Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling

Peter Schuck

Molecular Interactions Resource, Bioengineering and Physical Science Program, ORS, National Institutes of Health, Bethesda, Maryland 20892 USA

ABSTRACT   A new method for the size-distribution analysis of polymers by sedimentation velocity analytical ultracentrifugation is described. It exploits the ability of Lamm equation modeling to discriminate between the spreading of the sedimentation boundary arising from sample heterogeneity and from diffusion. Finite element solutions of the Lamm equation for a large number of discrete noninteracting species are combined with maximum entropy regularization to represent a continuous size-distribution. As in the program CONTIN, the parameter governing the regularization constraint is adjusted by variance analysis to a predefined confidence level. Estimates of the partial specific volume and the frictional ratio of the macromolecules are used to calculate the diffusion coefficients, resulting in relatively high-resolution sedimentation coefficient distributions $c(s)$ or molar mass distributions $c(M)$. It can be applied to interference optical data that exhibit systematic noise components, and it does not require solution or solvent plateaus to be established. More details on the size-distribution can be obtained than from van Holde–Weischet analysis. The sensitivity to the values of the regularization parameter and to the shape parameters is explored with the help of simulated sedimentation data of discrete and continuous model size distributions, and by applications to experimental data of continuous and discrete protein mixtures.

INTRODUCTION

The characterization of the size distribution of polymers is one of the principal problems in the study of biological macromolecules and of synthetic polymers. Numerous techniques based on a variety of different principles have been developed for this task, ranging, for example, from high-resolution mass spectrometry, dynamic light scattering, analytical ultracentrifugation, size-exclusion chromatography, field-flow fractionation, to gel electrophoresis. Analytical ultracentrifugation is the oldest of these techniques and has been surpassed by others with respect to precision and rapidity. However, for several reasons, a considerable interest in the use of ultracentrifugation for the characterization of size distributions still remains. First, it is attractive for its theoretical simplicity and firm basis on first principles. Hydrodynamic theory (and thermodynamics in sedimentation equilibrium) can be directly applied, and, for the separation of subpopulations of different size, no interaction with matrices, surfaces, or a bulk flow is required. Second, it is experimentally powerful and very versatile: the macromolecules are characterized in solution, and can be studied at a large range of concentrations, provided for by fluorescence (Laue et al., 1997; Schmidt and Riesner, 1992), interference (Laue, 1994; Schachman, 1959; Yphantis et al., 1994), absorbance (Giebeler, 1992; Hanlon et al., 1962; Schachman et al., 1962), and Schlieren optical detection systems (Svedberg and Pedersen, 1940). It can be

applied to an extremely large macromolecular size range by adjustment of the rotor speed. Third, analytical ultracentrifugation experiments generally provide a large quantity of data with relatively high precision, and a significant amount of experience in this technique has been accumulated during the last seven decades.

Both sedimentation equilibrium and sedimentation velocity methods have been used in the long history of the characterization of the particle size distributions by analytical ultracentrifugation (Baldwin and Williams, 1950; Bridgman, 1942; Fujita, 1962; Lechner and Mächtle, 1992; Mächtle, 1999; Scholte, 1968; Signer and Gross, 1934; Stafford, 1992; Svedberg and Pedersen, 1940; van Holde and Weischet, 1978; Vinograd and Bruner, 1966). Sedimentation equilibrium analysis (Lechner and Mächtle, 1992; Scholte, 1968) seems intrinsically more problematic because of the difficulty involved in unraveling the sedimentation equilibrium exponentials, and, in some cases, the analysis has been constrained to parameterized model distributions (Lechner and Mächtle, 1992). Sedimentation velocity experiments provide a richer database, because they observe the strongly size-dependent time course of migration, although here the size-distribution information is convoluted by the hydrodynamic properties of the particles.

Several different sedimentation velocity methods have been developed. For very large particles where separation is achieved during the time of the experiment, a well-conditioned high-resolution analysis can be performed based on the spatial derivative of the sedimentation profiles, dc/dr (Baldwin and Williams, 1950; Bridgman, 1942; Fujita, 1962; Signer and Gross, 1934; Svedberg and Pedersen, 1940), or by the related method of observing the time course of sedimentation at a single radial position (Mächtle, 1999). For smaller particles, however, diffusion broadens the sed-

imentation boundary, which makes it more difficult to resolve subpopulations of the distribution. In this regime, an established and very useful method for analyzing size distributions is the apparent sedimentation coefficient distribution g*(s) (Frigon and Timasheff, 1975; Rivas et al., 1999; Schachman, 1959; Schuster and Toedt, 1996; Stafford, 1997), using dc/dr, or the more recently introduced time derivative dc/dt of the sedimentation profiles (Stafford, 1992). However, the apparent sedimentation coefficient distribution obtained is convoluted by a Gaussian due to diffusional broadening. An elegant and powerful method to overcome diffusional broadening has been described by van Holde and Weischet (Demeler et al., 1997; van Holde and Weischet, 1978). Here, by extrapolation of the apparent sedimentation coefficients of sedimentation boundary fractions to infinite time on a $t^{-0.5}$ scale, diffusion-free integral sedimentation coefficient distributions $G(s)$ are obtained.

All the established sedimentation velocity methods for size-distribution analysis are similar in that they use different transformations of the sedimentation data that have been analytically shown to reveal, under the condition of long solution columns, the sedimentation coefficient distribution. This approach has the virtue of a model-free analysis. In general, if a model for the sedimentation behavior of macromolecules is available, however, it is widely accepted that an analysis by directly fitting the model to the raw data can be superior in information and precision of the derived parameters, although this is frequently computationally more difficult. For example, more information can be obtained from long-column sedimentation equilibrium experiments of mixtures of ideal species by multiexponential decomposition of the raw data in global analyses, as now commonly in use, when compared to the more traditional ln(c) versus r^2 transformations of a single data set. The present study is concerned with the problem of formulating and exploring the properties of an explicit boundary model for the size-distribution analysis in sedimentation velocity experiments, based on numerical solutions to the equations that govern sedimentation and diffusion, the Lamm equations (Lamm, 1929). This allows larger data sets in the analysis of a single experiment and in global analyses of multiple experiments, and the incorporation of prior knowledge on the distribution, which, as will be demonstrated, can lead to a better resolution of size distributions.

Numerical solutions to the Lamm equations and their use for direct fitting of ultracentrifuge data have been developed previously in several laboratories (among them, Cann and Kegeles, 1974; Claverie et al., 1975; Cox and Dale, 1981; and Dishon et al., 1966). More recently, enabled by the increased computational resources, this became an efficient and readily available tool for sedimentation-velocity data analysis (Demeler and Saber, 1998; Schuck, 1998; Schuck et al., 1998; Stafford, 1998). Lamm equation analysis can take into account all boundary conditions of the finite length of the centrifugal cell and of the effects of diffusion, but, at

present, can only be applied to a few discrete species. This paper describes an extension of the Lamm equation analysis for the characterization of continuous size distributions of macromolecules. The problem is stated as an integral equation, and regularization is used for its numerical inversion. The properties of the method in the application to discrete distributions, and to broad, continuous size distributions are explored.

THEORY

In the absence of interactions between the macromolecules (or particles), the experimentally observed sedimentation profiles of a continuous size distribution can be described as a superposition of the contributions of each subpopulation $c(M)$ of particles with sizes between M and $M + $ dM. If $L(M, r, t)$ denotes the sedimentation profile of a monodisperse species of size M at radius r and time t, the problem is described by a Fredholm integral equation of the first kind,

$$a(r, t) = \int c(M)L(M, r, t)\, dM + \epsilon, \qquad (1)$$

where $a(r, t)$ denotes the experimentally observed signal, with an error of measurement ϵ. This equation is encountered in similar form in problems of polymer characterization in many other techniques. In the following, first the calculation of the kernel $L(M, r, t)$ will be outlined, and then a detailed description of the method used for inverting Eq. 1 by regularization will be given. This will closely follow the method applied by Provencher (1982a,b) in the program CONTIN.

Solution of the Lamm equation for a monodisperse subpopulation

In the case of sedimentation velocity ultracentrifugation of dilute solutions of a polymer, the kernel $L(M, r, t)$ of Eq. 1 is the solution of the Lamm equation (Lamm, 1929),

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

This partial differential equation describes the migration and diffusion of a dilute solution of monodisperse particles with concentration $\chi(r, t)$ in a sector-shaped cell under the influence of the centrifugal field generated at a rotor angular velocity ω. $s(M)$ and $D(M)$ are the sedimentation and diffusion coefficient of the particle, respectively. They are both strongly dependent on the molar mass, and are related by the Svedberg equation,

$$s(M) = D(M)\frac{M(1 - \bar{v}_M\rho)}{RT}, \qquad (3)$$

where ρ denotes the solvent density, R denotes the gas constant, and T denotes the rotor temperature (Svedberg and Pedersen, 1940). The partial specific volume $\bar{v}_M$ of the solute may also be dependent on the macromolecular size, but, in most cases, only weakly or even negligibly (such a weak dependence will be indicated by the subscript M).

It can be seen at this point that the sedimentation velocity analysis of particles with continuous size distributions is complicated by the fact that it requires knowledge of at least two functional dependencies on size: in addition to $c(M)$, it requires either the sedimentation coefficient $s(M)$, or, equivalently, the diffusion coefficient $D(M)$. Because the problem of Eq. 1 is ill-posed, even if it is known how the sedimentation coefficient changes with size, it seems impossible to calculate both distributions $c(M)$ and $s(M)$ from noisy experimental data. As will be described in the following, this problem is addressed by assuming prior knowledge of the partial specific volume $\bar{v}_M$ and the frictional ratio $(f/f_0)_M$ (i.e., prior knowledge of the hydrodynamic shape) of the macromolecules, which will allow calculation of $s(M)$ and $D(M)$. (Only in favorable cases of very narrow monomodal distributions or negligible diffusion does it seem feasible to treat either $\bar{v}_M$ or $(f/f_0)_M$ as a fitting parameter to be determined through the data analysis.)

