(s)$  distributions (going from concentrations significantly lower than  $K_d$  to concentrations significantly higher than  $K_d$ ). If the peak positions are changing, the reaction is fast on the time-scale of sedimentation, and the distribution reflects the features of the reaction boundary. If the peak positions are constant but changing only in heights, this suggests stability of the complex on the time-scale of sedimentation, and supports the interpretation of peaks as reflecting parameters of the sedimenting species. For fast reactions, the interaction can be characterized by integrating the two characteristic peaks and modeling the isotherm of  $s$ -values and boundary heights on the basis of Gilbert-Jenkins theory<sup>32</sup>, while for slow reactions the isotherms of populations of the individual species can be determined by integration of  $c(s)$ , which can be modeled by applying mass action law and mass conservation to the loading mixture. In either case, as shown in<sup>34</sup>, weight-average sedimentation coefficients can be obtained rigorously from the integration of the complete  $c(s)$  profiles. For any of the isotherm analyses, the integration can be conducted such that peaks from contaminating species can be excluded if they sediment outside the range of  $s$ -values of the interacting species.

## Discussion

The examples above illustrate that a detailed model for the sedimenting macromolecular mixture can be extracted from the least-squares modeling of the concentration profiles observed in sedimentation velocity analytical ultracentrifugation. This is consistent with the results reported previously on this methodology<sup>12,16,20</sup>, and with the results of applications to the study of many macromolecular mixtures<sup>24</sup>. Because the  $c(s)$  distributions are superpositions of Lamm equation solutions, the hydrodynamic resolution is not confined to clearly visible sedimentation boundaries, but extends to the discrimination of macromolecular species within broad boundaries, where the rms displacement from diffusion during the sedimentation process exceeds the separation from differential sedimentation<sup>20</sup>.

Data from the entire sedimentation experiment can be modeled, and the analysis can be combined with the algebraic methods to calculate best-fit estimates of the systematic noise contributions. This method provides the least bias and noise amplification, and allows to subtract the estimated systematic noise contributions from the raw data. Although these estimates are model-dependent (and therefore the consideration systematic noise offsets must be maintained during the subsequent analyses of the corrected data), it should be noted that degrees of freedom in the data interpretation resulting from the unknown offsets are inherent in the data and cannot be eliminated by any data pre-treatment without bias. However, we found that the visualization of the fit in the raw data space can be very instructive, and it can be very helpful for the design and interpretation of sedimentation experiment to verify the relationship of different boundary components or other data aspects to the peaks in the  $c(s)$



distribution. In our experience, details of the sedimentation coefficient distributions should be interpreted only after a critical inspection of the raw data and the quality of the fit, ensuring that all features of the sedimentation process in the sample have been described in the model. The latter also includes the sedimentation of unmatched buffer salt.

Sedimentation velocity can be a very sensitive tool for the detection of trace components, a topic of high importance in biotechnology applications<sup>36</sup>. In principle, if a sedimenting oligomeric species sediments faster than the leading edge of the diffusion envelope of a monomer, its presence will manifest itself in a slope (or migrating step) in the plateau region of the monomer. (Theoretically, the highest rotor speeds should promote this configuration.) Considering the large number of data points reporting on the shape of the plateau during the course of the experiment, it can easily be seen how the detection sensitivity can be far below the noise for a single data point. The  $c(s)$  and  $ls-g^*(s)$  distributions can extract this information as they permit the detailed modeling of data sets – including the plateau region – and since they can describe fast-moving species with large boundary displacement between scans. The characterization of species sedimenting within the limits of the diffusion broadened boundary of the monomer is difficult. The use of prior knowledge, such as the existence of discrete species with molar mass constraints combined with segments of continuous distributions (Figure 3B) is a possibility to improve the analysis. In this approach, the continuous segments may only describe the species of secondary interest outside the  $s$ -range of the discrete species, but they ensure that those are properly modeled and not bias the estimate of the species of interest.

The description of diffusion in the sedimentation profiles constituting the basis for the distribution (Figure 1) provides for the deconvolution of diffusion in the sedimentation coefficient distributions, which is the source of the relatively high resolution. The extent of diffusion for each species is estimated by the assumption that, in first approximation, all species have a very similar frictional ratio. Since the numerical value of the frictional ratio usually does not vary greatly, this will accommodate a large range of shapes. (It should be noted that there is no assumption of an actual specific shape or frictional ratio, since the value of  $(f/f_0)_w$  is determined by regression from the data.) By design, this approach will work best at the highest rotor speeds, when the extent of diffusion is small compared to sedimentation. Nevertheless, as illustrated in Figure 4, it can also work very well for resolving species with a large diffusional spread, caused either by low rotor speeds or by small macromolecule size.

