Size-Distribution Analysis of Proteins by Analytical Ultracentrifugation: Strategies and Application to Model Systems

Peter Schuck[#][*], Mattthew A. Perugini[%], Noreen R. Gonzales[&], Geoffrey J. Howlett[%], Dieter Schubert[$]

[#]Division of Bioengineering and Physical Science, ORS, National Institutes of Health, Bethesda, Maryland, USA, [&]Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, Maryland, USA, [%]Department of Biochemistry and Molecular Biology, The University of Melbourne, Parkville, Australia, [$]Institut für Biophysik, Johann Wolfgang Goethe-Universität, Frankfurt am Main, Germany

Running Title: Size-Distribution Analysis of Proteins

keywords: analytical ultracentrifugation, sedimentation velocity, size-distribution, regularization, maximum entropy

*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

Abstract

Strategies for the deconvolution of diffusion in the determination of size-distributions from sedimentation velocity experiments were examined and developed. Based on four different model systems, we studied the differential apparent sedimentation coefficient distributions by the time-derivative method, g(s*), and by least-squares direct boundary modeling, *ls-g*(s)*, the integral sedimentation coefficient distribution by the van Holde-Weischet method, G(s), and the previously introduced differential distribution of Lamm equation solutions, c(s). It is shown that the least-squares approach *ls-g*(s)* can be extrapolated to infinite time by considering area divisions analogous to boundary divisions in the van Holde-Weischet method, thus allowing the transformation of interference optical data into an integral sedimentation coefficient distribution G(s). However, despite the model-free approach of G(s), for the systems considered, the direct boundary modeling with a distribution of Lamm equation solutions c(s) exhibited the highest resolution and sensitivity. The c(s) approach requires an estimate for the size-dependent diffusion coefficients D(s), which is usually incorporated in the form of a weight-average frictional ratio of all species, or in form of prior knowledge of the molar mass of the main species. We studied the influence of the weight-average frictional ratio on the quality of the fit, and found that is well-determined by the data. As a direct boundary model, the calculated c(s) distribution can be combined with a nonlinear regression to optimize distribution parameters, such as the meniscus position, and the weight-average frictional ratio. Although c(s) is computationally the most complex, it has the potential for the highest resolution and sensitivity of the methods described.

Introduction

Analyzing the size-distribution of biological or synthetic macromolecules in solution, for

example, for the study of the oligomeric state of proteins, is a very important application of

analytical ultracentrifugation with a long history (Signer and Gross, 1934; Svedberg and

Pedersen, 1940; Bridgman, 1942; Baldwin and Williams, 1950; Vinograd and Bruner, 1966;

Scholte, 1968; van Holde and Weischet, 1978; Stafford, 1992a). Because of the relatively large

size-dependence of macromolecular migration in a gravitational field, sedimentation velocity

studies have the potential for high resolution. In the last decades, two approaches for the

determination of sedimentation coefficient distributions have become most popular: the integral

sedimentation coefficient distributions $G(s)$ by van Holde and Weischet (in the following

abbreviated as vHW) (van Holde and Weischet, 1978) and the differential apparent

sedimentation coefficient distribution $g*(s)$ obtained as a transform of the time-derivative of the

signal, dc/dt (Stafford, 1992a). Exploiting the increased computational capabilities now

available, we have recently proposed new methods for obtaining the apparent differential

sedimentation coefficient distribution $g*(s)$, termed ls-$g*(s)$ (Schuck and Rossmanith, 2000), and

for calculating a differential sedimentation coefficient distribution $c(s)$ in which corrections for

diffusion are made (Schuck, 2000). Both methods are based on direct least-squares modeling of

the sedimentation boundary, using linear combinations of sedimentation profiles for non-

diffusing species or linear combinations of Lamm equation solutions, respectively. They can be

applied to larger data sets and, by virtue of regularization, exhibit substantially less noise in the

calculated distributions than previous methods. First applications of the methods indicate a high

versatility and advantages in sensitivity and resolution over the classical methods (Perugini et al.,

; Schuck et al., 2000; Schuck and Rossmanith, 2000; Cole and Garsky, 2001; Sedlák and Cölfen,

2001; Hatters et al., 2001). However, a systematic exploration of differences and relationships among the different approaches, the statistical and experimental prior knowledge needed, as well as the implicit assumptions and their practical relevance is not available at present.

One of the classical problems in the theory of ultracentrifugal sedimentation is the treatment of diffusion in size-distribution analysis. The deconvolution of boundary diffusion can significantly increase the amount of detail that can be learned from a sedimentation experiment, and how diffusion is described constitutes a central difference between the existing approaches. One difficulty is that a set of two parameters (such as sedimentation and diffusion coefficient) is required to describe the sedimentation of each species in the distribution. While the apparent sedimentation coefficient distributions $g(s^*)$ and $ls\text{-}g^*(s)$ do not allow consideration of diffusion, the vHW method for calculating G(s) is designed to take diffusion into account in a model-independent way by extrapolation of boundary fractions to infinite time. In the present paper, we describe a similar model-free method for calculating G(s) by extrapolation of the $ls\text{-}g^*(s)$ distributions to infinite time. Although complete model-independence can be appealing, the extrapolation process can limit the resolution. Further, we will show that this extrapolation strategy for the deconvolution of diffusion fails for heterogeneous mixtures of species with overlapping sedimentation boundaries. The Lamm equation method $c(s)$ employs an intermediate strategy, by estimating size-dependent diffusion coefficient via the Stokes-Einstein and Svedberg relationships and by utilizing prior knowledge, such as the weight-average frictional ratio, the molar mass of a main component, or a similar information relating size and shape of the distribution. It also employs maximum entropy regularization, a Bayesian strategy to achieve numerical stability and optimal resolution, which is routinely used in many other fields of physics and biophysics for problems of similar mathematical structure. In contrast to

the more classical data transformations, direct boundary models such as c(s) provide a criteria for

the goodness-of-fit, which has potential use in nonlinear regression of distribution parameters.

So far, however, this has remained unexplored.

In the present communication, we apply the different methods (*g(s*)*, *ls-g*(s)*, *G(s)* by

vHW and by extrapolation of *ls-g*(s)*, and *c(s)*) to four different data sets with different

broadness of the size-distribution and different extent of diffusion. First, we examine a

previously proposed theoretical model system with four species of closely spaced sedimentation

coefficients (Stafford, 1992b). We then compare the information obtained from the analysis of

experimental data from a predominantly single species with a trace impurity, a self-associating

protein with several discrete oligomeric states, and a continuous distribution of large lipid

emulsion particles. Besides questions of sensitivity and of resolution of species that do not

exhibit clearly distinguishable sedimentation boundaries, we also examine the stability of the *c(s)*

approach with respect to the prior assumption needed. In particular, we show how the

knowledge of the weight-average frictional ratio can be extracted from the experimental data by

combination with nonlinear regression.

Experimental

Analytical ultracentrifugation

For sedimentation velocity experiments, a Beckman Optima XL-I analytical

ultracentrifuge with absorbance and interference optical detection system was used. Epon

double-sector centerpieces were filled with 400 μl of sample solution and PBS, respectively, and

centrifuged at a rotor speed of 40,000 rpm or 55,000 rpm and at rotor temperatures of 5 °C or 20

°C, respectively. Absorbance data were acquired at a wavelength of 280 nm or 230 nm,

respectively, and in time intervals of 2 min, with the radial increment set to 0.002 cm and taking

2 averages per scan; interference scans were taken in time intervals of 1 min. Buffer viscosity,

protein partial specific volumes and frictional ratios were calculated using the software

SEDNTERP (Laue et al., 1992).

Sedimentation equilibrium studies were conducted in a Beckman Optima XL-A equipped

with absorbance optics. Double-sector or six-channel charcoal-filled epon centerpieces were

filled with 140 µl of sample at loading concentration between 0.1 and 0.6 mg/ml, respectively.

Sedimentation equilibrium was attained at a rotor temperature of 4 °C at rotor speeds of 10,000

rpm and 13,000 rpm, respectively, and absorbance profiles were acquired at wavelengths of 230

nm and 280 nm. Extinction coefficient ratios at different wavelengths were estimated

spectrophotometrically.

Data analysis

Lamm equation modeling

Sedimentation velocity data analysis was performed with the program SEDFIT (which can

be obtained from http://www.AnalyticalUltracentrifugation.com). For direct boundary modeling

with distributions of Lamm equation solutions (Schuck, 2000), the measured absorbance or

interference profiles $a(r,t)$ were modeled as an integral over the differential concentration

distribution $c(s)$

$$a(r,t) = \int c(s)\chi(s, D(s), r, t)ds + \varepsilon \qquad \text{(Eq. 1)}$$

with ε denoting noise components, and $\chi(s,D,r,t)$ denoting the solution of the Lamm equation for a single species (Lamm, 1929)

$$\frac{d\chi}{dt} = \frac{1}{r}\frac{d}{dr}\left[rD(s)\frac{d\chi}{dr} - s\omega^2 r^2\chi\right]$$ (Eq. 2)

(where r denotes the distance from the center of rotation, and ω the rotor angular velocity) which was solved by finite element methods on a static or moving frame of reference as described in (Claverie et al., 1975; Schuck, 1998; Schuck et al., 1998). For each species, the diffusion coefficient $D(s)$ was estimated as a function of the sedimentation coefficient s based on the known partial specific volume of the protein, and on an estimated anhydrous frictional ratio f/f_0 (Laue et al., 1992; Schuck, 2000). In some cases, this can be combined with a predetermined interval of s-values for which diffusion coefficients $D(s)$ are calculated from the Svedberg equation, using prior knowledge of the buoyant molar mass (Schuck et al., 2000).