Although $\bar{v}_M$ and $(f/f_0)_M$, in general, will also depend on the macromolecular size in many cases, either reasonable estimates or measurements can be made. In some cases, it may be a reasonable approximation that $\bar{v}_M$ and/or $(f/f_0)_M$ does not change with size; this may hold approximately true, for example, for particles such as random coils of polymers, lipid vesicles, emulsions, or, in a first approximation, even for mixtures of globular proteins. Alternatively, a parameterized model for $(f/f_0)_M$ could be used, such as the model of rodlike particles at a length-to-radius ratio that increases linearly with M. Similarly, if the particles can be approximated by multisubunit assemblies with regular geometry, values of $(f/f_0)_M$ could be derived with the help of hydrodynamic bead modeling (Bloomfield et al., 1967; de la Torre, 1992). In some cases, D may be constant, allowing the direct use of Eq. 3 to derive s as a function of the buoyant molar mass (an example of this, ferritin, is shown below). Finally, the values for $\bar{v}_M$ and $(f/f_0)_M$ may be measured in additional experiments for several fractionated subpopulations of the particles, which then can be combined with polynomial interpolation of the obtained values to approximate $(f/f_0)_M$ at any size. How possible errors in $\bar{v}_M$ and $(f/f_0)_M$ affect the calculated distributions $c(M)$ and $c(s)$ will be examined below.

Given $\bar{v}_M$, one can calculate the radius R of an equivalent sphere with the same volume as the particle by simple geometrical relationships (Laue et al., 1992). This leads to the minimum hydrodynamic frictional coefficient of an equivalent sphere. With the shape information of the particle expressed through the frictional ratio $(f/f_0)_M$, the diffu-

sion coefficient of the particle then follows from the Stokes–Einstein relationship as

$$D(M) = \frac{kT}{6\pi\eta_0\eta_r(f/f_0)_M R(M, \bar{v}_M)}, \qquad (4)$$

where k denotes the Boltzmann constant, and η_0 and η_r denote the standard and relative viscosity of the solution, respectively. This result can then be inserted into the Svedberg equation (Eq. 3) to obtain $s(M)$. Given $s(M)$ and $D(M)$ and their inverses $M(s)$ and $M(D)$, the size distribution $c(M)$ can then easily be transformed into a sedimentation coefficient distribution $c(s) := c(M(s))$ and a diffusion coefficient distribution $c(D) := c(M(D))$. These are basically equivalent descriptions of the distribution, although they represent different aspects of the particle size distribution.

After calculating s and D for a particle of size M, the numerical integration of the Lamm equation was started with the initial condition of a uniform concentration $\chi(r, 0) = 1$, and with graphically predetermined positions of the meniscus and bottom of the solution column (these can also be treated as floating parameters to be optimized in the nonlinear regression). Lamm equation solutions were calculated on a grid of between 200 and 500 radial points. For low values of $\omega^2 s$, the finite element method developed by Claverie et al. (1975) was used, combined with a Crank–Nicholson scheme (Crank and Nicholson, 1947) and an algorithm for adaptive step sizes in time (Schuck et al., 1998). For higher values of $\omega^2 s$, the moving grid finite element method (Schuck, 1998) was used. The later method is particularly well suited for the simulation of sedimentation of large particles with low diffusion coefficient, because it remains both numerically stable and relative efficient for very small values of D.

Analysis of the size distribution c(M)

For very large particles, the influence of diffusion flux on the particle distribution during the time of the sedimentation experiment is negligible compared to the sedimentation flux. As a consequence, $L(M, r, t)$ can be approximated by a step function $L(M, r, t) = \exp(-2\omega^2 s_M t) \times H(r - r^*(M, t))$ at a position $r^*(M, t) = r_m\exp(\omega^2 s_M t)$ (with the meniscus position r_m) (Fujita, 1962). In this limiting case, $r^*(M, t)$ can be used to change the integration variable in Eq. 1, and differentiation with respect to the radius r directly solves the integral. Therefore, the derivative of the measured concentration profiles at any time can be directly related to the particle size distribution (Baldwin and Williams, 1950; Bridgman, 1942; Fujita, 1962; Signer and Gross, 1934; Svedberg and Pedersen, 1940). Unfortunately, this approximation holds well only for larger macromolecules and is not suitable for many biopolymers.

The consideration of diffusion increases the complexity of Eq. 1, and the smoothness of the sedimentation bound-

aries of single species $L(M, r, t)$ makes Eq. 1 an ill-posed problem. As is characteristic for such problems, a large set of different $c(M)$ distributions may fit the data equally well, and a straightforward discretization and inversion usually leads to large, artificial high-frequency oscillations in $c(M)$.[1] It was observed that, for the present problem (in particular, in the case of narrow distributions), the condition of non-negativity imposed on $c(M)$ suppresses most of these oscillations. For further stabilization, regularization was used. Following the maximum entropy method, a term can be added to the inverse problem of Eq. 1,

$$\underset{c(M)}{\text{Min}}\left\{ \sum_{i,j}\left[a(r_i, t_j) - \int c(M)L(M, r_i, t_j)\, \mathrm{d}M \right]^2 \right.$$

$$\left. + \alpha \int c(M)\ln c(M)\, \mathrm{d}M \right\}, \quad (5a)$$

that maximizes the information entropy of $c(M)$ (Press et al., 1992; Smith and Grandy, 1985). For any positive value of α, this penalty term increases the rms error of the fit as compared to the optimal fit in the absence of regularization ($\alpha = 0$), and the increase of the ratio of the variance, $\chi^2(\alpha)/\chi^2(\alpha = 0)$, can be correlated with a probability P via F-statistics (Johnson and Straume, 1994). Therefore, F-statistics can be used to automatically adjust the regularization parameter α such that the quality of the fit still remains statistically indistinguishable from the unconstrained fit, based on a given confidence level P and on the level of the noise of the data (Bevington and Robinson, 1992; Provencher, 1979; Provencher, 1982a). The maximum entropy principle introduces the statistical prior probability that, in the absence of additional information, all sizes M are equally likely (this can be modified to incorporate more specific prior knowledge on the size distribution). The effect of the maximum entropy regularization term is the selection of the distribution $c(M)$ with the minimal amount of information in $c(M)$ required to fit the raw data.

Alternatively, Tikhonov–Phillips regularization with the term (Phillips, 1962)

$$\underset{c(M)}{\text{Min}}\left\{ \sum_{i,j}\left[a(r_i, t_j) - \int c(M)L(M, r_i, t_j)\, \mathrm{d}M \right]^2 \right.$$

$$\left. + \alpha \int |c''(M)|^2\, \mathrm{d}M \right\} \quad (5b)$$

was used. In contrast to maximum entropy, this procedure distinguishes the solutions $c(M)$ according to their smooth-

ness. But, when α is adjusted by the variance $\chi^2(\alpha)/\chi^2(\alpha = 0)$, it also selects from the set of all distributions $c(M)$ that lead to a statistically indistinguishable fit to the raw data the one distribution that exhibits the highest parsimony. As has been pointed out by Provencher (1982a), this procedure selects the solution that has the least amount of detail, but it ensures that the detail that is contained in the final distribution $c(M)$ is essential to describe the data, and therefore less likely to be an artifact.

For the numerical calculations, first the continuous molar mass distribution $c(M)$ of Eq. 1 is approximated by considering the concentrations c_k on a grid of N molar mass values M_k,

$$a(r_i, t_j) \cong b_i + \beta_j + \sum_{k=1}^{N} c_k L(M_k, r_i, t_j), \quad (6)$$

usually with $N = 100\text{--}200$. Depending on the optical system used for centrifugal data acquisition, which determines the noise structure in $a(r_i, t_j)$, Eq. 6 includes the algebraic time-invariant noise components b_i, and the radial-invariant jitter components β_j, as described in detail in Schuck and Demeler (1999). This allows the direct fitting of interference optical data from samples at low loading concentrations, where the systematic noise components due to optical imperfections are significant. After solution of the Lamm equations, normal equations of the least-squares problem Eq. 6 are formed,

$$\mathbf{y} = \mathbf{Ac} \quad (7)$$

$$\text{with} \quad y_k = \sum_{i,j} \hat{a}(r_i, t_j)\hat{L}(M_k, r_i, t_j)$$

$$A_{kl} = \sum_{i,j} \hat{L}(M_k, r_i, t_j)\hat{L}(M_l, r_i, t_j),$$

(using matrix notation) where the hat on $\hat{a}$ and $\hat{L}$ denotes the algebraic transformations required for the calculation of the systematic noise parameters (Schuck, 1999), and $\mathbf{c}$ denotes the vector of the concentrations $c_1, \ldots, c_N$. The maximum entropy minimization problem of Eq. 5a then leads to

$$\underset{c_k}{\text{Min}}\left\{ \mathbf{cAc} - 2\mathbf{yc} + \alpha \sum_k c_k \ln c_k \right\}, \quad (8a)$$

which can be solved by a Levenberg–Marquardt algorithm as described in (Press et al., 1992). In the case of the Tikhonov–Phillips regularization Eq. 5b, the resulting minimization remains a linear problem that can be directly solved as

$$\mathbf{y} = (\mathbf{A} + \alpha\mathbf{B})\mathbf{c}, \quad (8b)$$

where $\mathbf{B}$ denotes the square of the second difference operator as given in Eq. 18.5.12 of Press et al. (1992). Non-negativity of the concentrations c_k was achieved algebra-

---

[1]This can be understood by considering the lemma of Riemann-Lebesgue, as $\delta_m \equiv \int_a^b K(x, y)\sin(my)\, \mathrm{d}y \rightarrow 0$ as $m \rightarrow \infty$ for an integrable kernel (Phillips, 1962; Provencher, 1982a).

ically through the algorithm NNLS by Lawson and Hanson (1974), adapted for use with the normal equations and Cholesky decomposition.

