ELSEVIER

Biophysical Chemistry 108 (2004) 231–243

Biophysical Chemistry

www.elsevier.com/locate/bpc

Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants

Walter F. Stafford*, Peter J. Sherwood

Analytical Ultracentrifugation Research Laboratory, Boston Biomedical Research Institute, 64 Grove Street, Watertown, MA 02472, USA

Abstract

Analytical ultracentrifugation (AUC) has played and will continue to play an important role in the investigation of protein–protein, protein–DNA and protein–ligand interactions. A major advantage of AUC over other methods is that it allows the analysis of systems free in solution in nearly any buffer without worry about spurious interactions with a supporting matrix. Large amounts of high-quality data can be acquired in relatively short times. Advances in software for the treatment of AUC data over the last decade have eliminated many of the tedious aspects of AUC data analysis, allowing relatively rapid analysis of complicated systems that were previously unapproachable. A software package called SEDANAL is described that can perform global fits to AUC sedimentation velocity data obtained for both interacting and non-interacting, macromolecular multi-species, multi-component systems, by combining data from multiple runs over a range of sample concentrations and component ratios. Interaction parameters include both forward and reverse rate constants, or equilibrium constants, for each reaction, as well as concentration dependence of both sedimentation and diffusion coefficients. SEDANAL fits to time-difference data to eliminate time-independent systematic errors inherent in AUC data. The SEDANAL software package is based on the use of finite-element numerical solutions of the Lamm equation.

Keywords: Analytical ultracentrifugation; Nonlinear curve fitting; Time-difference data; Heterologous interacting systems; Kinetics; Thermodynamics

1. Introduction

Analytical ultracentrifugation (AUC) has played and will continue to play an important role in the investigation of protein–protein, protein–DNA and protein–ligand interactions. Many of these types of interaction occur in the regulation of metabolic processes. Recombinant DNA techniques have made it possible to produce many proteins in sufficient quantities that they may be studied by classical quantitative biophysical techniques. A major advantage of AUC over many other methods is that it allows the analysis of systems free in solution in nearly any buffer without worry about artifacts caused either by chemical modification or by spurious interactions with a supporting matrix. Large amounts of high-quality data can be acquired

*Corresponding author. Tel.: +1-617-658-7808; fax: +1-617-972-1753.

E-mail address: stafford@bbri.org (W.F. Stafford).

0301-4622/04/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2003.10.028

in relatively short times. Advances over the last decade in software for the treatment AUC data have eliminated many of the tedious aspects of AUC data reduction, allowing relatively rapid analysis of complicated systems that were previously unapproachable. A curve-fitting algorithm is described here that can perform global fits to AUC sedimentation velocity data obtained for both interacting and non-interacting, macromolecular multi-species, multi-component systems. Data from several runs over a range of sample concentrations and component ratios can be combined for global fitting. Interaction parameters, including both forward and reverse rate constants, or equilibrium constants, for each reaction as well as concentration dependence of both sedimentation and diffusion coefficients, can be obtained. The algorithms are embodied in a software package called SEDANAL. SEDANAL is based on the use of finite-element numerical solutions of the Lamm equation first developed by Claverie [1,2]. Fits are carried out on time-difference data to eliminate systematic errors. The first use of these numerical solutions for least-squares fitting to sedimentation velocity data was reported by Todd and Haschemeyer [3]. Since then several researchers have reported development of software packages based on the Todd–Haschemeyer method [4–8].

Mass-action equilibria are treated by solving either the kinetic differential equations or the equilibrium mass-action equations corresponding to each reaction scheme and stoichiometry. The main feature of this approach, which distinguishes it from several other methods, is that the fits are performed on time-difference data. The use of time-difference data to eliminate time-independent systematic errors from the data was first implemented in the time derivative method of analyzing sedimentation velocity data [9]. Since then time-difference treatment of interference data has also been used by Philo's direct boundary fitting method in software called SVEDBERG [10] and again by Stafford in ABCD _ FITTER [6,7]. The process of computing the time-difference data completely removes all time-independent systematic errors in the data, thereby eliminating one of the impediments to the analysis of this type of data. In addition to the algorithms used, this communication describes a software package called SEDANAL that will perform the fits on PCs running the Windows operating system.