The integral equation 1 was solved numerically by discretization into a grid of 100 – 200 sedimentation coefficients and calculating the best-fit concentrations for each species in a linear least squares fit. Systematic noise components of the data were estimated by using an algebraic method (Schuck and Demeler, 1999) (see below). Numerical stability was achieved by the maximum entropy method (Amato and Hughes, 1991),

$$\underset{c(s)}{\text{Min}}\ \left\{\sum_{i,j}\left[a(r_i,t_j) - \int c(s)\chi(s,r_i,t_j)ds\right]^2 - \alpha\int c(s)\ln c(s)ds\right\}$$ (Eq. 3)

which is a regularization procedure that minimizes not only the deviation between model and data at each radius value r_i and each time t_j, but simultaneously maximizes the information entropy $\int c\ln c\,ds$. The maximum entropy constraint α is adjusted such that the increase in the χ^2 of the constrained fit as compared to the unconstrained fit ($\alpha = 0$) corresponds to a confidence

level of one or two standard deviations (P = 0.68 or 0.95, respectively) as calculated by F-statistics (Provencher, 1982a; Schuck, 2000; Johnson and Straume, 1994; Bevington and Robinson, 1992). This method has the virtue of providing the simplest distribution of all possible distributions that are consistent with the data, while allowing for sharp peaks in the distribution if necessary for modeling the data. Alternatively, when prior knowledge is available about the smoothness of the distribution, the maximum entropy constraint can be replaced by a Tikhonov-Phillips term $\int (d^2 c / ds^2)^2 \, ds$ that minimizes the second derivative of the distribution. Because of its linearity, this term can be implemented in the computationally simpler matrix form. However, it cannot describe isolated peaks as well as the maximum entropy constraint, and tends to produce smoother distributions, which can make it more robust.

For enabling the critical inspection of the quality of fits to sedimentation velocity data, we have developed a two-dimensional picture representation of the residuals, in order to avoid the usual loss of information on systematic deviations in the common overlay presentations. The bitmap representation of the residuals was calculated in the following way: The residual values of all points in all scans $R(r,t)$ were transformed to a gray value $n(r,t)$ between 0 and 255, with n = 0 for $R(r,t) <= -0.05$, $n = 255$ for $R(r,t) >= 0.05$, and with a linear transformation for the residuals $-0.05 < R(r,t) < 0.05$. This transformation results in neutral gray ($n = 128$) for a perfect fit with $R(r,t) = 0$, brighter pixels for positive and dark pixels for negative residuals. In the bitmap, the pixels were ordered in rows that correspond to the scan number, and columns that correspond to the radius values. This representation of the residuals results in a uniformly gray picture without any structure for a good fit with randomly distributed residuals. If a structure is visible, this corresponds to systematic residuals. In this way, systematic residuals from vibrations of the camera, which produce vertical patterns, can be distinguished from those of an

imperfect boundary model resulting in diagonal structures. Also, the presence of isolated bad

scans can be diagnosed from horizontal lines, and they can be identified from a separate output

file of SEDFIT of the local rms error for each file.

Systematic noise analysis

     Components of systematic time-invariant and radial-invariant noise were calculated using

the algebraic approach described previously (Schuck and Demeler, 1999). In brief, the time-

invariant baseline signal b_i at each radius r_i was minimized by least-squares according to

$$\underset{\{p\},b_i}{\text{Min}} \sum_{i,j} \left[a(r_i,t_j) - \left(b_i + S(\{p\},r_i,t_j)\right)\right]^2 \qquad \text{(Eq. 4)}$$

In this equation, S denotes any model for the sedimentation boundary, such as the solution of the

Lamm equation (Eq. 2), or the sedimentation of a size-distribution (Eq. 3), which might in

general depend on a set of parameters $\{p\}$. It can be easily shown that for any given value of $\{p\}$

(or if S has no further parameters), the best- fit time-invariant noise is given by

$$b_i(\{p\}) = \bar{a}_i - \bar{S}_i(\{p\}) \qquad \text{(Eq. 5)}$$

where the quantities

$$\bar{a}_i = (1/N_s)\sum_j a(r_i,t_j), \quad \bar{S}_i(\{p\}) = (1/N_s)\sum_j S(\{p\},r_i,t_j) \qquad \text{(Eq. 6)}$$

(with the total number of scans N_s) represent an average scan, and an average sedimentation

model, respectively. This leads to a least-squares problem for the calculation of the remaining

parameters $\{p\}$

$$\underset{\{p\}}{\text{Min}} \sum_{i,j} \left[\left(a(r_i,t_j) - \bar{a}_i\right) - \left(S(\{p\},r_i,t_j) - \bar{S}_i(\{p\})\right)\right]^2 \qquad \text{(Eq. 7)}$$

An analogous procedure can be used for the radial-invariant signal-offsets (Schuck and Demeler, 1999). For the distribution analysis it is straightforward to solve the problem Eq. 7 directly by linear least squares methods (Schuck, 2000; Schuck and Rossmanith, 2000).

Inspection of Eq. 7 shows that the only information that can be extracted from the data is that of a time-difference (here with an 'average scan' as a reference). It should be noted that no explicit estimate of the time-invariant noise is used in Eq. 7. Nevertheless, for fundamental reasons, modeling the time-difference introduces new degrees of freedom into the data analysis, which is a consequence of the unknown radial-dependent baseline offsets. This can lead to slight correlation with parameters of the boundary model, in particular with those describing very slow sedimentation processes (Schuck and Demeler, 1999; Kar et al, 2000). Such correlation can be minimized by using a large data set that includes large boundary displacement (Kar et al, 2000). The calculation of an explicit estimate of the baseline parameters with Eq. 5 follows after the nonlinear regression in Eq. 7 and allows comparing the sedimentation model with the data in the original data space (direct boundary modeling). It follows from Eq. 5 that the best estimate of the time-invariant signal is an average over all scans of the residuals profiles. Therefore, it is dependent on the parameters of the sedimentation model, and realistic estimates of the time-invariant baseline signal are obtained only if the sedimentation model fits the data well. However, this step does not introduce any additional correlation in the estimates of the sedimentation parameters {p}.

Calculation of the sedimentation coefficient distributions g*(s) and G(s)

For obtaining the apparent sedimentation coefficient distribution g*(s), the direct boundary model for a distribution of non-diffusing particles *ls-g*(s)* (Schuck and Rossmanith,

2000) was used, as implemented in the software SEDFIT. In this method, *ls-g*(s)* is calculated

using the same concepts and numerical framework as the distribution *c(s)* described above, but

by replacing the Lamm equation solution **c***(s,D(s),r,t)* in Eq. 1 with the theoretical sedimentation

profiles of non-diffusing species, i.e. step-functions *U(s,r,t)*

$$a(r,t) \cong \int g*(s)U(s,r,t)ds \qquad \text{(Eq. 8)}$$

$$U(s,r,t) = e^{-2\omega^2 st} \times \begin{cases} 0 & for \quad r < r^*(t) = r_m e^{\omega^2 st} \\ 1 & else \end{cases} \qquad \text{(Eq. 9)}$$

U(s,r,t) describes the ideal behavior of initially uniformly distributed particles with

sedimentation coefficient *s*, but without any diffusion, during sedimentation and radial dilution in

the sector-shaped solution column in the centrifugal field (with the meniscus position of the

solution column r_m, and a boundary position *r*(t)*) (Fujita, 1962; Stafford, 1992a; Schuck and

Rossmanith, 2000). Regularization was applied using the Tikhonov-Phillips method, at a

confidence level of P = 0.95, unless stated otherwise.

Eq. 8 can be combined with Eqs. 7 and 5 for systematic noise analysis. As described

above, as a consequence of modeling the time difference in Eq. 7, some additional correlation

and increased error of the distribution at small s-values can occur. Both can be minimized best

in the *ls-g*(s)* analysis by using data sets with large boundary displacement and scans where the

boundary has cleared the meniscus. It should be noted that if used for the analysis of small

molecules where diffusion is not negligible, Eq. 9 is not a good approximation, and dependent on

the size of the data set, large residuals and relatively poor estimates of the time-invariant signal

may be obtained.

The differential sedimentation coefficient distribution defined by Eqs. 8 and 9 is termed

ls-g*(s) to indicate its basis on the least-squares data modeling ('ls'), and the neglect of diffusion

('g*'). A similar differential sedimentation coefficient distribution can be calculated using the time-derivative dc/dt, as approximated by the time difference Dc/Dt (Philo, 2000; Stafford, 1992a). This is termed g(s*) in order to indicate the use of a transformation of the radial variable r into an apparent sedimentation coefficient s* in the course of its calculation. In this method, the pairwise time-difference between scans is used to eliminate systematic time-invariant noise. The final apparent sedimentation coefficient distribution g(s*) from dc/dt can be transformed back into a boundary model (with rhs of Eq. 8 and 9) for comparison of model and data, and explicit estimates of the time-invariant noise can be calculated via Eq. 5 (data not shown). Because both methods *g(s*)* and *ls-g*(s)* are based on equivalent definitions of the apparent sedimentation coefficient distribution, when applied to the same data sets, this leads to equivalent results for both the g*(s) distribution (Schuck and Rossmanith, 2000) and the time-invariant noise estimates (data not shown). However, the absence of a differentiation step in *ls-g*(s)* allows larger boundary displacements between the scans, and avoids artificial broadening effects that can be introduced by the approximation of *dc/dt* by Dc/Dt (Schuck and Rossmanith, 2000; Philo, 2000). For comparison, where possible, g(s*) analysis of time-difference sedimentation data was performed with the program DCDT+ (J.S. Philo, 3329 Heatherglow Ct., Thousand Oaks, CA 91360).