As described above, the regularization parameter α was adjusted to reach the predetermined variance ratio calculated by F-statistics. Because of the large number of data points involved in the analysis, the influence of the constraints on the degrees of freedom is neglected. Unless noted otherwise, for any given data set, the variance ratio was calculated corresponding to a probability $p = 0.95$. With the usual number of experimental data points in the order of 10^4–10^5, the variance increase due to regularization is typically in the order of 1%. Finally, the distribution c_k was rescaled by trapezoidal integration such that the integral over $c(M)$ equals the total loading concentration.

The computational cost of the method is an important factor for practical use. It is determined mainly by two procedures: the N solutions of the Lamm equation and the $N \times N$ summations over all data points of the pairwise products of L involved in the calculation of the elements A_{kl} of the normal equations. The latter increases quadratically in N, and therefore determines the computation time for large N and large data sets. (The importance of this can be seen by considering a typical set of interference data: with 100 scans, 1000 data points per scan, and $N = 100$, 10^9 summation and multiplication operations are required to build the entire matrix A_{kl}.) For a relatively low number of data points or a lower resolution in $c(M)$, the solutions of the Lamm equations determine the computation time. During Monte Carlo simulations, these two steps need only to be calculated once. The inversion of Eq. 8 and the adjustment of α can be accomplished comparatively rapidly. In the current implementation of the program SEDFIT, when using moderate amounts of data (e.g., 1–2×10^4 data points, $N = 100$), the distribution can be calculated with a fast PC, typically, in significantly less than one minute, and one Monte Carlo iteration in a few seconds.

The sedimentation coefficient distribution analysis in the absence of diffusion was performed by replacing the Lamm equation distributions in Eq. 1 by the well-known step functions $U(r, t) = \exp(-2\omega^2 s_M t) \times H(r - r^*(M, t))$ at a position $r^*(M, t) = r_m \exp(\omega^2 s_M t)$ (Fujita, 1962; Stafford, 1992). This is closely related to the conventional $g^*(s)$ approximation of the sedimentation coefficient distribution in the absence of diffusion (Stafford, 1992), and, if applied to a data set from a small time interval, the numerical results are equivalent to those derived from dc/dt analysis (P. Schuck and P. Rossmanith, submitted).

EXPERIMENTAL

Sedimentation velocity experiments were performed with a Beckman Optima XL-A analytical ultracentrifuge equipped with absorbance optics. Horse spleen apoferritin (Sigma A3641) and horse spleen ferritin (Boehringer 197742) were diluted into PBS, and epon double-sector centerpieces were filled with 300 μl of the protein sample and PBS, respectively. Using an An50-Ti rotor, the samples were centrifuged at a rotor speed of 15,000 rpm at a temperature of 24°C. Scans were acquired at a wavelength of 230 nm in time intervals of 210 sec. The partial specific volume of 0.73 ml/g for apoferritin monomers was calculated based on the amino acid composition using the program SEDNTERP (Laue et al., 1992). $g^*(s)$ analysis was performed with the program DCDT+ (J. S. Philo, 3329 Heatherglow Ct., Thousand Oaks, CA 91360). Dynamic light scattering experiments were conducted with a DynaPro-MSTC200 (Protein Solutions, Charlottesville, VA), with the temperature control adjusted to 24°C.

Van-Holde–Weischet analyses were performed according to methods described in detail in Demeler et al. (1997) and van Holde and Weischet (1978). Briefly, the sedimentation boundaries were divided in N_f fractions of the plateau signal c_0, and the best least-square radial positions of the boundary fractions were calculated by averaging the radii of all data points a_f with absorbance values within the limits of each fraction (i.e., $(f - 1)^*c_0/N_f < a_f < f^*c_0/N_f$ for fraction f). The first and last fraction was omitted in the further analysis because of their larger noise in their calculated radial positions. N_f was chosen such that all fractions had at least one data point in each scan. Apparent s-values were calculated, and s-values at infinite time were determined by least-squares extrapolation in a $t^{-0.5}$ scale, as described in van Holde and Weischet (1978), defining an integral sedimentation coefficient distribution $G(s)$.

All computational methods were implemented into the Windows-based ultracentrifugal analysis program SEDFIT, which is available on request, or can be downloaded from http://www.biochem.uthscsa.edu/auc/software, and from the RASMB network at ftp://rasmb.bbri.org/rasmb/spin/ms_dos/.

RESULTS

The resolution of the method will be examined first for the case of relatively small molecules, where the influence of diffusion is comparatively large and no visual separation during the sedimentation process is achieved. Figure 1 *A* shows simulated sedimentation profiles of a discrete mixture of two spherical molecules of molar masses 30,000 and 50,000, and sedimentation coefficients of 3.4 and 4.78 S, respectively, at loading concentrations of 0.5 for each species, superimposed by a normally distributed error of 0.01. Also shown in Fig. 1 *A* are the best-fit single-component sedimentation profiles (*dashed lines*), which resulted in an apparent molar mass of 25,700 and a sedimentation coefficient of 4.1 S. The unphysical combination of such a low value for the apparent molar mass (or high value for the apparent diffusion coefficient, respectively) and this relatively high value for the sedimentation coefficient is a result of the broadening of the sedimentation boundary due to the
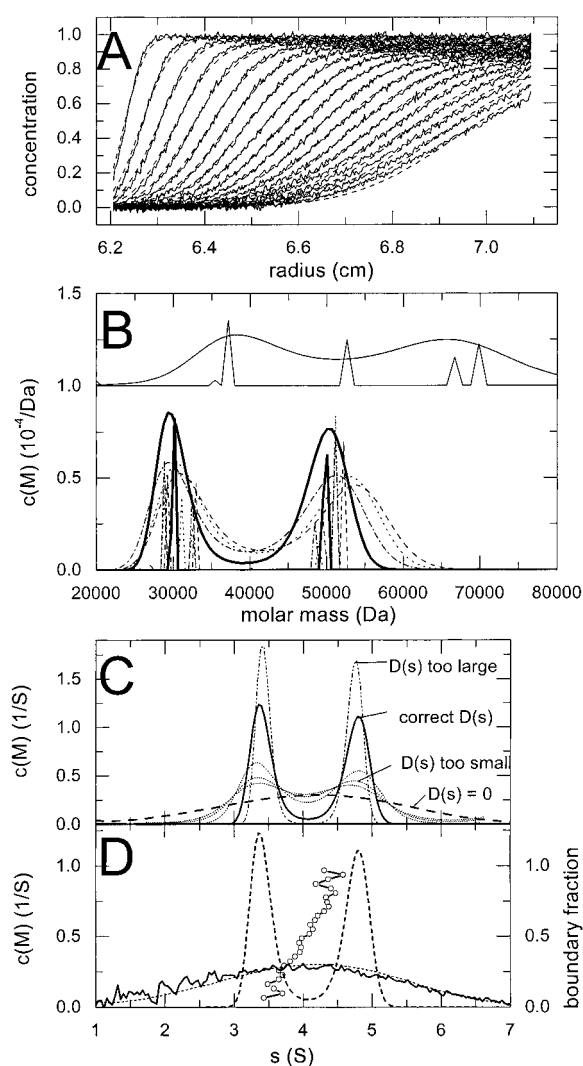
FIGURE 1 (A) Simulated sedimentation profiles of a discrete mixture of two components of molar masses 30,000 and 50,000, and sedimentation coefficients of 3.4 and 4.78 S, respectively, each with a partial specific volume $\bar{v} = 0.73$ cm³/g and a frictional ratio $f/f_0 = 1$, at a loading concentration of 0.5 (*solid lines*). Simulated conditions: $\rho = 1.0067$ g/cm³, $\eta_r = 1$, $\omega = 50,000$ rpm, $T = 20°$C, radial data interval 0.003 cm, and time-interval of scans 500 sec, Gaussian distributed error of measurement of 0.01. Included are the best-fit single-component sedimentation profiles (*dashed lines*), with a molar mass of 25,700 and a sedimentation coefficient of 4.1 S, with an rms error of 0.0155. (*B*) Calculated distributions $c(M)$ based on Eq. 8, with $N = 100$, and a baseline offset as a floating parameter. Shown are results for $\alpha = 0$ (distributions with spikes, at 10-fold reduced scale) and with maximum entropy regularization and α adjusted to a probability of $p = 0.68$ (*smooth curves*). The distributions calculated using the correct parameters for particle density and frictional ratio are shown as bold solid lines. Results of the analysis with incorrect parameters: $f/f_0 = 1.03$ (*dotted lines*), $f/f_0 = 1.04$ (*dashed lines*), $f/f_0 = 1.05$ and $\bar{v} = 0.72$ cm³/g (*dash-dot lines*). Results of larger deviations of the frictional ratio ($f/f_0 = 1.2$) are offset by 1×10^{-4}. (*C*) Transformation in sedimentation coefficient distributions $c(s)$ with α adjusted to $p = 0.683$. Using a value for the frictional ratio of $f/f_0 = 1.0$ (*bold solid line*) leads to the correct relationship $D(s)$ between the diffusion and the sediment coefficient. Models with $f/f_0 = 1.1$, $f/f_0 = 1.3$, and $f/f_0 = 1.5$ lead to an underestimation of the diffusion coefficients (*dotted lines*). The limit of no diffusion is presented as the dashed line. Using the parameters $f/f_0 = 1.0$ with $\bar{v} = 0.70$

underlying heterogeneity. It should be noted that the fit is not of acceptable quality (rms error = 0.0155), because a single-component model cannot describe the initially sharp but rapidly broadening sedimentation boundary well. This distinct difference between the diffusion broadening of the sedimentation boundary of a single sedimenting component, and the boundary shape of a heterogeneous mixture, provides the potential for gaining information on the size distribution. This difference will be larger when a larger relative separation of the size of the species is present, and in cases of larger particles with smaller diffusion coefficients (see Fig. 4 *B* below).