Whether one chooses to use equilibrium or velocity methods to study a particular interacting system will depend on several factors. Among those factors are the concentration range over which either significant dissociation or association can be observed and the kinetics of the interactions. In many cases, it will be found useful to analyze any given system by both methods. A few specific points are worth highlighting. In general, somewhat lower concentrations can be observed by velocity methods than that by equilibrium methods thereby extending the observable range over which measurements can be made. Samples that are unstable over the times required by an equilibrium experiment often can be analyzed successfully by velocity methods. A good example of this type of system is the isodesmic self-association of tubulin described in this issue by Sontag et al. [11]. Systems that may take several days to reach thermodynamic equilibrium can be treated as a quasi-static system in a velocity run but may not reach equilibrium in an equilibrium experiment in a practical length of time.

2. Theoretical background

A previous version of SEDANAL, called ABCD _ FITTER, was designed to analyze heterologous pair-wise interactions that can be represented by a reaction of the following type:

$$A + B = C \quad K_{eq} = \frac{k_f}{k_r} = \frac{[C]}{[A][B]} \tag{1}$$

where the equilibrium constant is K_{eq} and k_f and k_r are the forward and reverse rate constants, respectively.

SEDANAL is designed to analyze sedimentation velocity data from interacting systems composed of either single or multiple macromolecular components. In the context of this discussion, we define a component as an electroneutral macromolecule that can be added or removed, at least conceptually, from a solution independently of other molecules. A component may be composed

of several macromolecular species. Species interact through chemical reactions. In general, the number of components is equal to the number of species minus the number of chemical reactions. At constant temperature and pressure, or at constant temperature in an incompressible system, the number of degrees of freedom is equal to the number of components. For any given values of the equilibrium constants, the composition of the solution at equilibrium (i.e. the amount of each of the species) is determined by the total concentration of each component at each position in the centrifuge cell.

It is worthwhile to make the distinction between species and components clear; consider the following examples. A monomer–dimer, self-associating system in chemical equilibrium is an example of a single component, multispecies system. The monomer and dimer are in equilibrium with each other and, therefore, cannot be added to or removed from the solution independently. This is why they cannot be separated on a gel filtration column, for example. As soon as some dimer and monomer become separated, either the monomer re-associates to form more dimer or the dimer dissociates to form more monomer. At any given temperature and pressure, the amount of monomer and dimer present is determined completely by the total concentration of the component and the equilibrium constant for the dimerization reaction. Therefore, a rapidly reversible monomer–dimer system is a one-component system composed of two species with one chemical reaction between them. To extend this one step further, if this was a monomer–dimer–tetramer system, it would be a one-component, three-species system with two chemical reactions between them. It would have one degree of freedom, such that the amounts of monomer, dimer and tetramer at each point in the boundary would be uniquely determined by the equilibrium constants and the total macromolecular concentration at each point.

Now, consider a heterologous interacting system composed of two components, A and B that interact to form a complex, C. Let C interact with another molecule of B to form D. This is a two-component system which is composed of four species that are related by two chemical reactions. The system has two degrees of freedom, the total concentrations of A and B at each point in the boundary; it can be represented by the following equations:

$$A+B=C \quad K_{1,\mathrm{eq}}=\frac{k_{1,\mathrm{f}}}{k_{1,\mathrm{r}}}=\frac{[C]}{[A][B]} \tag{2}$$

$$C+B=D \quad K_{2,\mathrm{eq}}=\frac{k_{2,\mathrm{f}}}{k_{2,\mathrm{r}}}=\frac{[D]}{[C][B]}$$

However, if D were composed of one mole of A and two moles of B but was not in equilibrium with C and B, it would be a third, independent component. This system would have three degrees of freedom, the total concentrations of A, B that participate in the reactions and the total concentration of D. In principal, D could be removed from the system by gel filtration without being reformed from C and B. D might be a covalently cross-linked aggregate, for example. Therefore, non-interacting components can also be added to the fitting model to account for aggregates or other species not participating in the reactions of interest. The software also allows one to include and fit for incompetent species. For example, incompetent B (i.e. some B that is not participating in the reaction), would be a third component. By global fitting over a range of loading concentrations, one can estimate the fraction of incompetent B that is present in the sample because its fraction relative to the total concentration in each cell would be independent of the loading concentration.