The integral sedimentation coefficient distribution *G(s)* was calculated as described by van Holde and Weischet (van Holde and Weischet, 1978). This method is based on the Faxén-type approximate solution of the Lamm equation, which can be written as

$$c(r,t) = \left(\frac{c_0 e^{-2sw^2 t}}{2} \right) \left[1 - \Phi\left(\frac{r_m \ln r*(t) - r_m \ln r}{\sqrt{2Dt}} \right) \right] \qquad \text{(Eq. 10)}$$

with the boundary position of a non-diffusing species $r^*(t)$ as defined in Eq. 9, and the error

function Φ (van Holde and Weischet, 1978; Demeler et al., 1997). Following the derivation of

van Holde and Weischet (van Holde and Weischet, 1978), the radial positions R_i of fractional

plateau concentrations c_i ($c_i = ic_p/N$, with i denoting the fraction number, and N denoting the

total number of divisions of the plateau concentration c_p) can be transformed in apparent

sedimentation coefficients $s^*_{app,i} = \ln(R_i/r_m)/\omega^2 t$, so that

$$\frac{2i}{N} = 1 - \Phi\left((s - s^*_{app,i}) \frac{w^2 r_m}{2\sqrt{D}} \sqrt{t} \right) \qquad \text{(Eq. 11)}$$

With the inverse error function Φ^{-1} applied to both sides of Eq. 11, we arrive at

$$s^*_{app,i} = s - \frac{2\sqrt{D}}{w^2 r_m} \Phi^{-1}\left(1 - \frac{2i}{N} \right) \times \frac{1}{\sqrt{t}} \qquad \text{(Eq. 12)}$$

(van Holde and Weischet, 1978), which shows that a linear extrapolation of $s^*_{app,i}$ on a $t^{-0.5}$ scale

to infinite time allows determining s, and effectively deconvoluting of diffusion effects on the

sedimentation boundary. The resulting s-values from the different boundary fractions form the

integral sedimentation coefficient distribution $G(s)$ (van Holde and Weischet, 1978; Demeler et

al., 1997).

We have used the implementation of the vHW approach outlined earlier (Schuck, 2000).

In brief, after determination of the plateau signals for each curve, the boundary was divided in N

(usually 20 to 50) fractions of equal concentration increments dh. The radial position of the

boundary fraction is calculated as R_i = mean {r, with $dh´(i-0.5)< c(r) < dh´(i+0.5)$}, i.e. as the

average of the radial values of all data points that have signal values as defined by the limits of

the boundary fraction. This method is designed for a high number of fractions, where the

boundary increment dh for each fraction are comparable in size to the noise of the data, and it

extracts the boundary positions in a least-square sense, not requiring smoothing of the data. In the algorithm implemented in SEDFIT, it is ensured that all boundary fractions in all scans have at least one data point, otherwise the number of boundary fractions N is automatically reduced.

As an alternative strategy for the calculation of $G(s)$, we have implemented the following extrapolation of ls-$g*(s)$ to infinite time: The total set of scans used for analysis was subdivided in sequential sets of scans, each taken at a time interval centered at t_i. (For example, sets of 10 scans were used for the analysis of interference optical data.) For each set, a differential sedimentation coefficient distribution ls-$g*(s)_i$ was calculated and divided into N equal area fractions A_j. Because the area under the ls-$g*(s)$ curves corresponds to the loading concentration (Eq 8), these fractional areas are equivalent to boundary fractions, and the average sedimentation coefficient $s_{ij}(A_j)$ in a given area fraction A_j at time t_i directly corresponds to the s-values $s*_{app,i}$ calculated for each boundary fractions in the van-Holde Weischet method. As a consequence, the same extrapolation procedure Eq. 12 can be applied to generate a distribution $G(s)$. (It should be noted that this method, like vHW, requires the existence of solution plateaus in order to define consistent area fractions, and that it requires some depletion at the meniscus to avoid correlation of ls-$g*(s)$ with baseline and systematic noise parameters.) Each of the ls-$g*(s)_i$ curves can be calculated taking into account time-invariant and radial-invariant systematic noise (Eq. 7), but best results were obtained if only time-invariant noise was considered (vertical alignment of the scans, e.g. close to the meniscus, may be achieved separately). After the linear regression with Eq. 12, the best-fit values of $s_{ij}(A_j)$ can be transformed back into equivalent boundary positions, and a step-function model of the boundary in the original data space can be generated (Eq. 9). This allows calculating overall best-fit estimates for the systematic noise contributions via Eq. 5 (see above).

There is a subtle difference between the two procedures. The division of the boundary into equal fractions of the plateau signal introduces small errors in the linear approximation of $s_{app,i}$ versus $t^{-0.5}$. However, by substituting Eq. 10 with an improved approximate solution of the Lamm equation, a more complex relationship $s_{app,i}$ versus $t^{-0.5}$ can be derived (Eq. 17 of (van Holde and Weischet, 1978)). In the division of the $ls\text{-}g^*(s)$ area, the division is made in units of equivalent loading concentrations, and the fractions are propagated in time according to Eq. 9, such that they generate boundary fractions that have experienced different radial dilutions. This can be expected to slightly alter the precision of the linear approximation. However, in studies with synthetic data, we found this to produce similar accuracy of the linear extrapolation $s_{app,i}$ versus $t^{-0.5}$ (both have errors < 1%, data not shown). An independent theoretical justification of the $ls\text{-}g^*(s)$ approach can be derived from the observation that the apparent sedimentation coefficient distributions exhibit a sharpness increasing with time, and by interpreting Eq. 10 as describing diffusional spread in a space of 'apparent sedimentation coefficients'. In this view, the division of $ls\text{-}g^*(s)$ in area fractions and linear extrapolation on a $t^{-0.5}$ scale can be understood solely as a rational method for extrapolating $ls\text{-}g^*(s)$ to infinite time (with the transformation to the integral $G(s)$ distribution providing increased numerical stability of the extrapolation). However, because of the close relationship of the methods, we interpret the extrapolation of area fractions of the $ls\text{-}g^*(s)$ distributions as an extension of the vHW method, in a sense that $ls\text{-}g^*(s)$ can be calculated easily in the presence of systematic time-invariant as well as radial-invariant noise of interference optical data by applying algebraic noise decomposition (Schuck and Demeler, 1999).

Sedimentation equilibrium analysis

Sedimentation equilibrium data were analyzed by global modeling of 3 to 6 data sets

obtained at different loading concentrations and rotor speeds using the commercial mathematical

modeling software MLAB (Civilized Software, Silver Spring, MD). Least-squares fits of the

measured absorbance profiles *a(r)* were calculated using models based on the exponential

equilibrium distribution of ideally sedimenting oligomeric species (Svedberg and Pedersen,

1940)

$$a(r) = \sum_{\{i\}} c_i(r_0) i \varepsilon_\lambda d \exp\left[iM(1-\bar{v}\rho) \frac{\omega^2(r^2 - r_0^2)}{2RT} \right] \qquad \text{(Eq. 13)}$$

where r_0 is an arbitrary reference radius, and with the monomer molar mass and partial specific

volume M and $\bar{v}$, respectively, the solvent density ρ, the absolute temperature T, the gas

constant R, the molar extinction coefficient ε_λ at wavelength λ, and the thickness of the

centerpiece d (1.2 cm). Dependent on the particular model used, the molar concentrations of the

i-mer $c_i(r_0)$ were coupled by mass action law. In all models, the absence of partial specific

volume changes upon oligomerization was assumed.

A transformation of the absorbance distribution (of a single scan) into a continuous molar

mass distribution *c(M)* combined with maximum entropy regularization was achieved by

replacing the kernel in Eq. 3 by sedimentation equilibrium exponentials (Eq. 13) for a single

species. In order to allow for a rational comparison of the concentration of the different species,

average loading concentrations were used as concentration units, obtained by integration of each

species from meniscus to bottom. This method is similar to the Laplace-transform with

regularization described by (Wiff and Gehatia, 1976). Because of the high sensitivity of the

shape of *c(M)* on the location of the bottom of the solution column, for this analysis, the

meniscus and bottom were predetermined using intensity scans.

Dynamic light scattering

Dynamic light scattering experiments were conducted using a Protein Solutions DynaPro

99 instrument with a DynaPro-MSTC200 microsampler (Protein Solutions, Charlottesville, VA).