The calculation of the molar mass distribution is performed using Eq. 8, on a grid of $N = 100$ molar-mass values between 20,000 and 70,000. For the calculation of both sedimentation and diffusion coefficients $s(M)$ and $D(M)$ according to Eqs. 3 and 4, first spherical particles ($f/f_0 = 1$) with a partial-specific volume $\bar{v} = 0.73$ cm³/g were assumed (the values identical to those used for generating the data). In the absence of regularization ($\alpha = 0$) this results in sharp peaks at the correct molar masses underlying the simulation (Fig. 1 *B*). However, the location of these peaks depends strongly on the details of the simulated data and of the model. This is illustrated by the effect of using slightly incorrect frictional ratios, which leads to shifts of the location of these sharp peaks, or to fragmentation into two groups of 2–3 peaks, without significantly changing the rms error of the fit (<0.0101) (Fig. 1 *B*). This clearly demonstrates that the direct solution of Eq. 6 without regularization results in an unreliable level of detail. When the parameter α for the maximum entropy regularization is adjusted to a probability of $p = 0.68$, significantly smoother curves are obtained, which are much more robust against small errors in the model. The two components can still be clearly resolved (Fig. 1 *B*). Because the rms error of the fit is not significantly worse (<0.0104) than the fits without regularization, these curves reflect much better the information that can be extracted from the distribution analysis of the sedimentation data. Similar results were found when studying discrete distributions in a larger size range, or when using the Tikhonov–Phillips regularization (data not shown). Under comparable conditions with simulated noisy data, two discrete species with a 30% relative difference of the molar mass in the range of 100,000 and a 20% relative difference in the range of 1,000,000 could be resolved (data not shown).

cm³/g leads to an overestimation of the diffusion (*dash-dot-dot line*). (*D*) Result of a $g^*(s)$ analysis based on the dc/dt data transformation (*solid line*), and, for comparison, the distributions obtained for $D = 0$, and for the correct $D(s)$ from (*C*) (*dashed lines*). The integral sedimentation coefficient distribution of the van Holde–Weischet analysis is plotted as a function of the boundary fraction (–o–).

If the assumptions on the shape of the particles implied by $f/f_0 = 1$ and the assumed value of $\bar{v}$ do not lead to a reasonable approximation of $s(M)$ and $D(M)$, however, the regularized distributions $c(M)$ are significantly broader, limiting the resolution of the two species (Fig. 1 B, *offset data*). It should be noted that this is accompanied by an increase of the rms error of the fit (9% increase for the data shown with $f/f_0 = 1.2$ in Fig. 1 B). In cases where at least the assumption of shape similarity among the species is correct, this increase of the rms error can be used as the basis for nonlinear regression and fitting for the parameter f/f_0. (In the implementation used in SEDFIT with a simplex routine, $N = 50$ and $p = 0.68$, this converged rapidly to a best-fit value for f/f_0 of 1.005.)

An alternative representation of the distribution is the transformation into a sedimentation coefficient distribution $c(s)$ using the $s(M)$ relationship from the Svedberg equation (Eq. 3) (Fig. 1 C). The distributions $c(s)$ are much more robust than $c(M)$ against poor assumptions for the shape of the molecules: whereas errors in the frictional ratios lead to overall translations of $c(M)$, these errors only affect the resolution in $c(s)$, but not the location of the peaks. If Eq. 1 is used in the limit of no diffusion, a broad apparent sedimentation coefficient distribution is obtained. With estimates of the hydrodynamic parameters that lead to a good approximation of the diffusion coefficient, $c(s)$ results in two distinct peaks. Errors in the frictional ratios that produce too low diffusion coefficients in Eq. 4 led to broadening of $c(s)$, whereas errors that produce too large diffusion coefficients (such as the case of the too small value of $\bar{v} = 0.70$ cm³/g shown in Fig. 1 C) led to artificially sharp distributions $c(s)$. A comparison with the established methods shows that the results at $D = 0$ are very similar to those from the time-derivative g*(s) analysis (Stafford, 1992), which produces apparent sedimentation coefficient distributions in the approximation of no diffusion, whereas even moderately precise estimates of the frictional ratios (or diffusion coefficient, respectively) lead to peak sedimentation coefficients consistent with those obtained by the van Holde–Weischet method, which corrects for diffusion broadening of the boundary (Fig. 1 D) (van Holde and Weischet, 1978).

The influence of the regularization parameter on the calculated $c(M)$ in the case of discrete distributions (δ-functions) is that of a broadening of the peaks in $c(M)$ (in case of second-derivative regularization approximately Gaussian shaped), with a half-width that increases with α (Figs. 1 B and 2 C). For the discrete distribution of Fig. 1, increasing the regularization from $p = 0.68$ to 0.95 still allows clear distinction of the two peaks (with a ratio of $c(s)$ height at the peak to the enclosed minimum of ~4:1, data not shown).

To study the effect of regularization for broader, continuous distributions, noisy sedimentation data based on model distributions in different size ranges were simulated. Figure

2, A and B, shows the analysis of a step-function model for a homogeneous size distribution in the molar mass range between 30,000 and 70,000 (Fig. 2 A) and at 1000-fold higher molar masses (Fig. 2 B). Without regularization ($\alpha = 0$), as can be expected, a series of sharp peaks were obtained, which, in their location and height, strongly depend on the noise of the data. Already with a very small degree of regularization (a variance increase of $\Delta\sigma_\alpha/\sigma_0 < 0.1\%$, adjusted to $p = 0.55$), the analysis resulted in continuous distributions. However, they still can exhibit oscillations that mimic a structured, apparently multimodal distribution (this was observed in particular with the second derivative regularization, data not shown). When the regularization parameter was increased to a level corresponding to $p = 0.68$ ($\Delta\sigma_\alpha/\sigma_0 \sim 1\%$) or $p = 0.95$, which selects the most parsimonious of all $c(M)$ distributions that lead to statistically comparable fits of the sedimentation data, in most cases, a relatively unstructured distribution was obtained in which misleading peaks were absent. Further increase of the regularization parameter to a significantly larger value of $\Delta\sigma_\alpha/\sigma_0 = 10\%$ ($p > 0.99$) only slightly worsened the resemblance of the calculated and the underlying model distributions (Fig. 2, *dash-dot lines*). As illustrated in Fig. 2, A and B, the resolution increased slowly with increasing size of the particles. Also, the results were found to improve when studying model distributions with higher degree of smoothness. This is illustrated in Fig. 2 C, where the calculated $c(M)$ distributions are shown for simulated noisy sedimentation data that are based on a size-distribution model combining a Gaussian and a δ-function.

Again, if the distributions are transformed into a $c(s)$ distribution, they can be easily compared with the integral sedimentation coefficient distributions $G(s)$ from the van Holde–Weischet analysis (insets of Fig. 2). The results of both methods were found to be very consistent. However, the distributions $c(s)$ appear to have higher information content in the description of the shape of the distributions: whereas the $G(s)$ curves from the van Holde–Weischet analysis of Figs. 1 D and 2 A are qualitatively similar, the corresponding $c(s)$ profiles resolve the difference between a broad continuous and a discrete bimodal distribution better.