3. Sedimentation theory

3.1. Sedimentation transport theory

The Lamm equation, which describes the sedimentation process in differential form, is also known as the continuity equation and is a statement of conservation of mass in the centrifuge cell. For a single component system composed of a single macromolecular species, we have

$$\left(\frac{\partial c}{\partial t}\right)_r=-\nabla J_{\mathrm{total}} \tag{3}$$

which in cylindrical coordinates becomes [12]

$$\left(\frac{\partial c}{\partial t}\right)_r = 2\frac{\partial}{\partial \xi}\left[D\xi\left(\frac{\partial c}{\partial \xi}\right)_t - \omega^2 sc\xi\right] \quad (4)$$

where c is the concentration as a function of t and ξ, $\xi = r^2/2$, t is time, D is the diffusion coefficient and s is the sedimentation coefficient. Concentration dependence of s and D can be incorporated in the following forms to first order in total local concentration, $c(r)$

$$s(c) = \frac{s_0}{(1+K_s c)}; \quad D(c) = \frac{D_0(1+2BM_1 c)}{(1+K_s c)} \quad (5)$$

where $(1+K_s c)$ is the concentration dependence of the frictional coefficient and $(1+2BM_1 c)$ is a factor expressing thermodynamic non-ideality through the second virial coefficient. It should be noted that this treatment of non-ideality has ignored the cross terms usually considered in the more general case. Therefore, the concentration range over which this treatment would apply is limited. However, it was felt that the inclusion of cross terms would introduce too many parameters into the fits, although it could be done. Dividing the concentration dependencies of s and D in this way has been discussed previously [13].

The Lamm equation can be solved numerically, very rapidly, by the finite-element method described by Claverie [1,2]. The finite-element method combined with a nonlinear least-squares fitting procedure was first applied to the analysis of sedimentation velocity data by Todd and Haschemeyer [3]. The performance of the finite-element algorithm has been improved significantly by Peter Schuck, who has introduced several enhancements, the most notable of which are an adjustable time interval and a 'moving hat' moving frame of reference [5,8]. He has also developed a novel technique to fit for the systematic errors inherent in AUC data. These algorithms have been embodied in software called SEDFIT, available from Peter Schuck (http://www.analyticalultracentrifugation.com). In 1998, Stafford introduced software called ABCD _ FITTER for the global analysis of heterologous reversibly-interacting systems [6,7]. As explained below, the approach used by ABCD _ FITTER is fundamentally different from that used by either the Todd–Haschemeyer or the Schuck methods, which fit directly to concentration profiles. In contrast, ABCD _ FITTER fits to concentration time-difference curves to eliminate time-independent systematic errors inherent in the optical systems. An adjustable time interval for the time grid of the Claverie solutions has been incorporated into ABCD _ FITTER to improve performance, as well. ABCD _ FITTER was designed to fit to the system $A+B=C$; $C+B=D$ with or without an accompanying indefinite self-association of component B. Since then, ABCD _ FITTER has been expanded to fit to a wide range of interacting systems composed of several components and multiple species. ABCD _ FITTER has become a subset of the newer program called SEDANAL. A brief description of ABCD _ FITTER can be found in Rivas et al. [7]. SEDANAL is described in detail in this communication.

3.1.1. Use of concentration time-difference curves

Historically, the analysis of data from the analytical ultracentrifuge has been hampered by systematic errors that have been difficult to treat. There are two main types of systematic error that are characteristic of the interference optics on the XL centrifuges. One type is time-invariant noise (TI-noise) composed of variations that are a function only of radius. The other type is time-dependent, but radially-invariant (RI-noise), irregular vertical variations in fringe displacement (jitter) that are a function of time only and affect the entire fringe pattern. The absorbance optics, on the other hand, suffer only from TI-noise, that is usually of smaller magnitude. The TI-noise in the interference optics arises mainly from various components of the optical system, including compression of the cell windows, dirt on the optics, irregularities in the lenses, mirrors, optical flats and the CCD array of the camera. The TI-noise in the absorbance system arises mainly from dirt and oil deposited on the cell windows, lenses and slit assembly. The RI noise, which is peculiar to the interference optics, arises from mechanical and thermal variations caused predominantly by cycling of the Peltier heating and cooling system.