Protein samples were centrifuged for 5 min in a microcentrifuge to remove dust particles, and a

20 µl sample was inserted in the cuvette with the temperature control set to 20 °C. The light

scattering signal was collected at 90°, and autocorrelation coefficients were exported for analysis

with the software SEDFIT, adapted for dynamic light scattering analysis by replacing the Lamm

equation solutions with the following models for the field autocorrelation function:

$$g^{(1)}(\tau) = \exp[-Dq^2\tau] \qquad\qquad \text{(Eq. 14)}$$

where τ is the decay time and $q=(4\pi n/\lambda)\sin(\Theta/2)$, with the solvent refractive index *n*, the

wavelength of the incident light λ, and the scattering angle Θ (Murphy, 1997). Continuous size-

distributions were calculated from the autocorrelation data analogous to the analysis of the

sedimentation coefficient distribution, but using the correlation functions Eq. 14 as kernel in the

integral equation Eq. 3. This results in a distribution analysis that is similar to the maximum

entropy method described by Livesey et al. (Livesey et al., 1986) and similar to that implemented

by S. Provencher in CONTIN (Provencher, 1982b; Provencher, 1979) (which employs Tikhonov-

Phillips regularization). The regularization was adjusted to a confidence level between 0.45 and

0.55 (Provencher, 1992).

Results

Comparison of the resolution using synthetic data

To examine the potential of Lamm equation modeling for resolving species in

sedimentation velocity experiments when no clear sedimentation boundary is visible, we have

simulated data from the model system that was proposed earlier by Stafford (Stafford, 1992b)

(Figure 1). It consists of four species with sedimentation coefficients 6, 7, 8, and 9 S. For each,

theoretical sedimentation profiles were calculated with a starting concentration of 0.25 (arbitrary

signal units), and to the sum of their signals normally distributed noise at a magnitude of 0.01

was added. From inspection of the resulting broad sedimentation profiles the presence of a

heterogeneous mixture is obvious, but no distinct boundaries can be identified (Figure 1A).

The distributions obtained greatly differ for the different methods (Figure 1B): As can be

expected, the apparent sedimentation coefficient distributions (both *ls-g*(s)* and *g(s*)*) only

reveal the range of s-values, without finer structure in the size-distribution. The results obtained

with the vHW method for the integral sedimentation coefficient distribution *G(s)* (van Holde and

Weischet, 1978), as well as for *G(s)* by extrapolation of *ls-g*(s)* to infinite time are only slightly

better as they represent the range of s-values more accurately. Although they can, in principle,

unravel the effects of diffusion, the close spacing of the sedimentation coefficients clearly

exceeds the resolution. However, the four species can be discriminated with the c(s) method, if

combined with the optimization of the weight-average frictional ratio (see below).

The deconvolution of diffusion by G(s) and c(s) merits more detailed consideration. The

extrapolation of the boundary fractions of both G(s) methods are shown in Figure 2. In

comparison, the extrapolated *ls-g*(s)* distribution has fewer time points for the extrapolation

(due to the need for a whole set of curves for calculating a single *ls-g*(s)* distribution), but it can

be subdivided into more boundary fractions (Figure 2B). However, as expected, the results are very similar in both methods. Interestingly, although the linear regression of the boundary fractions appears to be of high quality, it is clear that they do not contain the information required for resolving the species underlying the model system.

The application of the distribution of Lamm equation solutions $c(s)$ is based on one piece of additional information that can link s and D, which can be obtained most conveniently by estimating a frictional ratio f/f_0 of the species under study. Because in most cases the data will not contain enough information for defining f/f_0 as a function of sedimentation coefficient (or even a distribution of f/f_0 for all species with the same sedimentation coefficient), the c(s) method is restricted to use only a single value, corresponding to a weight-average frictional ratio of all species. An estimate may initially be based, for example, on the expected hydrodynamic behavior for the type of macromolecule under study (e.g. for globular protein, or random coil). In the application to the data shown in Figure 1A, we started with an initial guess of f/f_0 of 1.2, which resulted in two peaks at 6.5 and 8.5 S, but with a poor fit with an rms deviation of 0.015, and clearly systematic residuals (data not shown). The model functions calculated from c(s) were too broad, particularly in the earlier scans, indicating too large diffusion coefficients predicted by a too small prior estimate of $f/f_0 = 1.2$. Therefore, we increased f/f_0 to a value of 2 (resulting in the distribution with three peaks shown as dotted lines in Figure 1B), and then floated this parameter to be optimized in a nonlinear regression. This resulted in a $c(s)$ distribution that exhibits peaks at the correct position of the species underlying the simulated data. Furthermore, the parameter f/f_0 converged to a value of 2.9, which is close to the weight-average frictional ratio of 2.7 for the simulated, highly elongated, species. This indicates that nonlinear regression of f/f_0 can be a useful technique to obtain good estimates for this parameter.

Nonlinear regression was also found to be a good technique for determining the exact meniscus position, as well as the sedimentation and diffusion coefficients of co-sedimenting small molecules (such as buffer components that are not matched in the reference and sample side of the ultracentrifugal cell; data not shown).

When inspecting the details of the resulting best-fit c(s) distribution, it should be considered that the maximum entropy regularization causes the peaks not to be very sharp, as one could expect for an ideal measurement. By design of the Bayesian procedure, it restricts the structure in the $c(s)$ distribution to the features that are essential for describing the raw data within the limits of the noise. According to this result, any sharper distributions would not lead to significantly better fits (at a confidence limit of 0.68). Therefore, the $c(s)$ distribution in Figure 1B depicts the information that can be extracted reliably from the data in Figure 1A, and by design it does not represent the best-fit, which would in most cases be too unstable. However, it should be noted that even in the presence of noise, all four species can be unraveled with the $c(s)$ analysis even without their boundary separation. Interestingly, the four species cannot be resolved if only the data subsets suitable for vHW analysis are taken into consideration. We have shown earlier that for broad continuous distributions, the regularization can produce artificial oscillations (Schuck, 2000). The analysis here also highlights another well-known property of maximum entropy regularization, an inherent tendency to merge closely spaced peaks, in particular in non-optimal fits (where the predefined F-ratio results in higher fractional increase in χ^2, and consequently higher regularization). This is shown in the dotted line in Figure 1B. In order to achieve optimal resolution, therefore, it appears important to balance the distribution analysis (the confidence level and prior estimate of f/f_0) with the inspection of the

quality of the fit. As indicated above, this can include optimization of f/f_0 by least-squares

regression.

Study on the sensitivity of the methods using experimental data from an immunoglobulin G

sample

Next, we compared the performance of the methods using data from a nearly

homogeneous immunoglobulin G (IgG) sample. The experimental fringe shift data are shown in

Figure 3A. A fit with discrete solutions of the Lamm equation reveals one main species with

6.60 S, but with a (statistically highly significant) 3% contamination of a dimeric species with

9.58 ($\pm$ 0.2) S. Similarly, a fit of dc/dt with the Fujita-MacCosham-Philo function (Philo, 1997;

Philo, 2000) converges at a value of 5% of a faster sedimenting species (data not shown). This

may serve as a test for the sensitivity in the detection of trace amounts of species.

Figure 4 shows differential sedimentation coefficient distributions that are uncorrected

for the effects of diffusion, which are g(s*) by dc/dt (solid line), and the *ls-g*(s)* distribution

(circles). If applied to the same data subset, both distributions are very similar, and both clearly

reveal the presence of the larger species.   Figure 3B and C presents the integral sedimentation

coefficient distributions by the vHW analysis, by extrapolation of *ls-g*(s)* to infinite time (circles

in Figure 3B and 3C, calculated systematic noise shown as dotted line in Figure 3A) and by

conventional analysis after subtraction of systematic noise components (taken, for comparison

only, from the best-fit with the discrete Lamm equation modeling) (open squares in Figure 3B).

It can be seen that the s-value of the main species is at ~ 6.7 S and that the highest fraction

indicates the presence of a faster component. However, quantitation is difficult because of the

relatively large noise present in the highest boundary (or area) fractions. Finally, the differential

sedimentation coefficient distribution c(s) with diffusion deconvoluted assuming an average

frictional ratio of 1.58 is shown as a dotted line in Figure 4.

As in the first example, the increase in resolution in the c(s) method is achieved in part

because of the larger number of files that can be included in the analysis, but mainly through

estimates of the extent of diffusion, which will be examined in the following. Figure 5 shows the

dependence of the quality of fit obtained at different values for the frictional ratio. It can be seen

that the rms error has a clearly defined minimum. With non-optimal values, the boundary shape

is not well described, as illustrated by the diagonal pattern in the residual bitmaps. It should be

noted that this occurs to a similar extent both at too low and too high values of f/f_0, and that the

assumption of the average shape to be spherical ($f/f_0 = 1$) is equally poor as the limit of non-

diffusing particles. (This limit of no diffusion is identical to the *ls-g*(s)* and g(s*) distribution).

In contrast, the residual bitmap shows very little systematic patterns at the optimal value of 1.58.

As a consequence, like in the first example, we can extract an average frictional ratio from the

data itself by virtue of the criterion of the quality of fit. How the different values of f/f_0 affect

the calculated c(s) distribution is shown in Figure 6. Consistent with previous observations

(Schuck, 2000), the position of the main peak remains essentially constant, while smaller values

of f/f_0 lead to sharper peaks. However, the location of the peak of the trace component was

found to be correlated with f/f_0. If the diffusion is over-corrected, information on the

contaminating faster sedimenting species is lost, and the smaller peak appears reduced in area

and at higher s-values (inset in Figure 6). On the other hand, this is accompanied by a sharp

decrease in the quality of the fit (Figure 5), which helps in determining the s-value of the faster

species. In the inset of Figure 6, the two solid lines indicate two distributions that are

indistinguishable on a confidence level of 0.9, suggesting an error estimate in the order of 0.3 S (best fit at 9.38 S).