As a first application of the method to discrete mixtures of globular proteins, the interference profiles from sedimentation experiments with myoglobin and gamma globulin were analyzed (Fig. 3, A–C). These data have been published before (Schuck and Demeler, 1999) in the context of demonstrating the validity of the algebraic systematic noise-reduction procedure developed for the analysis of interference optical data. In the previous analysis, known partial specific volumes and molar masses of the proteins had been used as prior knowledge. In the present context, to evaluate the robustness of the size-distribution analysis method, the data were reanalyzed without this prior knowledge, but instead making the assumption of having globular proteins with approximately spherical shapes ($f/f_0 = 1$), and with an
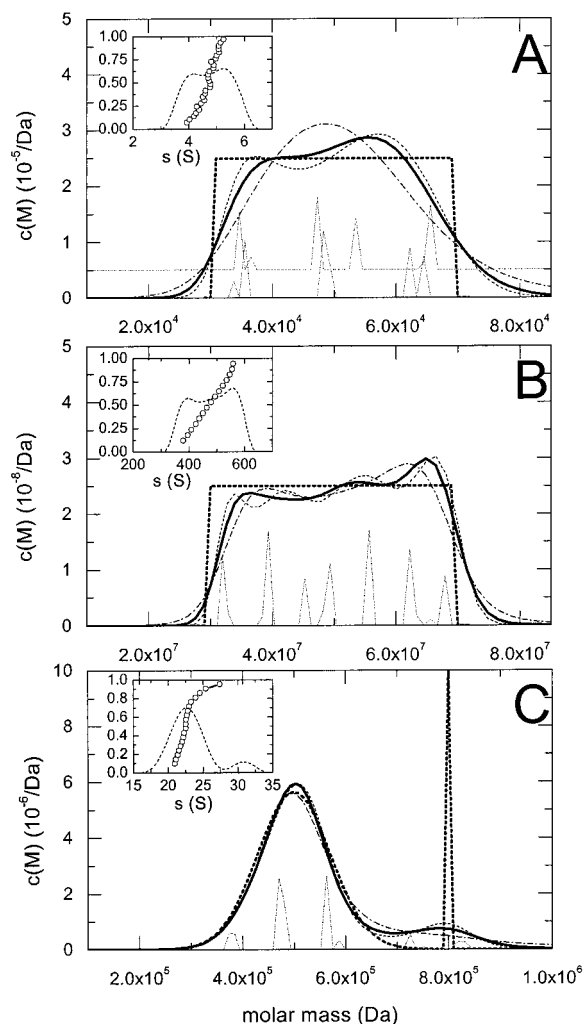
FIGURE 2 Calculated distributions $c(M)$ of simulated sedimentation velocity profiles of continuous model distributions. Sedimentation data were simulated for a solution column of 1-cm length, assuming spherical particles ($f/f_0 = 1$, $\bar{v} \times \rho = 0.73$, $\eta_r = 1$, $T = 20°C$) with the continuous molar mass distribution models (bold dotted line) given by (A) a step function centered at 50 kDa, (B) a step-function centered at 50 MDa, and (C) a Gaussian at 500 kDa combined with a delta-function at 800 kDa, at rotor speeds of 50,000, 5,000, and 30,000 rpm, respectively. The total loading signal was 1, and normally distributed noise of 0.01 was added. Twenty profiles were included in the sedimentation analysis, spaced in time intervals of 500, 300, and 300 sec, respectively, producing sedimentation patterns similar to those in Fig. 1 A. The analysis was performed including a floating baseline parameter, and using the correct hydrodynamic parameters. Data are shown without regularization ($\alpha = 0$, thin dash-dot-dot lines, reduced in scale by a factor of 20), and with maximum entropy with α adjusted to $p = 0.55$ ($\Delta\sigma_\alpha/\sigma_0 < 0.1\%$) (dashed line), α adjusted to $p = 0.68$ ($\Delta\sigma_\alpha/\sigma_0 \sim 1\%$) (bold solid line), and α adjusted to $\Delta\sigma_\alpha/\sigma_0 = 10\%$ ($p > 99.9\%$) (dash-dot line). Panel A shows a second distribution without regularization (offset by 0.5), which is based on replicated simulation, differing only in the details of the normally distributed noise (at the same rms). The insets show the integral sedimentation coefficient distributions $G(s)$ from a van Holde–Weischet analysis, plotted as boundary fraction versus s-value, and, for comparison, a (rescaled) transformation of the calculated size distributions at $p = 0.68$ into $c(s)$ distributions.
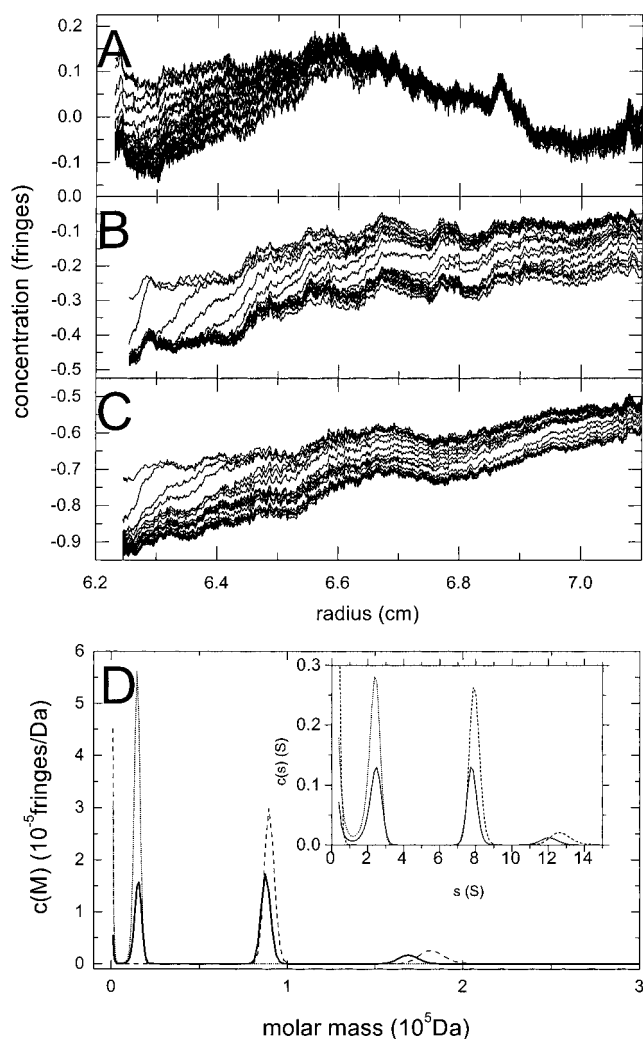
FIGURE 3 Interference fringe patterns of sedimentation velocity experiments with (A) myoglobin, (B) gamma globulin, and (C) a mixture of both ($\omega = 40,000$ rpm, $T = 25°C$, 20 scans in time intervals of 500 sec were analyzed). (D) calculated molar mass distributions $c(M)$ with maximum entropy regularization at $p = 0.68$ (a variance increase of 0.6%), with a resolution of $N = 150$ molar mass values from 1,000 to 500,000, using a constant f/f_0 of 1 and a partial specific volume value of 0.73 cm³/g ($\rho = 1.004$ g/cm³, $\eta_r = 0.9$). The algebraic systematic noise was calculated according to Schuck and Demeler (1999). The analysis resulted in an rms error of 0.0072 fringes for myoglobin (dotted line), 0.0067 fringes for gamma globulin (dashed line), and 0.0049 fringes for the mixture (solid line), respectively. The inset shows the sedimentation coefficient distributions $c(s)$ obtained with α adjusted to a confidence limit of $p = 0.95$.

estimate of the partial specific volume of 0.73 cm³/g. As is shown in Fig. 3 D, the calculated distributions $c(M)$ and $c(s)$ exhibit well-defined, sharp peaks, as can be expected for this discrete mixture of proteins. Because these proteins are not truly spherical, the molar mass values at the peak maxima of $c(M)$ (13,500–15,800 for myoglobin, 87,400–93,700 for gamma globulin monomer, 174,000–190,000 for the dimer) do not coincide with the true molar masses of these species, but instead represent their molar masses ap-

proximately reduced by the frictional ratio ($f/f_0 \sim 1.5$ for the IgG species, based on the earlier results). This problem is absent in the sedimentation coefficient distribution $c(s)$. Both distributions $c(M)$ and $c(s)$ give an excellent fit to the data, and reflect the main features of the samples, i.e., the presence of a small component, and the presence of two larger components with a molar mass ratio of 2:1. When using the Tikhonov–Phillips regularization (Eq. 5b), the resulting distributions suggest the presence of a small amount of aggregates much larger than the IgG dimer (data not shown), but this cannot be resolved well, and is not observed with the maximum entropy regularization. In both methods, an artifact is visible at very small molar masses in the distributions where the sedimentation profiles are correlated with the baseline parameters.

As an example for the application of the method to a continuous mass distribution, the sedimentation velocity profiles of a ferritin sample were studied. Ferritin is well-known to exhibit a broad distribution in the iron content (see, e.g., Leapman and Hunt, 1995). Apoferritin and ferritin do not differ in their sizes, but only in their molar masses and partial specific volumes, depending on the number of iron molecules in the core. As a consequence, the diffusion coefficient should remain constant, and the sedimentation coefficient distribution $s(M)$ according to Eq. 3 can be directly related to the buoyant molar mass distribution $c(M^*)$. Dynamic light-scattering experiments with the ferritin and the apoferritin samples gave autocorrelation functions that were very well described by that of a single species with nearly identical diffusion coefficients of 3.37×10^{-7} and 3.11×10^{-7} cm^2/sec, and hydrodynamic radii of 6.4 and 6.7 nm, respectively. This is consistent with the radius of $\sim$6.5 nm measured for murine ferritin by electron microscopy (Ohkuma et al., 1976). In the analysis of the sedimentation profiles of apoferritin (Fig. 4 A), when constraining the diffusion coefficient to a value of 3.37×10^{-7} cm^2/sec, a reasonable fit was obtained with a sedimentation coefficient $s_{w, 20}$ of 18.9 S, which corresponds to a molar mass of $\sim$540,000 (rms error = 0.0113 OD; a slightly better fit of rms error = 0.0100 OD could be obtained by taking into account free monomers of ferritin). In contrast, the sedimentation velocity profiles of the iron-loaded ferritin could not be well described by the single-species model with the predetermined diffusion coefficient (rms error = 0.0321, s_w = 67.1 S, Fig. 4 B), because the broadening of the sedimentation boundary is much larger than that of a species with $D = 3.37 \times 10^{-7}$ cm^2/sec, indicated by the dashed line in Fig. 4 B. This suggests strong heterogeneity of the ferritin sample.