3.1.1.1. Elimination of RI-noise. The RI noise can be nearly completely eliminated by aligning the

fringe patterns in the air–air space centripetal to the meniscus, where no sedimentation is occurring. This is accomplished in SEDANAL by least-squares fitting of all the patterns to each other over a small region in the air–air space: numerically, the fringe deflections over that small region are averaged, and the average is subtracted from the entire fringe pattern for that scan. The TI-noise, on the other hand, is completely eliminated by subtracting pairs of scans to create time-difference scans. The time-difference scans, which contain only stochastic noise, are then fit to corresponding time-difference scans generated by the finite-element method.

3.1.1.2. Elimination of TI-noise. The signal obtained from the centrifuge, call it $S(r, t)$, whether it is fringe displacements, absorbance or fluorescence, is a function of both time and radius. After $S(r, t)$ has been preprocessed to remove optical jitter (RI-noise) and integral fringe shifts, it is composed of the contribution from the true concentration distribution, $C(r, t)$, which is also a function of time and radius, as well as a background optical systematic component, $B(r)$, that is time-independent and a function of radius only (Fig. 1). The signal, $S(r, t)$, also has stochastic noise superimposed on it; we can write

$$S(r,\ t) = \alpha C(r,\ t) + B(r) + \text{noise} \tag{6}$$

where α is the conversion factor between concentration and the signal type, either fringes, absorbance or fluorescence.

We can remove completely the time-independent background component, $B(r)$, of the signal by subtracting any two experimental curves obtained at, say, times t_1 and t_2.

$$\begin{aligned} S(r,\ t_2) &= \alpha C(r,t_2) + B(r) + \text{noise} \\ S(r,\ t_1) &= \alpha C(r,t_1) + B(r) + \text{noise} \\ \text{----------} & \\ \Delta S(r,\ t_1,\ t_2) &= \alpha \Delta C(r,\ t_1,\ t_2) + 0 + \sqrt{2}\text{noise} \end{aligned} \tag{7}$$

The time-difference curves, $\Delta S(r, t_1, t_2)$, are proportional to the concentration difference curves, $\Delta C(r, t_1, t_2)$, but now have no time-independent systematic error. As mentioned above, the irregular, time-dependent, systematic error from the optics (often referred to as jitter) will have been removed at the preprocessing stage. The data to be fitted, $\Delta S(r, t_1, t_2)$, have only stochastic errors making them suitable for least-squares fitting.

SEDANAL fits to the time-difference curves, $\Delta S(r, t_1, t_2)$ for any set of parameter guesses by generating concentration curves, $C(r, t_1)$ and $C(r, t_2)$, corresponding to times t_1 and t_2, using numerical solutions to the Lamm equation and subtracting them to form $\alpha\Delta C(r, t_1, t_2)$. The root-mean-square residual is computed as the triple sum over all the points, scans and cells.

$$\text{rmsd} = \left[\frac{\sum_{k=1}^{L} \left\{ \sum_{j=1}^{M} \left(\sum_{i=1}^{N} (\Delta S_{k,j}(r_i,\ t_j,\ t_{j+M}) - \alpha \Delta C_{k,j}(r_i,\ t_j,\ t_{j+M})[P])^2 \right) \right\}}{LMN} \right]^{1/2} \tag{8}$$

where k is the cell number, L is the number of cells, j is the difference curve index, M is the number of difference curves and $2M$ is the total number of scans, j and $j+M$ are the indices of the scans being subtracted, N is the number of radial points in each difference curve, and i is the radial point index. The value of the rmsd is minimized with respect to the parameters of interest using the simplex-directed search method of Nelder and Mead [14], where $[P]$ is the parameter vector composed of both global and local parameters. See below for a definition of global and local parameters.

3.1.2. Global fitting to several data sets

SEDANAL can perform global fits to multiple data sets. A data set is meant to refer to scans obtained from a single centrifuge cell, or from the same cell obtained at different wavelengths or optical systems. Parameters are divided into either global or local categories. Global parameters, those that apply to all data sets, are the primary physical molecular parameters of interest and include sedimentation coefficient, molecular weight, equilibrium constants, rate constants and non-ideality