Application to the study of preparations of the herpes simplex capsid protein VP5

A more complex problem is the analysis of a protein with extended, slow oligomerization. This is illustrated by experiments with preparations of the herpes simplex capsid protein VP5 (the biological implications will be discussed elsewhere). Previous reports of sucrose gradient centrifugation suggested that the 149 kDa protein is monomeric (Newcomb et al., 1999). However, Figure 7A shows typical sedimentation profiles with a sloping plateau region, indicating the existence of large aggregates, together with two separate major boundaries from discrete smaller species. Experiments at different loading concentrations and rotor speeds led to similar distributions, but with slightly different peak areas, consistent with a slow and at least partially reversible oligomerization.

With the data of Figure 7A, both versions of G(s) are not applicable because of the absence of a solution plateau. No consistent boundary fractions (or area fractions, respectively) can be defined. Therefore, only the differential sedimentation coefficient distributions can be compared (Figure 7C and 7D). All distributions clearly show two maxima corresponding to the two visible separating boundaries, with the peaks in c(s) clearly being the best resolved. In order to avoid broadening from large time-intervals only a small subset of absorbance scans can be utilized in the calculation of g(s*) by dc/dt (solid line in Figure 7C), limiting the range of g*(s) under conditions where the peaks are well resolved. Because the *ls-g*(s)* method does not require the approximation of dc/dt by $\Delta c/\Delta t$ (Schuck and Rossmanith, 2000), a much larger number of scans can be incorporated in the analysis, which shows the presence of a high number of larger

aggregates with s-values up to 25 S, consistent with the results from the c(s) analysis. However, although the c(s) distribution may suggest separate peaks for the larger species, a Monte-Carlo statistical analysis reveals that the apparent peaks at s-values larger than ~ 15 S may be induced by oscillations from the regularization procedure and not significant within the given level of noise in the data. However, this only refers to the exact position of the c(s) peaks, but not the existence of material at large s-values, which is highly significant.

Because of the formation of distinct boundaries, at least for the two slower sedimenting species, the oligomerization is slow on the time-scale of the sedimentation, and it seems possible to assign oligomeric states to the individual peaks. This, however, requires additional information which we have sought in sedimentation equilibrium and dynamic light scattering experiments. Global modeling of sedimentation equilibrium data at multiple rotor speeds and concentrations show that the majority of the protein is monomeric, but with significant contributions of small oligomers (data not shown; an example of sedimentation equilibrium profile modeled with monomer, dimer and tetramer is shown in the inset of Figure 8A). While the self-association scheme could not be identified, the data were consistent with an isodesmic association with contaminations of incompetent monomer. These results from sedimentation equilibrium show that the main peak of the c(s) corresponds to the monomer, and suggest the second peak to be a dimer (or possibly a trimer). With a monomer sedimentation coefficient of 6.8 S, we can calculate a Stokes radius (R_S) of 5.2 nm, and a frictional ratio of 1.5 (equivalent to a prolate ellipsoid with 2a = 25.4 and 2b = 4.5 nm). This value was applied in the c(s) analysis of Figure 7D for diffusional deconvolution. Nonlinear regression of the weight-average frictional f/f_0 (with a starting value of 1.0) converged to a best-fit value of 1.25, but with a final rms error of the fit that was not statistically different from that obtained with $f/f_0 = 1.5$. Further,

the c(s) curves calculated using both values virtually superimpose (data not shown). This confirms that a non-linear regression of f/f_0 leads to values sufficient in precision for the deconvolution of diffusion in the sedimentation coefficient distributions, although the obtained average frictional ratio itself is not suitable for the transformation of c(s) into precise molar mass distributions c(M).

Interesting from the methodological point of view is a transformation of the sedimentation equilibrium data into a 'model-free' molar mass distribution *c(M)*, as suggested earlier by Wiff and Gehatia (Wiff and Gehatia, 1976) (Figure 8A). This *c(M)* transform does not take advantage of our knowledge of the molar mass of the different oligomers, and it is mathematically equivalent to the continuous size-distribution analysis of the sedimentation velocity data. It shows a main peak at a molar mass ~ 200 kDa, distinctly higher than the molar mass of the monomer (149 kDa), clearly indicating the presence of oligomeric species. Unfortunately, the data at molar mass > 600 kDa are not reliable in this transformation because they are mainly governed by assumptions of sedimentation in the region of optical artifacts close to the bottom of the cell. In contrast to sedimentation velocity, the *c(M)* transform does not have sufficient information to resolve the different species. A similar situation is encountered in the interpretation of the dynamic light scattering data, which are commonly transformed to distributions of Stokes radii, R_S (Figure 8B). The scattering intensity has a peak at ~ 5 nm, but also extends to larger species. In order to better compare the relative abundance of the different species, the distribution was rescaled into relative weight concentrations as shown by the dotted line in Figure 8B. From these data, it appears that the scattering from particles with $R_S > 8$ nm are in very low abundance, despite their significant contribution to the scattered intensity. Like in the *c(M)* transform of the sedimentation equilibrium data, no resolution of the oligomers is

achieved. Nevertheless, the virtual absence of species with $R_S > 8$ nm indicates that the 10.3 S

peak seen in sedimentation velocity may be a dimer ($R_S = 6.9$ nm), and less likely a trimer ($R_S = $

10.3 nm) or even larger oligomers. Similarly, the 13.2 S peak may be a trimer ($R_S = 8.3$ nm) but

less likely a tetramer ($R_S = 10.6$ nm) or even a pentamer or a hexamer.

This example illustrates the current potential and limitations of the sedimentation coefficient

distributions from complex oligomeric mixtures. It demonstrates that complementary

information from sedimentation equilibrium and dynamic light scattering can be used (and is

required) for the detailed interpretation. This is despite the much lower resolution of these

methods due to the significantly more ill-conditioned analysis of exponentials as compared to the

Lamm equation solutions. Correspondingly, the additional information from the c(s) distribution

on the number and approximate size of species can be very important for the correct

interpretation of the sedimentation equilibrium data.

Analysis of continuous size-distributions of emulsions

As a last example, we analyzed a truly continuous size-distribution of lipid emulsion

particles. General physical characteristics of such particles and their use for the study of

apolipoproteins have been described by (MacPhee et al., 1977) and (M.A. Perugini, P. Schuck,

G.J. Howlett, submitted). In the current context, for illustrating the behavior of the size-

distribution we considered the data from a mixture of two different elution fractions after

sucrose-gradient ultracentrifugation. Figure 9A shows the experimental flotation data exhibiting

a bimodal boundary. For both of the fractions, we have measured the average diffusion

coefficient by dynamic light scattering (with hydrodynamic radii of 34 and 62 nm, respectively).

The dashed lines in Figure 9A are the calculated best-fit distributions based on two discrete

species with the predetermined diffusion coefficients. The comparison with the boundary spread

of the experimental data shows that the distributions of the fractions are broad, and that boundary

broadening by diffusion is relatively small, but cannot be neglected.

Here, the analysis with the *c(s)* method can be based on the knowledge that the emulsion

particles are spherical, i.e. that $f/f_0 = 1.0$. (For simplicity, we have used the mean partial specific

volume of 1.055 ml/g of the components of the emulsion mixture; a refinement taking into

account the full size-dependence of the partial-specific volume of the particles is included in

(M.A. Perugini, P. Schuck, G.J. Howlett, submitted)). The resulting size-distributions are shown

in Figure 9B. When using maximum entropy regularization, we obtained very noisy c(s)

distributions with several artificial spikes (dotted line). This is consistent with previous findings

that use of the maximum entropy method for broad continuous distributions can cause artificial

oscillations (Provencher, 1992). However, this difficulty can be circumvented effectively by the

use of Tikhonov-Phillips regularization (solid line). Through the second derivative minimization

of this procedure, one can make use of the broadness and smoothness of the distributions as prior

knowledge.

The *ls-g*(s)* method leads to a similar distribution, which is only slightly broader because

of the limited extent of diffusion (dashed line in Figure 9B). Further artificial broadening would

be expected from the approximation of *dc/dt* by $\Delta c/\Delta t$ in the g(s*) analysis, due to the large

boundary displacement between the absorbance scans (Schuck and Rossmanith, 2000; Philo,

2000). (This results in a convolution of the *g*(s)* distribution with a hyperbola segment of width

$\Delta s = s\Delta t/t$ (Schuck and Rossmanith, 2000); when restricting the analysis to scans 8 to 13, the

broadening for a single non-diffusing species would be ~ 100 S for the first peak, and ~ 200 S for

the second peak). The vHW analysis applied to a suitable data subset results in similar

information as *ls-g*(s)* or c(s) (Figure 9B, circles). However, less information on the faster

floating particles is obtained, and the two peaks from the two lipid emulsion fractions appear not

as well resolved as in the *c(s)* analysis.