The calculated buoyant molar mass distributions $c(M^*)$ of apoferritin, ferritin, and a mixture are shown in Fig. 5 A. All result in very good fits of the data, with rms errors of $\sim$0.009 OD. For the apoferritin, the majority of the material is in a single peak with a maximum at a buoyant molar mass of 140,000 (Fig. 5 A, dotted line). The presence of a small
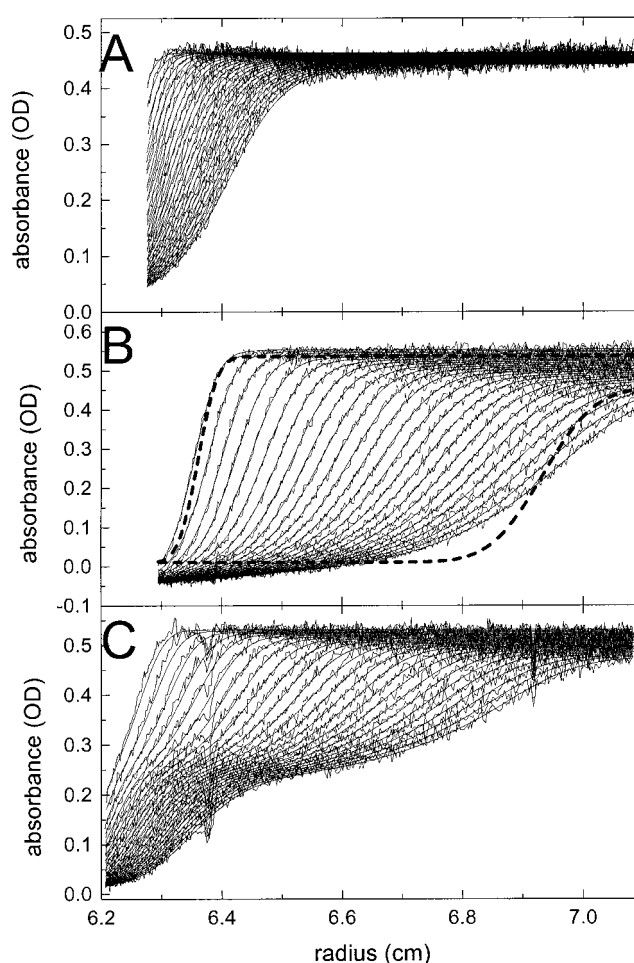

FIGURE 4 Sedimentation velocity absorbance profiles of (*A*) apoferritin, (*B*) ferritin, and (*C*) a mixture, obtained at a rotor speed of 15,000 rpm, rotor temperature of 24°C, scanned at a wavelength of 230 nm. Equivalent subsets with time increments of $\sim$200 sec are shown. The best-fit distributions from the $c(M^*)$ analysis (as described in Fig. 5 A) are superimposed on the experimental data. (*B*) also shows the best-fit distribution to the first and last scan using a single-species sedimentation model with the predetermined diffusion coefficient of 3.37×10^{-7} cm^2/sec (*dashed lines*).

fraction of material of approximately double the size of the main peak is suggested. The $c(M^*)$ distribution of ferritin is characterized by a broad, asymmetric peak with a maximum at a buoyant molar mass of 540,000, but also exhibiting a broad distribution of smaller material, including molecules of the size of apoferritin (Fig. 5 A, dashed line). For the mixture, the clearly bimodal sedimentation profiles of Fig. 4 C translate in the $c(M^*)$ distribution into a bimodal mass distribution, with maxima at buoyant molar mass values of 140,000 and 530,000 (Fig. 5 A, solid line). The features of the ferritin distribution seem to be reasonably well reproduced.

It should be noted that size-distribution of the mixture exhibits a small oscillatory finer structure, which does not appear in the ferritin distribution (Fig. 5 A, dashed and solid
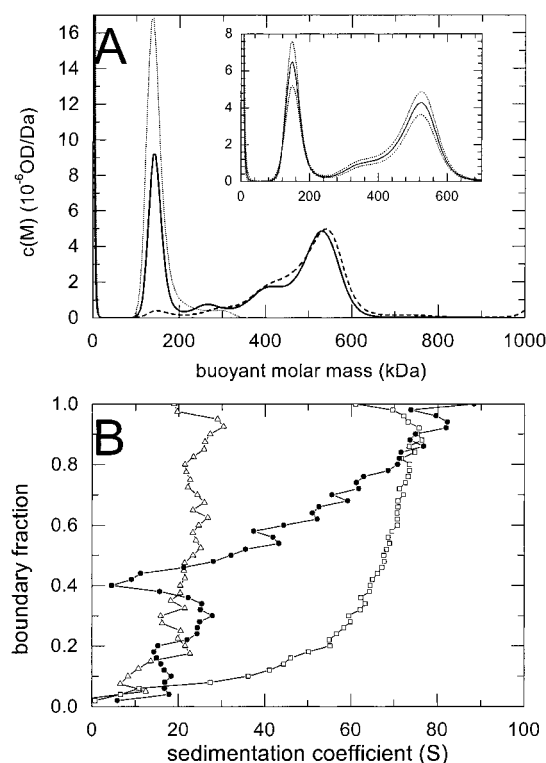
FIGURE 5 Buoyant molar mass distributions of apoferritin (*dotted line*), ferritin (*dashed line*), and the mixture (*solid line*), from the analysis of the data shown in Fig. 4, using the predetermined value of $D = 3.37 \times 10^{-7}$ cm^2/sec, with $N = 150$, maximum entropy α adjusted to $p = 0.95$. For comparison, the data of the mixture are scaled by a factor of 1.4. The inset shows the result of a Monte Carlo simulation (10^3 replicates) based on the best-fit calculated sedimentation data and rms error from the analysis of the apoferritin/ferritin mixture. For each molar mass value, the mean (M), the 5%, and the 95% levels of the set of $c(M)$ data obtained are shown. (*B*) Van Holde–Weischet analysis of appropriate data subsets of the same experiments with apoferritin ($\triangle$), ferritin ($\square$) and the mixture ($\bullet$).

line). To study whether these oscillations are essential features of the data, and how sensitive they are to the noise in the raw data, we performed Monte Carlo simulations. Simulated data sets were replicated ($n = 10^3$) based on the calculated best-fit sedimentation profiles as shown in Fig. 4 *C*, with normally distributed noise added in the magnitude of the rms error of the fit. The inset in Fig. 5 *A* shows the mean distribution $c(M^*)$ and the 5% and 95% contours, respectively. In this statistical average, $c(M^*)$ appears slightly smoother, which demonstrates that some of the oscillatory fine structure in $c(M^*)$ can be governed by noise in the data, and may not be features of the true underlying particle size distribution. Nevertheless, comparing the distributions obtained from the Monte Carlo analysis and the results from van Holde–Weischet analysis, although they are qualitatively consistent, it appears that a higher level of detail can be extracted from the Lamm equation model.

A basic assumption of the distribution analysis is that the observed sedimentation data are a simple superposition of

the sedimentation profiles of noninteracting macromolecules (Eq. 1). However, because of the practical importance of this case for the study of proteins, the results obtained when applied to a system of interacting species was investigated. The sedimentation process was simulated for a rapid monomer–dimer and monomer–trimer self-association, using the Lamm equation methods described in Cox, (1969) and Schuck (1998), with 1% normally distributed noise. The conditions of the sedimentation were chosen to generate profiles generally similar to those in Fig. 1 *A*; as can be expected for these systems, no separation of sedimentation boundaries was achieved (data not shown). The sedimentation profiles of these self-associating systems could be fitted very well by the continuous mass distribution (data not shown), which, in the absence of regularization, resulted in a large number of discrete peaks in $c(M)$ (see, e.g., the dotted line in Fig. 6 *C*). But, in contrast to discrete mass distributions of noninteracting species, small regularization at a level $p = 0.68$ already led to very broad, smooth distributions. This result was qualitatively independent on the regularization procedure (data not shown). Compared to the relatively sharp distributions obtained from a superposition of noninteracting species under identical conditions (Fig. 6, *dashed lines*), the spreading of the sedimentation boundary that is caused by the rapid self-association results in an apparent population of macromolecules with a broad range of intermediate sizes (Fig. 6, *bold lines*), with the positions of the maxima dependent on the loading concentration and the association constant. This is analogous to the results from the van Holde–Weischet analysis, where the case of interacting and noninteracting monomers and dimers can also be clearly distinguished from the positive slope and the range of sedimentation coefficients in $G(s)$ (Fig. 6 *B*).

DISCUSSION

The present paper describes a method for direct boundary modeling for the size-distribution analysis in sedimentation velocity analytical ultracentrifugation. Because the continuous size distributions are approximated by a superposition of Lamm equation solutions, the effects of diffusion can be taken into account, and a relative high resolution can be achieved for small molecules in the size range of proteins.