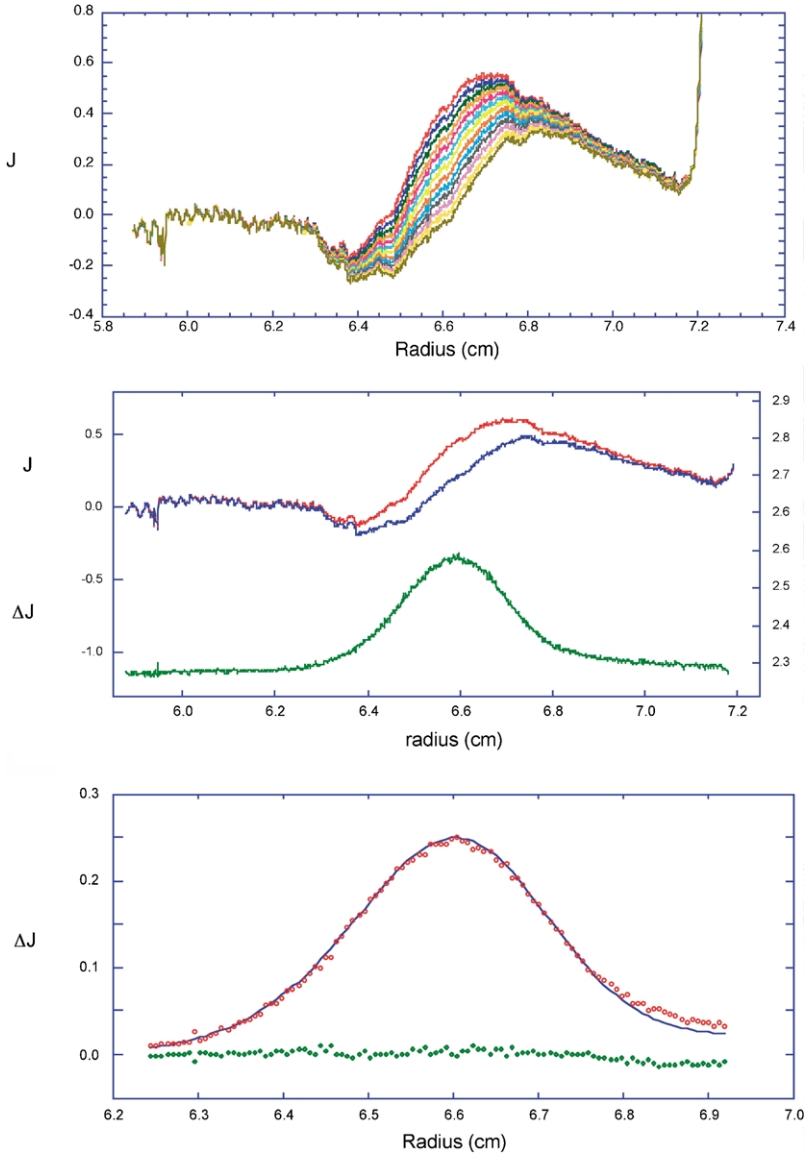

Fig. 1. Schematic representation of the fitting procedure. (top panel) In this example, forty scans are obtained and combined into a single data set (only every other one shown). After a data set has been preprocessed to remove optical jitter by aligning them by least-squares fitting over a small radial span in the air–air region above the meniscus, the scans are subtracted in pairs to produce a set of time-difference curves which are now free of time-independent systematic error. An even number of scans are subtracted pair-wise to give $\Delta S(r(i), t(j), t(j+M)) = S(r(i), t(j)) - S(r(i), t(M+j))$, where $2M$ is the total number of scans. (middle panel) One of the pairs of scans and the resulting difference scan, $\Delta S(r(i), t(j), t(j+M))$. The subtraction process removes the time-independent background component which is a function of radius only. (bottom panel) One of the twenty difference patterns from this fit showing the fitted curve and the residuals. In a typical fitting procedure, as many as 500 difference scans from each cell might be included in the actual fit. The time-difference curves are fit by the nonlinear least-squares procedure as described in the text.

coefficients. The local parameters are the cell loading concentrations, and ratios of components. The ratio of components can be treated as either local or global parameters depending on the experimental procedure. For example, the component ratios would become global parameters if the several data sets represented a dilution series of the same original solution. On the other hand, the component ratios would be treated as local parameters if they were intended, known or suspected to be different for each cell. The current version of SEDANAL can treat up to three components, seven species and six chemical reactions in a global fit to 10 data sets.