If prior knowledge is available regarding the molar mass or dissimilar frictional shape of certain species (or if the course of the analysis leads to such conclusion), more flexibility of the  $c(s)$  analysis is possible by dividing the range of  $s$ -values into different regimes, and by combination with discrete species with or without molar mass constraints. Possible pitfalls of such more complex parameterizations of the distribution are correlations, but these can be avoided by considering the characteristic shapes of the sedimentation profiles provided for by the different segments and species assembled in the distribution. Further, we have shown previously that the discrete species in the hybrid distributions can be described by global parameters in the framework of global analysis of several sedimentation velocity data sets<sup>25</sup>. This can be the global analysis of different samples with common discrete species<sup>25</sup> or the global analysis of the same sample at different rotor speed<sup>37</sup>.

Although in the discussion above the purpose of estimating the extent of diffusion was essentially a means to arrive at higher resolution sedimentation coefficient distributions, of course, the diffusion itself can provide valuable information on the nature of the sample. As illustrated in the example of Figure 4A and B (corresponding to the dotted and dashed line in Figure 4D) the  $c(s)$  distribution may be converted to a  $c(M)$  molar mass distribution using the Svedberg relationship to calculate the molar mass from each  $s$ -value in the distribution and its estimated diffusion coefficient. In our experience this works well if a single major peak is present in the distribution (usually within 10% of the correct molar mass value, provided the correct density and partial-specific volume is known). This will ensure

that the weight-average frictional ratio is representing well the major species of interest. In many cases, such a determination of molar mass by  $c(M)$  analysis can be superior to the modeling with single or few discrete species<sup>16</sup>. In contrast, the peaks of less abundant fractions of the sample may be at an incorrect molar mass position if they exhibit significantly different frictional ratio from the calculated weight-average. Such a dependence of peak positions on the value of  $(f/f_0)_w$  is usually not observed in the  $c(s)$  distribution<sup>20</sup>. The example of the mixture in Figure 4C shows how  $c(M)$  can lead to incorrect results for mixtures of similarly abundant species with dissimilar frictional ratio. Better results are obtained when the peak regions of  $c(s)$  are substituted by discrete species, combined with continuous sections in between (using again the hybrid discrete/continuous model in SEDPHAT, Eq. 5), such that the hydrodynamic scaling law Eq. 4 and assumptions of shape similarity are abandoned for the species of interest (see above).

A third source of information from the sedimentation process specifically for the characterization of heterogeneous macromolecular mixtures is the different signal contributions to the interference optical and absorbance optical detection system, and/or to the absorbance at different wavelengths. This was demonstrated by Steinberg and Schachman for the simultaneous analysis of constituent sedimentation coefficients. (This has been implemented in SEDPHAT for the global analysis of isotherms of signal-average sedimentation coefficient obtained at different loading compositions and observed at different optical signals). It has been shown recently that characteristic distinguishable optical signals can be exploited to unravel the sedimentation coefficient distributions of the individual components in mixtures,  $c_k(s)$ <sup>26</sup>. Due to the large data basis in sedimentation velocity, differences in the number of aromatic amino acids in proteins can be sufficient for distinguishing their sedimentation for mixtures of two or three different proteins mixtures, if data are acquired by refractive index sensitive interference optics and absorbance optics at 280 and 250 nm<sup>26</sup>. This can be enhanced by chromophoric labeling of selected protein components. First applications have shown the potential for the combined spectral and hydrodynamic resolution to unravel the stoichiometry of multiple co-existing protein complexes, such as in the multi-step assembly of vitronectin and plasminogen activator inhibitor-1<sup>38</sup>, and in the assembly of the multi-protein complex of adaptor proteins after T-cell activation<sup>39</sup>.