Although unraveling of diffusion effects in this way was found to be similarly effective as the extrapolation to infinite time in the van Holde–Weischet method (van Holde and Weischet, 1978), the direct boundary modeling can offer several advantages. First, because the Lamm equation method can take into account the end effects of the solution column, there is no requirement for a solvent and solution plateau to be established, which allows the analysis of the data from an entire sedimentation experiment. This ability to make maximal use of the information of the boundary spreading observed over a large time period enhances the ability for distinguishing boundary spreading due to size
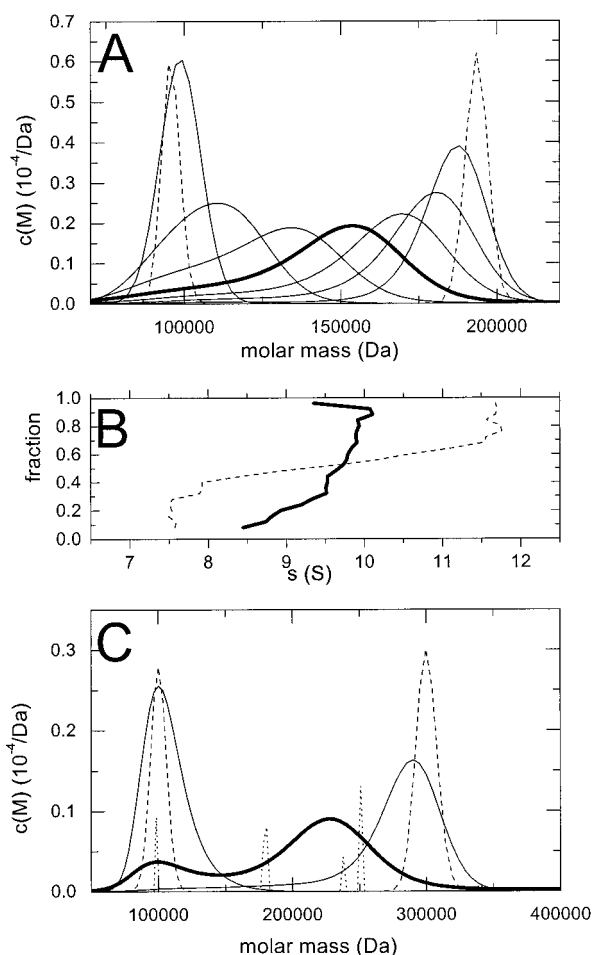
FIGURE 6 Analysis of simulated sedimentation profiles of discrete components in rapid monomer–dimer (*A* and *B*) and monomer–trimer (*C*) self-association equilibrium. (*A*) Simulated monomer–dimer data based on a monomer molar mass of 100,000, $f/f_0 = 1.0$, and $\bar{v} = 0.73$ cm^3/g. Sedimentation profiles were generated at a total loading concentration of 1, with 0.01 normally distributed noise, and with dimerization constants leading to initial monomer/dimer ratios of 26, 5.9, 2.2, 1, 0.5, 0.25, and 0.11, respectively (*solid lines*). Shown are the calculated mass distributions $c(M)$ with $N = 100$, maximum entropy regularization with α adjusted to $p = 0.68$. The distributions obtained at equal loading concentrations of monomer and oligomer are highlighted (*bold lines*), and, for comparison, the distributions from the sedimentation profiles of a noninteracting mixture of species at equal concentrations are shown (*dashed lines*). (*B*) The same data analyzed by van Holde–Weischet analysis. The integral sedimentation coefficient distribution $G(s)$ is given for a mixture of monomer and dimer at equal loading concentrations in rapid self-association equilibrium (*bold line*) and noninteracting (*dashed line*). (*C*) Monomer–trimer system, with a monomer molar mass of 100,000, $f/f_0 = 1.3$, and $\bar{v} = 0.73$ cm^3/g. Simulations were performed with an association constant for trimer formation of 4, and total concentrations of 0.1, 1, and 10, respectively, each with 1% normally distributed noise (each $c(M)$ was scaled to a total loading concentration of 1). The distribution at a total concentration of 1 is also shown without regularization (*dotted line*).

heterogeneity from simple diffusional spreading. This, combined with better statistical properties of a direct fit, seems to be the origin of the higher level of detail in the $c(M)$ as

compared to the $G(s)$ curves. The new method can also be applied to experiments of mixtures that include small and rapidly diffusing material, or samples with a very high degree of heterogeneity. Second, as an explicit boundary model, the method can use the algebraic noise decomposition techniques (Schuck and Demeler, 1999), and be directly applied to interference optical data where a significant systematic time-invariant background profile can be superimposed to the macromolecular sedimentation profiles. Third, the analysis also lends itself to be extended to a global fit of several experiments, and allows to incorporate knowledge on the distribution into the analysis.

The method presented here could be considered intermediate between a more conventional direct boundary fitting method that uses an explicit single- (or few-) component Lamm equation model (Demeler and Saber, 1998; Philo, 1997; Schuck, 1998; Schuck et al., 1998), and a relatively model-free data transformation, such as the van Holde–Weischet method to obtain $G(s)$ (Demeler et al., 1997; van Holde and Weischet, 1978), or the dc/dr (Baldwin and Williams, 1950; Bridgman, 1942; Fujita, 1962; Signer and Gross, 1934; Svedberg and Pedersen, 1940) and dc/dt (Stafford, 1992) transformations used to obtain $g^*(s)$. The size-distribution analysis proposed here is model-free in a sense that it imposes virtually no constraints on the number and size of the species present. However, in contrast to the data transformations involved in the van Holde–Weischet method and in the $g^*(s)$ methods, it requires prior knowledge on the approximate density and shape of the molecules, and the density and viscosity of the solvent. When available, this knowledge can be used to enhance the resolution of the sedimentation coefficient distribution and transform it into a size distribution $c(M)$.

The relationship and the resolution of the different methods can be understood by considering different degrees of diffusion incorporated into the Lamm equation model (Fig. 1, *C* and *D*). In the absence of any diffusion, as can be expected, the distribution $c(s)$ resembles an apparent sedimentation coefficient distribution $g^*(s)$. Even moderately precise estimates of the hydrodynamic shape and relatively low estimates of the diffusion coefficient leads to a substantial increase in resolution of $c(s)$, which then defines a range of sedimentation coefficients of the sample consistent and comparable with $G(s)$. It is important to note that the van Holde–Weischet method is very powerful in indicating the range of the true sedimentation coefficients of the sample, without further assumptions. If prior knowledge can be used, however, $c(s)$ seems to have a higher resolution. This is indicated by a comparison of the $G(s)$ from the bimodal discrete distribution of Fig. 1 *D* and the broader distribution of Fig. 2 *A*, where qualitatively very similar $G(s)$ distributions were obtained, whereas the corresponding $c(s)$ could distinguish the distributions better. Also, the comparison of the $c(M^*)$ and the $G(s)$ distributions of the ferritin experiment (Fig. 5) indicates slightly higher information content

of the Lamm equation analysis. However, because of the well-known tendency of the inversion of integral equations to produce oscillations, some of the details in $c(M)$ can be deceptive. This is illustrated by the Monte Carlo simulations in Fig. 5, and represents a major technical difficulty with the presented approach.

The underlying problem is that Eq. 1 is an ill-posed problem for smooth kernels (Phillips, 1962). This has been extensively studied (Amato and Hughes, 1991; Hansen, 1992; Phillips, 1962), and is well known to occur in many biophysical techniques, for example, in dynamic light scattering (Provencher, 1979). Because, in a direct inversion, the size-distribution analysis in Eq. 1 tends to produce large oscillations in $c(M)$, the analysis requires regularization and adjustment to the level of detail that one can reliably extract from the experimental data. The approach used here closely followed the technique of adjusting the regularization parameter by controlling the variance increase of the fit that is introduced by the regularization constraint, a technique developed by Provencher and implemented in the program CONTIN (Provencher, 1982b). As regularization methods, maximum entropy regularization and Tikhonov–Phillips regularization with a second derivative operator were studied. Maximum entropy performed slightly better, because it could create sharper peaks for discrete size distributions and had a somewhat lower tendency to exhibit oscillations for broader distributions. Overall, the results are consistent with the previous findings from numerical simulations (Amato and Hughes, 1991) and from studies of broadly distributed biopolymers by light scattering (Provencher, 1992). Alternative numerical methods to avoid artificial peaks, such as described by Provencher (1992) could be adapted.

If one compares the physical processes observed for particle-size analysis in sedimentation velocity ultracentrifugation with those of dynamic light scattering, centrifugation has a strongly size-dependent directed migration in the centrifugal field in addition to the diffusion. Therefore, it appears that this additional source of information in centrifugal data should make the choice of the regularization procedure less critical. In both the experimental and the simulated data with continuous distributions, it was found that it is advantageous to slightly increase the regularization parameter to suppress artificial oscillations. This may be due to the inactivity of the non-negativity constraints in the case of broad distributions. Clearly, the adjustment of the regularization parameter to obtain the optimal degree of information will require experience, and may depend on knowledge of the type of sample under study. This problem of ill-posed analysis can also be minimized experimentally at high rotor speeds and long solution columns, where the diffusional spreading of the sedimentation boundaries is smallest, and their translation is largest.

Another difficulty in the presented approach is the requirement for estimates of the partial specific volume and the hydrodynamic frictional ratio, which basically provides the information on how much boundary spreading one would expect for a discrete single species of any given size. Fortunately, there are many alternative ways of obtaining such hydrodynamic prior knowledge, and many cases in which good approximations can be made (see above) (Frigon and Timasheff, 1975). Although the quantitative aspects of the calculated distributions depend on this prior knowledge, the method was found robust in reporting the essential qualitative features of the distributions even when using slightly incorrect size and shape assumptions (Fig. 3 D). Furthermore, if the distribution is transformed to a sedimentation coefficient distribution $c(s)$, only the resolution of the distribution was found to suffer slightly from errors in f/f_0 and $\bar{v}$. (In $c(M)$, these errors additionally influence the location of the peaks, which will depend linearly on f/f_0, and approximately with a ⅔ power dependence on $\bar{v}$.) However, even with relatively gross estimates of the hydrodynamic parameters, the sedimentation coefficient distribution $c(s)$ can have a higher resolution than a sedimentation coefficient distribution that is uncorrected for diffusion (Figs. 1, C and D, and 3 D).

The utility of the method, if applied to protein mixtures, is illustrated in Fig. 3, where all three species, including a small component, can be clearly resolved. Beyond the calculation of molar mass estimates for species in discrete protein mixtures, which could be useful for diagnostics of protein homogeneity and aggregation, the method may also allow the diagnosis of the presence and stoichiometry of fast reversible interactions. Similar to the broadening of the $G(s)$ distributions obtained from van Holde–Weischet analysis of a reaction boundary, very broad apparent $c(M)$ and $c(s)$ distributions were obtained for interacting systems (Fig. 6). In both methods, this broadening can be observed without macroscopic separation of the species, such as in monomer–dimer and monomer–trimer self-associations (Gilbert, 1955; Gilbert and Gilbert, 1973), which is a reflection of the boundary spreading in a reaction boundary being different from that of noninteracting species. The possibility of a detailed quantitative analysis of these apparent $c(M)$ curves beyond their diagnostic utility is unclear. However, once the stoichiometry of an interaction has been identified, more direct modeling with the Lamm equation expanded by the reaction terms seems advantageous and practical (see, for example, Claverie, 1976; Cox, 1969; Frigon and Timasheff, 1975; Goad and Cann, 1969; Schuck, 1998; Stafford, 1998).