3.1.3. Combining data from different optical systems and wavelengths

SEDANAL allows data from several wavelengths and optical systems to be combined in a global fit, and an extinction coefficient to be specified for each species and for each data set. It also allows the determination of extinction coefficients if both interference and absorbance data from a single cell are fit simultaneously, by allowing the value of the extinction coefficient for the absorbance data set to float in the fitting process.

3.1.4. Multiple species fits

When a system is comprised of multiple species, SEDANAL computes in parallel a separate finite-element solution for each species such that

$$\left(\frac{\partial c}{\partial t}\right)_r = \sum_{i=1}^{N}\left(\frac{\partial c_i}{\partial t}\right)_r$$

$$= 2\sum_{i=1}^{N}\frac{\partial}{\partial \xi}\left[D_i\xi\left(\frac{\partial c_i}{\partial \xi}\right)_t - \omega^2 s_i c_i \xi\right] \quad (9)$$

After each time increment, the total concentration of each component at each radial position is computed and the concentrations of all species in equilibrium are recomputed. For kinetically limited (i.e. non-equilibrium) cases the differential equations for the kinetics are solved for that time interval and the new species concentrations are used for the next time increment of the finite-element solution. The kinetic equations are solved by the Bulirsch-Stoer algorithm using Richardson rational polynomial extrapolation (Numerical Recipes in FORTRAN [15]).

3.1.5. Testing with synthetic data

The method has been tested with synthetic data generated by two basically different implementations of the Claverie algorithm. One approach uses the equilibrium mass-action equations and the other solves the kinetic differential equations. SEDANAL has a simulator function that can simulate sedimentation of either rapidly reversible or kinetically limited reversibly interacting systems. Several synthetic model cases with added normally-distributed random noise were tested.

Table 1
Comparison of fits of equilibrium models to kinetically limited data

Model	EQ	EQ, hold S1	EQ, hold S1, M1	Kinetic	Correct values
rmsd	8.46E−03	9.28E−03	1.03E−02	5.71E−03	0.00
M1	4.05E+04	3.73E+04	[4.00E+04]	3.98E+04	4.00E+04
Co1	2.52E−05	2.74E−05	2.54E−05	2.58E−05	2.50E−05
Co2	7.22E−06	7.87E−06	7.38E−06	7.54E−06	7.50E−06
Co3	2.31E−06	2.57E−06	2.43E−06	2.52E−06	2.50E−06
Co4	6.42E−07	7.53E−07	7.16E−07	7.57E−07	7.50E−07
K_{eq} or k_f	2.51E+05	1.17E+06	8.59E+05	1.03E+02	1.00E+02
k_r				1.01E−04	1.00E−04
S1	4.75E−13	[3.60E−13]	[3.60E−13]	3.57E−13	3.60E−13
S2	5.81E−13	5.82E−13	5.87E−13	5.80E−13	5.80E−13

Notes: Reaction simulated was kinetically limited monomer–dimer, $2A = A_2$, with an equilibrium constant of 1×10^6 M^{-1}. The residual plots are shown in Fig. 2. Noise added to the simulated data was 0.005 fringes.