Discussion

In this paper, we have systematically explored and extended different size-distribution

analysis techniques for sedimentation velocity, and compared their resolution, sensitivity, and

assumptions made. For small molecules, the resolution depends centrally on the deconvolution

of diffusional broadening of the sedimentation boundary.

The vHW method is widely considered the most general approach, as no a priori

assumptions are necessary that relate sedimentation to diffusion of the individual boundary

fractions, yet, by extrapolation to infinite time, in theory, true sedimentation coefficient

distributions are obtained. While it has been proven an invaluable technique for a large class of

problems, there are limitations to this approach in practice. First, it is restricted to data that have

a well-defined solution plateau over the entire time-period analyzed, in order to allow the

division of the sample in equivalent boundary fractions (Demeler et al., 1997). Our results show

that this constraint to small time intervals can be crucial for resolving closely spaced species

(example 1), whereas for broad distributions no suitable scans may be found (Figure 7A).

Another well-known limitation to date is the requirement for absence of systematic noise, such

as superimposed to the fringe shift data of the interference optics. However, we have shown that

this can be overcome by using the analogy of the boundaries formed by non-diffusing species

(step functions) in the apparent sedimentation coefficient distribution *ls-g*(s)* (Schuck and

Rossmanith, 2000), and the boundary divisions in the vHW method (van Holde and Weischet,

1978). This analogy leads to a natural extension of the vHW method to interference optical data

(see below). Although the step-functions of non-diffusing species in *ls-g*(s)* lead to a slightly

different boundary division, because the differential radial dilution is taken into account, whereas

this effect is neglected in the vHW boundary division by separate reference to the plateau

concentration for each scan (van Holde and Weischet, 1978), both approaches obtain similar

precision with model data derived from numerical solutions of the Lamm equation (Eq. 2) and

from Faxén solutions (Eq. 10).

A maybe more fundamental limitation is illustrated in results from Figure 1 and 2, where

vHW was not successful in deconvoluting diffusion in the presence of heterogeneity. Similar

results can be obtained for other mixtures of relatively small molecules under experimentally

accessible conditions. These findings are consistent with those reported earlier by Demeler et al

(Demeler et al., 1997). A way for improving the precision of vHW has been suggested in theory

by introduction of a higher-order term in the Faxén solution of the Lamm equation, which leads

to more complex extrapolations (van Holde and Weischet, 1978). However, in the problem of

Figure 1, we found the conventional Faxén approximation Eq. 10 not to be limiting (by using Eq.

10 for simulating model boundaries, data not shown). The more fundamental reason for the

limited resolution appears to be that the linear extrapolation to infinite time is only accurate

when the boundary fraction represents a homogeneous population of species; for this case the

extrapolation has been shown to conform with the Lamm equation solution (van Holde and

Weischet, 1978). For a mixture of two species with different sedimentation coefficients, the

right hand side of Eq. 11 will be composed of a sum of two error functions, and at times when

the two species still have overlapping boundaries, both will be significant for the intermediate

fractions. As a consequence, the simple inversion of the error function (i.e. Eq. 9 of (van Holde and Weischet, 1978)) breaks down, and the extrapolated s-value has an s-value intermediate to the s-values of two species. This can be seen in Figure 10, where the overlap of two species was modulated by the different diffusion coefficients. This shows that only when the boundaries are essentially separated (e.g. when the sum of the rms diffusion distance of the species at a given time is much smaller than the difference of the average sedimentation distance of the species) can the species be separated and G(s) approaches a piecewise vertical line. Interestingly, the extreme upper and lower boundary fractions approach homogeneity of the population much earlier, such that no s-values are obtained outside the range defined by the true sedimentation coefficients of the species present. Therefore, the diffusional deconvolution of vHW defines the range of s-values, but cannot resolve different species unless they exhibit clearly separated sedimentation boundaries. For very large particles with low diffusion coefficients, this is easily fulfilled within the space and time-scales accessible in ultracentrifugation experiments. For smaller macromolecules, however, the G(s) curves at present appear to have mainly qualitative diagnostic information on the heterogeneity.

The extrapolation of g*(s) to infinite time has been suggested before by others (e.g., (Gosting, 1952; Fujita, 1975; Stafford, 1992b)), but difficulties in the choice of the appropriate extrapolation procedure were described, including the problem of extrapolations to negative distribution values (Stafford, 1992b). One of the key differences between these earlier (Stafford, 1992b) and the present approach is that we performed the extrapolation of g*(s) not at constant s, but at constant area (which is similar to the extrapolation of the integral apparent sedimentation coefficient distribution to infinite dilution described by Gralén and Lagermalm (Gralén and Lagermalm, 1952)). By analogy of the area fractions and boundary divisions, one can use the

theoretical argument in vHW which leads to $t^{-1/2}$ as the extrapolation scale (van Holde and Weischet, 1978). A second significant advantage of the current extrapolation procedure compared to previous methods is the absence of artificial broadening in the *ls-g*(s)* distributions (Schuck and Rossmanith, 2000). Figure 3 illustrates that this method can work very well with experimental data. Interestingly, this procedure also results in an estimate of the systematic signal contributions of the interference data. However, we also observed a sensitivity to outlier distributions *ls-g*(s)$_i$* (data not shown). Empirically, we found that with interference data taken in rapid succession (such as in Figure 3, which shows only every 10th scan), at least 10 to 20 scans per *ls-g*(s)$_i$* curve were required to obtain stable results. Further, inclusion of too early scans, as well as allowing a too broad s-range in the calculation of *ls-g*(s)$_i$* were detrimental for the resulting *G(s)* extrapolation. In part, these difficulties may be due to correlated baseline parameters, which becomes more problematic when analyzing smaller molecules. In some cases, we found it helpful not allow simultaneously for time-invariant and radial-invariant noise. For fundamental reasons, however, we believe that the extrapolation of *ls-g*(s)* to infinite time cannot overcome the limitations described above for the vHW procedure, in particular those for small molecules. It is important to note that the extrapolation of boundary fractions neglects the (nonlinear) details of the evolution of the boundary position, as well as the information on the radial dilution of the boundary, which by itself can reveal the weight-average sedimentation coefficient (Fujita, 1962). This information is utilized, however, in the c(s) method.

Although c(s) is computationally the most complex, it has the potential for highest resolution and sensitivity of the methods described. It is also very stable, and due to its use of numerical solutions of the Lamm equation, it has no constraints in the data set that can be considered for analysis, and is the most general with respect to size-range for the distribution.

However, there are some key assumptions made in this technique in the form of statistical and experimental prior knowledge, which must be properly adjusted in order to avoid poor or misleading results:

(1) The strategy uses hydrodynamic prior knowledge in order to estimate the extent of diffusion for each species. The first example shows how this can affect qualitatively the distribution (compare dotted and solid lines in Figure 1); the second shows how it can influence quantitatively the resolution and sensitivity (Figure 6). Fortunately, there are many different ways of entering the prior assumptions to match our knowledge of the macromolecular system under study, ranging from estimated similarity in the hydrodynamic shapes, known diffusion coefficients, to knowledge of the molar masses of the main species (Schuck, 2000; Schuck et al., 2000). Very important in this respect is that $c(s)$, like the ls-$g^*(s)$ method, is a direct boundary model rather than a data transformation. The quality of the fit is a very good criterion for the validity of the prior assumptions, and we have shown in the present paper that nonlinear regression can be used to obtain estimates for the numerical value of the parameters used for distribution analysis, in our case the weight-average frictional ratio. The same strategy applied to the related $c(M)$ sedimentation velocity analysis (Schuck, 2000) will reduce distortions in this molar mass distribution that can be a result of poor estimates for the frictional ratio. Other parameters that can be optimized in the direct boundary models are the exact meniscus position, as well as the sedimentation and diffusion coefficient of an additional discrete species (such as unmatched buffer salts; data not shown). It is particularly important to obtain a good fit of the data when interpreting details of the calculated distribution, such as minor peaks. Because of the importance of the inspection of the quality of fit, we have introduced a new tool to visualize the residuals in the form of a bitmap, that allows to distinguish systematic deviations that evolve in

parallel with the boundary movement from instrumental imperfections, which are predominantly stationary or periodic.

(2) Another, more subtle prior assumption is of a statistical nature and enters the analysis in the regularization algorithms. This technique allows the inversion of otherwise highly ill-conditioned integral equations, by selecting the most parsimonious of the (in general infinite) set of possible distribution that fit the data similarly well. Consistent with results in many other problems of this type in other fields, we found that the maximum entropy principle, which is the default choice in SEDFIT, gives generally good results. This algorithm is based on the assumption that the distribution with the least amount of information (or largest informational entropy) is least likely to contain artificial details (Livesey et al., 1987), which would otherwise typically dominate the distribution. It is non-local, and therefore can tolerate isolated peaks well, which is desirable when dealing with a sample that likely contains a discrete set of species. On the other hand, the selection of the distribution with the maximal entropy did not give good results with broad, truly continuous distributions (Figure 9), where it tends to produce artificial oscillations. Better results were obtained in this case with the Tikhonov-Phillips regularization, which in our implementation favors the distribution with smallest second derivative. The underlying assumption in this case is the absence of isolated peaks, which fits well the expectation for smooth distributions, such as those generated by random processes. As a consequence, adjustment of this second prior assumption should be unproblematic in most cases.