The characterization of broad continuous distributions is an important application, and a ferritin sample was used as a model system. In conventional sedimentation velocity analysis apoferritin showed a molar mass that is slightly high compared to the previously determined value of 460,000 (Stefanini et al., 1982), this could be due to a small fraction of aggregates, such as those observed by flow field flow fractionation in Pauck and Cölfen (1998). Interestingly, in the size-distribution analysis of the apoferritin, a small peak in the size range of double the main peak was

found. In the size-distributions of the ferritin sample, a small degree of artificial oscillations in $c(M^*)$ was identified by Monte Carlo simulation. These problems could be characteristic of the study of broadly distributed macromolecules. However, the gross shape of $c(M^*)$ was reproducible and appears reasonable. The difference in the buoyant mass between apoferritin and ferritin corresponds to ~4000–5000 iron atoms per molecule, consistent with values for maximal loading found in Leapman and Hunt (1995). Further studies are in progress to explore the properties of the new method in more detail when applied to the characterization of broad distributions, such as emulsions (M. Perugini, P. Schuck, and G. Howlett, in preparation) and polystyrene.

In summary, it was shown that an explicit Lamm equation model for continuous size-distribution analysis can be formulated and applied. Equipped with maximum entropy regularization, it reveals relatively high-resolution distributions, which were tested with simulated and experimental data of both discrete and broad continuous distributions. Although the results obtained are qualitatively comparable to those from van Holde–Weischet analysis, providing a higher level of detail, the requirements and properties are much different from those of the van Holde–Weischet method and the dc/dt- or dc/dr-based $g^*(s)$ method. This suggests that it could be a useful new tool for ultracentrifugal studies. Furthermore, the described method can be extended to global analyses, for example, of several experiments at different rotor speeds, and can be adapted to different experimental protocols such as gravitational-sweep sedimentation (Mächtle, 1999). Also, the explicit Lamm equation model for continuous distributions could be used to combine the analysis of sedimentation velocity and dynamic light-scattering experiments, which may provide a significant increase in resolution, and possibly additional information on the hydrodynamic shape of the macromolecules as a function of their size.

REFERENCES

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

Baldwin, R. L., and J. W. Williams. 1950. Boundary spreading in sedimentation velocity experiments. *J. 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.

Bloomfield, V., W. O. Dalton, and V. H. K. E. 1967. Frictional coefficients of multisubunit structures. *Biopolymers.* 5:135–148.

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

Cann, J. R., and G. Kegeles. 1974. Theory of sedimentation for kinetically controlled dimerization reactions. *Biochemistry.* 13:1868–1874.

Claverie, J.-M. 1976. Sedimentation of generalized systems of interacting particles. III. Concentration-dependent sedimentation and extension to other transport methods. *Biopolymers.* 15:843–857.

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.

Cox, D. J. 1969. Computer simulation of sedimentation in the ultracentrifuge. IV. Velocity sedimentation of self-associating solutes. *Arch. Biochem. Biophys.* 129:06–123.

Cox, D. J., and R. S. Dale. 1981. Simulation of transport experiments for interacting systems. *In* Protein–Protein Interactions. C. Frieden, and L. W. Nichol (eds.), Wiley, New York.

Crank, J., and P. Nicholson. 1947. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Proc. Cambridge Philos. Soc.* 43:50–67.

de la Torre, J. G. 1992. Sedimentation coefficients of complex biological particles. *In* Analytical Ultracentrifugation in Biochemistry and Polymer Science, S. E. Harding, A. J. Rowe, and J. C. Horton (eds.), Royal Society of Chemistry, Cambridge, U.K. 333–358.

Demeler, B., and H. Saber. 1998. Determination of molecular parameters by fitting sedimentation data to finite element solutions of the Lamm equation. *Biophys. J.* 74:444–454.

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

Dishon, M., G. H. Weiss, and D. A. Yphantis. 1966. Numerical solutions of the Lamm equation. I. Numerical procedure. *Biopolymers.* 4:449–455.

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

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

Giebeler, R. 1992. The Optima XL-A: a new analytical ultracentrifuge with a novel precision absorption optical system. *In* Analytical Ultracentrifugation in Biochemistry and Polymer Science. S. E. Harding, A. J. Rowe, and J. C. Horton (eds.), The Royal Society of Chemistry, Cambridge, U.K. 16–25.

Gilbert, G. A. 1955. *Disc. Faraday Soc.* 20:68–71.

Gilbert, L. M., and G. A. Gilbert. 1973. Sedimentation velocity measurement of protein association. *Methods Enzymol.* 27:273–296.

Goad, W. B., and J. R. Cann. 1969. Theory of sedimentation of interacting systems. *Ann. N.Y. Acad. Sci.* 164:172–182.

Hanlon, S., K. Lamers, G. Lauterbach, R. Johnson, and H. K. Schachman. 1962. Ultracentrifuge studies with absorption optics. I. An automatic photoelectric scanning absorption system. *Arch. Biochem. Biophys.* 99:157–174.

Hansen, P. C. 1992. Numerical tools for analysis and solution of Fredholm integral equations of the first kind. *Inverse Prob.* 8:849–872.

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 (eds.), Birkhäuser, Boston, MA. 37–65.

Lamm, O. 1929. Die Differentialgleichung der Ultrazentrifugierung. *Ark. Mat. Astr. Fys.* 21B:1–4.

Laue, T. M. 1994. An on-line interferometer for the XL-A ultracentrifuge. *Prog. Coll. Polym. Sci.* 94:74–81.

Laue, T. M., A. L. Anderson, and B. W. Weber. 1997. Prototype fluorimeter for the XLA/XLI analytical ultracentrifuge. *In* Ultrasensitive Biochemical Diagnostics. SPIE Proceedings, G. Cohn, and S. Soper (eds.), SPIE, Bellingham, WA. 196–204.

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 (eds.), The Royal Society of Chemistry, Cambridge, U.K. 90–125.

Lawson, C. L., and R. J. Hanson. 1974. Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, New Jersey.

Leapman, R. D., and J. A. Hunt. 1995. Compositional imaging with electron energy loss spectroscopy. *JMSA.* 1:93–108.

Lechner, M. D., and W. Mächtle. 1992. Modern methods for determining the molar mass distribution of polymers. General considerations and application to sedimentation equilibrium. *Makromol. Chem, Makromol. Symp.* 61:165–175.

Mächtle, W. 1999. High-resolution, submicron particle size distribution analysis using gravitational-sweep sedimentation. *Biophys. J.* 76: 1080–1091.

Ohkuma, S., H. Noguchi, F. Amano, D. Mizuno, and T. Yasuda. 1976. Synthesis of apoferritin in mouse peritoneal macrophages. Characterization of 20 S particles. *J. Biochem. (Tokyo).* 79:1365–1376.

Pauck, T., and H. Cölfen. 1998. Hydrodynamic analysis of macromolecular conformation. A comparative study of flow field flow fractionation and analytical ultracentrifugation. *Anal. Chem.* 70:3886–3891.

Phillips, D. L. 1962. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.* 9:84–97.

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

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

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 (eds.), The Royal Society of Chemistry, Cambridge, U.K. 92–111.

Rivas, G., W. Stafford, and A. P. Minton. 1999. Characterization of heterologous protein–protein interactions via analytical ultracentrifugation. *Methods: A Companion to Methods Enzymol.* 19:194–212.

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

Schachman, H. K., L. Gropper, S. Hanlon, and F. Putney. 1962. Ultracentrifuge studies with absorption optics. II. Incorporation of a monochromator and its application to the study of proteins and interacting systems. *Arch. Biochem. Biophys.* 99:175–190.

Schmidt, B., and D. Riesner. 1992. A fluorescence detection system for the analytical ultracentrifuge and its application to proteins, nucleic acids, viroids and viruses. *In* Analytical Ultracentrifugation in Biochemistry

and Polymer Science. S. E. Harding, A. J. Rowe, and J. C. Horton (eds.), The Royal Society of Chemistry, Cambridge, U.K. 176–207.

Scholte, T. G. 1968. Molecular weights and molecular weight distribution of polymers by equilibrium ultracentrifugation. Part II. Molecular weight distribution. *J. Polymer 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. 1999. Sedimentation equilibrium analysis of interference optical data by systematic noise decomposition. *Anal. Biochem.* 272:199–208.

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.

Schuster, T. M., and J. M. Toedt. 1996. New revolutions in the evolution of analytical ultracentrifugation. *Curr. Opin. Struct. Biol.* 6:650–658.

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

Smith, C. R., and J. W. T. Grandy. 1985. Maximum-Entropy and Bayesian Methods in Inverse Problems. Reidel, Dordrecht, The Netherlands.

Stafford, W. F. 1992. Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Anal. Biochem.* 203:1–7.

Stafford, W. F. 1997. Sedimentation velocity spins a new weave for an old fabric. *Curr. Opin. Biotech.* 8:14–24.

Stafford, W. F. 1998. Time difference sedimentation velocity analysis of rapidly reversible interacting systems: determination of equilibrium constants by global non-linear curve fitting procedures. *Biophys. J.* 74: A301.

Stefanini, S., E. Chiancone, P. Arosio, A. Finazzi-Agro, and E. Antonini. 1982. Structural heterogeneity and subunit composition of horse ferritins. *Biochemistry.* 21:2293–2299.

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

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.

Yphantis, D. A., J. W. Lary, W. F. Stafford, S. Liu, P. H. Olsen, D. B. Hayes, T. P. Moody, T. M. Ridgeway, D. A. Lyons, and T. M. Laue. 1994. On-line data acquisition for the Raleigh interference optical system of the analytical ultracentrifuge. *In* Modern Analytical Ultracentrifugation. T. M. Schuster and T. M. Laue (eds.), Birkhäuser, Boston, MA. 209–226.