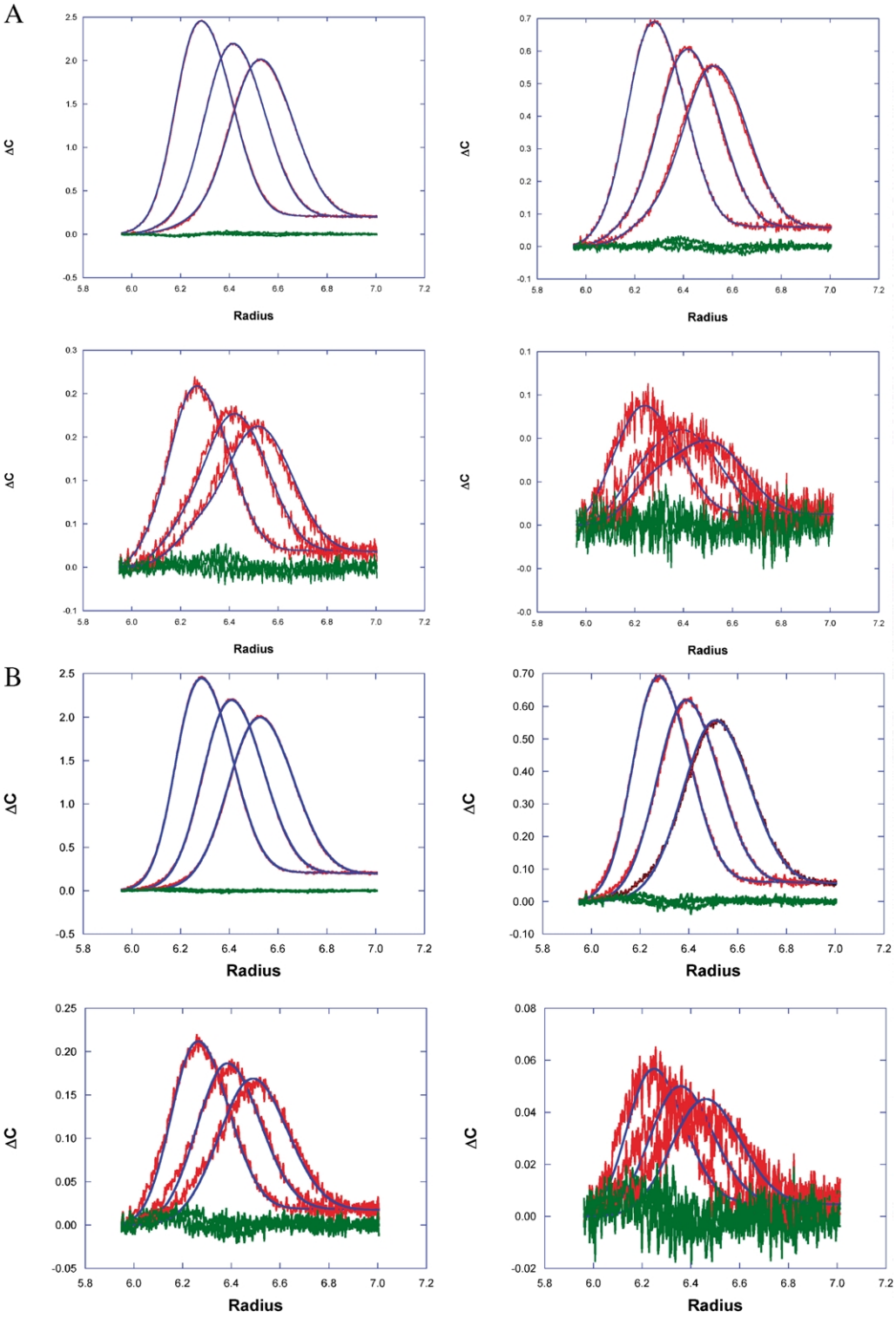

Fig. 2.

Both rapidly reversible and kinetically-limited systems were simulated. One of the questions to be answered is when can the difference be detected and what kinds of errors are to be expected if one were to fit a kinetically limited system with the rapidly reversible model. We address this below.

3.1.5.1. Model 1. To test the basic finite-element algorithm, a single-species system was generated and fitted and showed that the basic algorithm worked as expected. Results are not shown.

3.1.5.2. Model 2. To test the kinetic algorithms, synthetic data for a simple monomer–dimer self-associating system was generated and fitted. This algorithm, for a case in which $k_r = 0.01\ s^{-1}$, was compared to one using the analytical solution for instantaneous equilibration. Essentially identical results were obtained with fits to both sets of synthetic data, demonstrating the kinetic and equilibrium solutions to the mass-action equations gave the same results in the limits at very fast rates. When the reverse rate constant becomes less than approximately 0.001 s^{-1}, but greater than approximately $1 \times e^{-4}\ s^{-1}$, this system could be fit with the reversible case but with unreasonable values for the returned parameters (Table 1). For reverse rate constants less than approximately $1 \times 10^{-4}\ s^{-1}$, the fits are not good as judged by obviously systematic residuals and relatively high root mean squared deviations (Fig. 2).

$$2A = A_2 \quad K_{eq} = \frac{k_f}{k_r} = \frac{[A]}{[A]^2} \tag{10}$$

3.1.5.3. Model 3. Simple 1:1 heterologous complex formation (Eq. (1)) was simulated and fitted.

3.1.5.4. Model 4. To test more complicated, hetero-associations, a two-step 'antigen–antibody' type interaction was simulated and fitted. This model is considered in somewhat more detail below (Table 2), see Eq. (2). Consider the case in which the intrinsic association constants are the same for all steps such that $K_{1,eq} = 2K_{int} = 4K_{2,eq}$, where $K_{1,eq}$ and $K_{2,eq}$ are the macroscopic equilibrium constants. Given a value of K_{int}, this system is completely determined by the loading concentration of component A and the ratio of loading concentrations of components B to A, R_{BA}. For this discussion, we will assume that $R_{BA} = 2.0$. The relative composition of all systems of this type can be further parameterized in terms of the single dimensionless variable $\Theta = K_{int}[A]$.