In the present context, we have shown how the global multi-signal  $c_k(s)$  analysis can reveal the stoichiometry of the protein complex (Figure 6), extended to the combination of discrete species and continuous segments of the distribution (Eq. 8). Due to the large molar mass ratio, the spectral differences of complexes with different stoichiometry was relatively small, posing a very stringent test for the spectral decomposition. Nevertheless, the results clearly indicate the formation of 2:1 complexes, consistent with the previous results from the continuous  $c_k(s)$  analysis<sup>26</sup>. It should be noted that this does neither rely on the particular  $s$ -value of the complex, nor the boundary spreading, and therefore appears to be more robust against the onset of the influence of chemical reactions on the time-scale of sedimentation in reaction boundaries<sup>26,32</sup>.

It is a problem inherent to all known sedimentation coefficient distributions that they are based on equations describing the non-interacting superposition of species, which is strictly not suitable for the characterization of reversibly interacting macromolecules. Nevertheless, it can be a good approximation if the protein complexes are relatively stable on the time-scale of sedimentation, which could be either due to a low dissociation rate constant of the complex, or if the complex can be maintained in virtual saturation through use of concentrations far above the equilibrium dissociation constant. If these conditions are not fulfilled, the sedimentation is governed by the simultaneous migration and chemical conversion. Gilbert has described in detail the shapes of the resulting reaction boundaries for the case of instantaneous reactions in the absence of diffusion, and shown that they exhibit bimodal shape for two-component mixtures, with a leading boundary component that migrates

at a rate intermediate between that expected of a stable complex and stable free species<sup>31</sup>. Independently, Krauss et al. have shown that the reaction boundary for small reactants can be described approximately by a single diffusion coefficient<sup>5,30,40</sup>. Using computer simulations for the sedimentation of reacting systems with different complex stability and at different loading concentrations<sup>30,34</sup>, we have shown that whether or not  $c(s)$  peaks represent sedimenting species or features of the reaction boundary can be diagnosed from a series of experiments at different loading concentrations covering a range from significantly lower to significantly higher than  $K_D$ . If the peak positions of  $c(s)$  remain at a constant position but change in the ratio of peak area, a relatively slow interaction is indicated for which the species interpretation is a good approximation. On the other hand, dependence of the  $c(s)$  peak positions on the loading concentration indicates the presence of a fast reaction with  $c(s)$  reflecting the reaction boundary. However, it is also possible that the sedimentation data alone may not have sufficient information to permit the correct assessment of the interaction kinetics<sup>30</sup>.

In both cases, the integration of  $c(s)$  distribution over all peaks caused by species participating in the reaction will provide a correct measure of the weight-average  $s$ -value (or signal-average  $s$ -value, respectively). The isotherm of the dependence of the signal-average  $s$ -values on the loading concentration can be modeled to determine equilibrium constants<sup>34,41</sup>. This can be combined in a global model with isotherms utilizing the resolution of undisturbed and reaction boundary or the population of the different species, respectively, dependent on the time-scale of reaction kinetics.

As an alternative to the isotherm modeling approach following  $c(s)$  analysis for interacting systems, the raw sedimentation data may be modeled directly with Lamm equation solutions incorporating reaction terms<sup>42</sup>. Efficient numerical solutions of these Lamm equations have been recently introduced and implemented in SEDPHAT<sup>30</sup>. In this context, our results suggest that the application of the  $c(s)$  analysis to can be helpful initially to determine the correct interaction model, in particular, in the form of the multi-signal component sedimentation coefficient distributions  $c_k(s)$ . More details on the behavior of  $c(s)$  and multi-signal  $c_k(s)$  applied to reacting systems, and the comparison with Lamm equation modeling and isotherm analyses can be found in<sup>27,30,32,34</sup>.