It could be argued that the option of introducing additional information in the analysis of the sedimentation velocity profiles is a significant advantage rather than an unfortunate byproduct of the method. This is obvious in cases where such information is readily available in a simple form, such as the known molar mass values in the study of the conformational changes

of the rotavirus nonstructural protein NSP2 (Schuck et al., 2000), or the known shape and size-dependence of the partial-specific volume of the emulsion particles (M.A. Perugini, P. Schuck, G.J. Howlett, submitted), or a known diffusion coefficient (Schuck, 2000). Our studies on the herpes simplex capsid protein VP5 in the present paper indicate how additional data from sedimentation equilibrium and dynamic light scattering can contribute important information on the size-distribution, although, by themselves, these methods are of significantly lower resolution than sedimentation velocity. The difficulties of arriving at a consistent interpretation of the oligomeric distribution illustrate that it would be highly desirable to incorporate these data into a single global size-distribution analysis, thus also incorporating the advantages of high resolution obtained in sedimentation velocity into sedimentation equilibrium and/or dynamic light scattering modeling. Similarly, this could substitute the current hydrodynamic prior assumptions by the additional, complementary data. Importantly, because the *c(s)* method is based on the least-squares fit of an explicit boundary model, it has the potential for such extension.

Literature Cited

Amato, U., and W. Hughes. 1991. Maximum entropy regularization of Fredholm integral equations of the first kind. *Inverse Problems.* 7:793-808.

Baldwin, R. L., and J. W. Williams. 1950. Boundary spreading in sedimentation velocity experiments. *Am. Chem. Soc.* 72:4325.

Bevington, P. R., and D. K. Robinson. 1992. Data Reduction and Error Analysis for the Physical Sciences. Mc-Graw-Hill, New York.

Bridgman, W. B. 1942. Some physical characteristics of glycogen. *Am. Chem. Soc.* 64:2349-2356.

Claverie, J.-M., H. Dreux, and R. Cohen. 1975. Sedimentation of generalized systems of interacting particles. I. Solution of systems of complete Lamm equations. *Biopolymers.* 14:1685-1700.

Cole, J. L., and V. M. Garsky. 2001. Thermodynamics of peptide inhibitor binding to HIV-1 gp41. *Biochemistry.* 40:5633-5641.

Demeler, B., H. Saber, and J. C. Hansen. 1997. Identification and interpretation of complexity in sedimentation velocity boundaries. *Biophys J.* 72:397-407.

Fujita, H. 1962. Mathematical Theory of Sedimentation Analysis. Academic Press, New York.

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

Gosting, L. J. 1952. Solution of boundary spreading equations for electrophoresis and the velocity ultracentrifuge. *J. Am. Chem. Soc.* 74:1548-1552.

Gralén, N., and G. Lagermalm. 1952. A contribution to the knowledge of some physico-chemical properties of polystyrene. *J. Phys. Chem.* 56:514-523.

Hatters, D. M., L. Wilson, B. W. Atcliffe, T. D. Mulhern, N. Guzzo-Pernell, and G. J. Howlett. 2001. Sedimentation analysis of novel DNA structures formed by homo-oligonucleotides. *Biophys. J.* 81:371-381.

Johnson, M. L., and M. Straume. 1994. Comments on the analysis of sedimentation equilibrium experiments. *In* Modern Analytical Ultracentrifugation. T. M. Schuster and T. M. Laue, editors. Birkhäuser, Boston. 37-65.

Kar, S. R., J. S. Kinsbury, M. S. Lewis, T. M. Laue, and P. Schuck. 2000. Analysis of transport experiments using pseudo-absorbance data. *Anal. Biochem*. 285:135-142

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

Laue, T. M., B. D. Shah, T. M. Ridgeway, and S. L. Pelletier. 1992. Computer-aided interpretation of analytical sedimentation data for proteins. *In* Analytical Ultracentrifugation in Biochemistry and Polymer Science. S. E. Harding, A. J. Rowe, and J. C. Horton, editors. The Royal Society of Chemistry, Cambridge. 90-125.

Livesey, A. K., M. Delaye, P. Licinio, and J.-C. Bronchon. 1987. Maximum entropy analysis of dynamic parameters via the Laplace transform. *Faraday Discuss. Chem. Soc.* 83:247-258.

Livesey, A. K., P. Licinio, and M. Delaye. 1986. Maximum entropy analysis of quasielastic light scattering from colloidal dispersions. *J. Chem. Phys.* 84:5102-5107.

MacPhee, C. E., R. Y. S. Chan, W. H. Sawyer, W. F. Stafford, and G. J. Howlett. 1977. Interaction of lipoprotein lipase with homogeneous lipid emulsions. *J. Lipid Res.* 38:1649-1659.

Murphy, R. M. 1997. Static and dynamic light scattering of biological macromolecules: what can we learn? *Curr. Opinion Biotechnology.* 8:25-30.

Newcomb, W. W., F. L. Holma, D. R. Thomsen, B. L. Trus, N. Cheng, A. C. Steven, F. P. Booy, and J. C. Brown. 1999. Assembly of the herpes simplex virus procapsid from purified components and identification of small complexes containing the major capsid and scaffolding proteins. *J. Virol.* 73:4239-4250.

Perugini, M. A., P. Schuck, and G. J. Howlett. 2000. Self-association of human apolipoprotein E3 and E4 in the presence and absence of phopholipid. *Journal of Biological Chemistry.* 275:36758-36765.

Philo, J. S. 1997. An improved function for fitting sedimentation velocity data for low molecular weight solutes. *Biophys. J.* 72:435-444.

Philo, J. S. 2000. A method for directly fitting the time derivative of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions. *Anal. Biochem.* 279:151-163.

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

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

Provencher, S. W. 1982b. CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comp. Phys. Comm.* 27:229-242.

Provencher, S. W. 1992. Low-bias macroscopic analysis of polydispersity. *In* Laser light scattering in biochemistry. S. E. Harding, D. B. Satelle, and V. A. Bloomfield, editors. The Royal Society of Chemistry, Cambridge, U.K. 92-111.

Scholte, T. G. 1968. Molecular weights and molecular weight distribution of polymers by equilibrium ultracentrifugation. Part II. Molecular weight distribution. *J. Polym. Sci.* 6:111-127.

Schuck, P. 1998. Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys. J.* 75:1503-1512.

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

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

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

Schuck, P., and P. Rossmanith. 2000. Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers.* 54:328-341.

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

Sedlák, M., and H. Cölfen. 2001. Synthesis of double-hydrophilic block copolymers with hydrophobic moieties for the controlled crystallization of minerals. *Macromolecular Chemistry and Physics.* 202.

Signer, R., and H. Gross. 1934. Ultrazentrifugale Polydispersitätsbestimmungen an hochpolymeren Stoffen. *Helv. Chim. Acta.* 17:726.

Stafford, W. F. 1992a. Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Anal. Biochem.* 203:295-301.

Stafford, W. F. 1992b. Methods for obtaining sedimentation coefficient distributions. *In* Analytical Ultracentrifugation in Biochemistry and Polymer Science. S. E. Harding, A. J. Rowe, and J. C. Horton, editors. The Royal Society of Chemistry, Cambridge, U.K. 359-393.

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

van Holde, K. E., and W. O. Weischet. 1978. Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers.* 17:1387-1403.

Vinograd, J., and R. Bruner. 1966. Band centrifugation of macromolecules in self-generating density gradients. II. Sedimentation and diffusion of macromolecules in bands. *Biopolymers.* 4:131-156.

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

Figure Legends

<u>Figure 1</u>: Study of the resolution of *c(s)*, *ls-g*(s)*, and *dc/dt*-based *g(s*)* distributions for a four-component model system of elongated molecules as suggested in (Stafford, 1992b). Panel A: Sedimentation profiles of species with relative molar masses 341,000, 398,000, 441,000, and 490,000 and sedimentation coefficients of 6, 7, 8, and 9 S, respectively (corresponding to anhydrous frictional ratios of 2.93, 2.79, 2.61, and 2.49). The simulated rotor speed was 50,000 rpm, rotor temperature 20°C, partial specific volume 0.73 cm^3/g, and the loading concentration for each species 0.25. The simulated data (thin lines, only every 2nd data set shown) was generated by adding the distributions of all species and 0.01 normally distributed noise. *c(s)* analysis was performed with 150 species with sedimentation coefficients between 4.5 and 10.5 S, with regularization at a confidence level of 0.68, and a floating frictional ratio and floating baseline (similar results were obtained with and without consideration of time-invariant noise). The frictional ratio converged at a value of 2.9, close to the weight-average frictional ratio of 2.7 of the simulated species. The best-fit distributions are shown as bold dashed lines. Panel B: The calculated *c(s)* distribution (bold solid line) is shown, and, for comparison, the results of the *ls-g*(s)* analysis of scans 10 to 30 (bold dashed line), *dc/dt* analysis of scans 18 to 21 (thin solid line), vHW analysis of scans 1 to 25 (circles), and G(s) by extrapolation of *ls-g*(s)* to infinite time (diamonds). Also shown is the distribution *c(s)* obtained with a value of *f/f₀* of 2.0 (dotted line, reduced in scale by factor 0.2).