## References:

1. Svedberg, T. & Pedersen, K. O. *The ultracentrifuge* (Oxford University Press, London, 1940).
2. Schachman, H. K. *Ultracentrifugation in Biochemistry* (Academic Press, New York, 1959).
3. Lamm, O., *Ark. Mat. Astr. Fys.*, 1929, **21B(2)**, 1-4.
4. Holladay, L. A., *Biophys. Chem.*, 1979, **10**, 187-190.
5. Urbanke, C., Ziegler, B. & Stieglitz, K., *Fresenius Z. Anal. Chem.*, 1980, **301**, 139-140.
6. Philo, J. S., *Biophys. J.*, 1997, **72**, 435-444.
7. Schuck, P., *Biophys. J.*, 1998, **75**, 1503-1512.
8. Schuck, P., MacPhee, C. E. & Howlett, G. J., *Biophys. J.*, 1998, **74**, 466-474.
9. Demeler, B., Behlke, J. & Ristau, O., *Methods in Enzymology*, 2000, **321**, 36-66.
10. Behlke, J. & Ristau, O., *Biophys Chem*, 2002, **95**, 59-68.
11. Stafford, W. F. & Sherwood, P. J., *Biophys Chem*, 2004, **108**, 231-243.
12. Schuck, P., *Biophys. J.*, 2000, **78**, 1606-1619.
13. Schuck, P., *Biophys. Chem.*, 2004, **108**, 187-200.
14. Schuck, P., *Biophys. Chem.*, 2004, **187**, 201-214.
15. MacRaild, C. A., Hatters, D. M., Lawrence, L. J. & Howlett, G. J., *Biophys J*, 2003, **84**, 2562-9.
16. Dam, J. & Schuck, P., *Methods Enzymol*, 2004, **384**, 185-212.
17. Schuck, P. & Demeler, B., *Biophys. J.*, 1999, **76**, 2288-2296.

17. Schuck, P. & Demeler, B., *Biophys. J.*, 1999, **76**, 2288-2296.
18. Provencher, S. W., *Comp. Phys. Comm.*, 1982, **27**, 213-227.
19. Hansen, P. C. *Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion* (SIAM, Philadelphia, 1998).
20. Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J. & Schubert, D., *Biophys J*, 2002, **82**, 1096-1111.
21. Schuck, P. & Rossmann, P., *Biopolymers*, 2000, **54**, 328-341.
22. Philo, J. S., *Anal. Biochem.*, 2000, **279**, 151-163.
23. Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry. II. Techniques for the study of biological structure and function* (W.H. Freeman, New York, 1980).
24. [www.analyticalultracentrifugation.com/references.htm](http://www.analyticalultracentrifugation.com/references.htm)
25. Boukari, H., Nossal, R., Sackett, D. L. & Schuck, P., *Physical Review Letters*, 2004, **93**, 098106.
26. Balbo, A., Minor, K. H., Velikovsky, C. A., Mariuzza, R., Peterson, C. B. & Schuck, P., *Proc Natl Acad Sci U S A*, 2005, **102**, 81-86.
27. Balbo, A. & Schuck, P. in *Protein-Protein Interactions* (eds. Golemis, E. & Adams, P. D.) (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, in press).
28. [www.nih.gov/od/ors/dbeps/PBR/AUC.htm](http://www.nih.gov/od/ors/dbeps/PBR/AUC.htm)
29. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C* (University Press, Cambridge, 1992).
30. Dam, J., Velikovsky, C. A., Mariuzza, R., Urbanke, C. & Schuck, P., *Biophys. J.* in press.
31. Gilbert, G. A. & Jenkins, R. C., *Nature*, 1956, **177**, 853-854.
32. Dam, J. & Schuck, P., submitted.
33. Steiner, R. F., *Arch. Biochem. Biophys.*, 1954, **49**, 400-416.
34. Schuck, P., *Anal. Biochem.*, 2003, **320**, 104-124.
35. van Holde, K. E. & Weischet, W. O., *Biopolymers*, 1978, **17**, 1387-1403.
36. Shire, S. J., Shahrokh, Z. & Liu, J., *J. Pharmaceutical Sciences*, 2004, **93**, 1390-1402.
37. Taraporewala, Z. F., Schuck, P., Ramig, R. F., Silvestri, L. & Patton, J. T., *J Virol*, 2002, **76**, 7082-93.
38. Minor, K. H., Schar, C. R., Blouse, G. E., Shore, J. D., Lawrence, D. A., Schuck, P. & Peterson, C. B., submitted.
39. Houtman, J. C., Yamaguchi, H., Barda-Saad, M., Bowden, B., Titterence, R., Appella, E., Schuck, P. & Samelson, L. E., submitted.
40. Krauss, G., Pingoud, A., Boehme, D., Riesner, D., Peters, F. & Maass, G., *Eur. J. Biochem.*, 1975, **55**, 517-529.
41. Steinberg, I. Z. & Schachman, H. K., *Biochemistry*, 1966, **5**, 3728-3747.
42. Cox, D. J. & Dale, R. S. in *Protein-protein interactions*. (eds. Frieden, C. & Nichol, L. W.) (Wiley, New York, 1981).