<u>Figure 2</u>: Sedimentation coefficients of the boundary fractions of scans 1 to 25 of the data shown in Figure 1A, determined directly by the vHW method and by extrapolation of area fractions of *ls-g*(s)* to infinite time. For the vHW analysis (Panel A), the maximal number of

boundary fractions was 20, and shown are the apparent s-values (circles) and extrapolations

(lines) for fractions 1 to 19. (Fraction 20 near the plateau was dominated by noise, not shown).

The *ls-g*(s)* extrapolation (Panel B) was based on calculated *ls-g*(s)* distributions with a grid of

50 s-values between 1 and 15 S, with regularization at a confidence level of P = 0.9, allowing for

a floating baseline. This was applied to a 'sliding subset' of 5 sequential scans. Each *ls-g*(s)*

distribution was divided into 30 equal area fractions, for which the weight-average sedimentation

coefficient was determined (circles) and extrapolated to infinite time (solid line). The

distributions G(s) are shown in Figure 1B.

Figure 3: Van Holde Weischet analysis by extrapolation of *ls-g*(s)* of interference data from an

immunoglobulin sample. Panel A: Experimental fringe profiles at 40,000 rpm, 20 °C, scans 30

– 170. For clarity of presentation, only every 10th data set is shown, and the integral fringe shift

is removed. The best-fit time-invariant noise contribution from the G(s) analysis is shown as a

dotted line. Panel B: *ls-g*(s)* distributions (solid lines) calculated for sets of 20 scans, with 50 s-

values between 2 S and 11 S, and regularization at P = 0.9. The circles represent the integral

sedimentation coefficient distributions G(s), obtained by extrapolating 30 area fractions of *ls-

g*(s)* to infinite time. For comparison, G(s) is shown when derived from *ls-g*(s)* distributions

with lower resolution (30 s-values from 2 S to 11 S, dashed line), or from *ls-g*(s)* distributions

calculated on the basis of fewer scans (6, dotted line), and from conventional vHW analysis after

removing best-fit systematic noise components (fractions 2 – 29, squares). Panel C: s_{app} values

of the area fractions (circles) and linear extrapolation to infinite time on a $t^{-0.5}$ scale (lines).

Figure 4: Differential sedimentation coefficient distributions from the data shown in Figure 3A. Scans 110 to 139 were analyzed with the dc/dt-method (solid line), and with the *ls-g*(s)* method using 100 s-values between 4 S and 12 S, and Tikhonov-Phillips regularization at a confidence level of 0.68 (circles). The c(s) analysis (dotted line) was calculated on the basis of scans 1 to 200, with an average frictional ratio of 1.58, 200 s-values between 4 S and 14 S, and maximum entropy regularization at a confidence level of 0.95 (leading to an rms error of the fit of 0.00267 fringes).   The inset shows the same distributions at an expanded view for visualizing the contaminating larger species.

Figure 5: Dependency of the quality of fit with the c(s) method on the assumed average frictional ratio. Scans 1 to 200 of the immunoglobulin experiment shown in Figure 3A were analyzed with 200 s-values from 4 S to 14 S, P = 0.9, and using different values of f/f_0. (The resulting distributions are shown in Figure 6.) The circles represent the rms error of the fits obtained with different values of $\eta_r \times f/f_0$ (η_r = 1.02). A linear regression converges at a value of $\eta_r \times f/f_0$ = 1.58. For comparison, the rms error of the discrete 2-component model is shown as a horizontal dotted line. The corresponding residual bitmaps at representative values of $\eta_r \times f/f_0$ are shown as insets, with residual values at each data point depicted as bright (positive) or dark (negative) pixel. Different radius values are represented by pixel columns (meniscus left, bottom right), and scans are represented as pixel rows (first scans top, last scans bottom). The diagonal structures visible indicate large residuals propagating in time with the sedimentation boundary.

Figure 6: Effect of the deconvolution of diffusion on the calculated c(s) distributions for the immunoglobulin experiment. Distributions are calculated for scans 1 to 200, at P = 0.9, and

normalized at the main peak. The differences of the quality of the fit are shown in Figure 5. The

limiting case of no consideration of diffusion is illustrated by the apparent sedimentation

coefficient distribution *ls-g*(s)* (dashed line). This corresponds to an infinite value of $\eta_r \times f/f_0$.

With finite values, the main peak of the c(s) distribution decreases in width with decreasing

$\eta_r \times f/f_0$ at constant peak position (solid lines, values indicated in graph). The best-fit value of

$\eta_r \times f/f_0 = 1.58$ is shown as a bold line. c(s) curves at values smaller than 1.3 virtually

superimpose the 1.3 curve (data not shown). The inset shows the dependence of the location of

the minor peak of the c(s) curves on the value of $\eta_r \times f/f_0$: 1.70 (dotted line), 1.58 (bold solid line),

1.55 (solid line), 1.30 (dotted line). With a number of data points ~ 300,000, the curves for 1.58

and 1.55 are statistically indistinguishable on a 90% confidence limit, suggesting an error of the

peak position of ~ 0.3 S. The different heights of the peaks in the inset are a result of the

normalization of the distribution at the main peak.

Figure 7: Sedimentation velocity experiment with preparations of the herpes simplex virus

capsid protein VP5. The protein was prepared as described in (Newcomb et al., 1999) followed

by dialysis against PBS. The absorbance profiles were acquired at 230 nm, at a protein

concentration of 0.16 mg/ml, a rotor temperature of 8 °C and a rotor speed 55,000 rpm. Panel A:

Measured absorbance distributions (thin lines) and best-fit distribution from the Lamm equation

model c(s) (dashed bold lines). For clarity, only every third scan is shown. Panel B: Residuals

of the fit, which has a rms error of 0.0123 OD_{230}. Panel C: Apparent sedimentation coefficient

distributions *g(s*)* calculated by *dc/dt* (solid line) and distributions *ls-g*(s)* (dashed line). Panel

D: Best-fit *c(s)* sedimentation coefficient distribution (solid line), allowing for systematic time-

invariant noise. The 68% confidence band from Monte Carlo simulations (1,000 iterations, calculated at slightly lower resolution in s) is shown as dashed and dotted lines.

Figure 8: Panel A: Sedimentation equilibrium experiment with preparations of VP5 at a loading concentration of 0.26 mg/ml, a rotor speed of 10,000 rpm and a temperature of 4 °C, with data acquired at a wavelength of 230 nm. Shown is the calculated c(M) distribution obtained with maximum entropy regularization (P = 0.9). The inset shows the raw data (circles) modeled (as part of a global fit) with monomer, dimer and tetramer (solid line), and the calculated distributions for monomer (dotted line), dimer (dashed line), and tetramer (dash-dotted line). Panel B: Distribution of Stokes-radii as obtained upon analysis of the autocorrelation function from dynamic light scattering experiments. The distribution of scattering intensity versus radius is shown as a solid line and the estimated relative weight concentrations (based on a constant frictional ratio) are shown as a dashed line.

Figure 9: Flotation experiments with mixtures of fractionated lipid emulsions. Panel A: Experimental absorbance distributions of a mixture of two sucrose-gradient ultracentrifugation fractions, measured in intervals of 360 seconds (circles, shown is every second scan and every second data point of sets 1 – 25 analyzed). For experimental details, see (M.A. Perugini, P. Schuck, G.J. Howlett, submitted). To illustrate the extent of diffusion of the particles, the dashed lines show the best-fit models with separate species for each of the two fractions, using the diffusion coefficients as predetermined by dynamic light scattering for each fraction. Panel B: Differential sedimentation coefficient-distributions, calculated with a frictional ratio of 1.0, a partial-specific volume of 1.055 ml/g, and with regularization at P = 0.9. Shown are c(s) with

Tikhonov-Phillips regularization (solid line) and maximum entropy regularization (dotted line),
and *ls-g*(s)* (dashed line). The integral sedimentation coefficient distribution *G(s)* by the vHW
method, based on scans 7 – 14, is shown with solid circles, scaled to 0.003.

Figure 10: Dependence of the resolution of two species by vHW on boundary separation.
Simulated sedimentation boundaries in the Faxén approximation (Eq. 10) for two species with
sedimentation coefficients 5 S and 7 S at a rotor speed of 50,000 rpm (no noise was added). The
diffusion coefficient for both species was the same, and varied in different simulations from $D =$
1×10^{-7} cm^2/sec to 8×10^{-7} cm^2/sec in increments of 1×10^{-7} cm^2/sec. A set of suitable
sedimentation profiles centered at 4000 sec was taken for vHW analysis, the upper inset shows
these sedimentation boundaries for $D = 3 \times 10^{-7}$ cm^2/sec. The resulting sedimentation coefficient
distributions G(s) are superimposed (the data for $D = 1 \times 10^{-7}$ cm^2/sec are shown as dotted line,
those for $D = 3 \times 10^{-7}$ cm^2/sec as dashed line). For the conditions used, the separation of ideal
(non-diffusing) boundaries at 4,000 sec is 0.1 cm, compared to an rms diffusional spread of the
boundaries of 0.02 cm at $D = 1 \times 10^{-7}$ cm^2/sec, 0.034 cm at $D = 3 \times 10^{-7}$ cm^2/sec, and 0.056 cm at D
$= 8 \times 10^{-7}$ cm^2/sec. The lower inset shows the result of a c(s) analysis of simulated sedimentation
data with the highest diffusion coefficient in the presence of 1% noise (calculated with f/f0 = 1.0
and P = 0.9).

A

B