For various values of Θ, this system was fit at several signal-to-noise ratios. The results are summarized in Table 2. The values given in the table are the averages of the parameters from the abbreviated Monte Carlo analysis for 10 fits to a global fit to six cell loading concentrations.

Several kinetically limited cases were also fitted. When the reverse rate constants become smaller than approximately $10^{-3}\ s^{-1}$, it becomes necessary to explicitly account for the kinetics. As mentioned above, one would like to know how slow the kinetics can be before the equilibrium model will fail to fit the data reliably. To shed some light on that question, a monomer–dimer model with an equilibrium constant of 1×10^6 and a reverse rate constant of $1 \times 10^{-4}\ s^{-1}$ was simulated with 0.005 fringes typical Gaussian noise added. Four loading concentrations were combined in a global fit (Fig. 2). Fits were preformed using the correct model and three variations of the incorrect equilibrium model (Table 1). Internally SEDANAL uses analytical solutions to the monomer–dimer case for both the rapidly reversible and the kinetically limited cases.

3.1.6. Standard deviation of parameters

Parameter standard deviations can be estimated either by Monte Carlo analysis or by bootstrap with replacement. In Monte Carlo analysis, one

Fig. 2. Comparison of fits to a kinetically limited monomer–dimer system as described in the text and Table 1. The system was simulated with an equilibrium constant of $1 \times 10^6\ M^{-1}$ and a reverse rate constant of $1 \times 10^{-4}\ s^{-1}$ In both parts A and B the loading concentrations, clockwise from the upper left, are 25.0, 7.5, 2.5 and 0.75 μM, respectively. (A) Fit to the correct model allowing for kinetically slow re-equilibration (B) fit to an incorrect model assuming instantaneous re-equilibration. Red curves are time-difference plots from the simulated data. Black curves are the best-fit model, and the green curves are the residuals.

generates a synthetic data set using the parameters obtained with the best fit. Multiple sets of random noise are added to the synthetic data to create as many synthetic data sets, and then they are each fitted to obtain a distribution for each parameter. One can compute the confidence intervals, which may be asymmetric, from the parameter distribution function. The other method, called bootstrap with replacement, involves refitting subsets of the data by randomly selecting points from the original data set so that each new data set has the same number of points as the original. The random selection is done with replacement so that some points may be included more than once. To perform a complete analysis of either type, one might have to perform 500–1000 fits of the synthetic data to obtain a complete parameter distribution function. The amount of CPU time required by a program like SEDANAL would make either of these approaches impractical. Therefore, an abbreviated version of either technique can be used. By relying on the Central Limit Theorem [16], which states that distribution of the mean of any parameter will be normally distributed no matter what the shape of the underlying distribution of that parameter, and therefore, its standard deviation can be meaningfully computed. In this approach, the standard deviation of the estimated parameters from 10 to 20 fits of simulated data sets with different noise sets is computed. The standard deviation cannot be translated into confidence limits but gives a good idea of how well a particular parameter is determined. Full Monte Carlo or bootstrap analyses may be done when sufficiently fast computers become available. The current version of SEDANAL (v3.21) can be used to perform both Monte Carlo and Bootstrap with replacement error analyses.

As an example of the abbreviated Monte Carlo analysis of parameter errors is given in Table 2. Two interacting systems were studied. Both systems were of the two-step hetero-interacting type, $A+B=C$; $C+B=D$, with the same molecular parameters in different concentration ranges relative to the equilibrium constants; i.e. different values of θ. The purpose of this analysis is to show the benefits and increased reliability of parameter estimation by globally fitting to multiple data sets spanning several concentrations and component ratios. In the first case, six cells were fit with a range of three loading concentrations for a stoichiometric ratio of two moles of B to one mole of A, along with three variations in which the ratio of B to A, R, was varied from 4:1 to 8:1 to 1:2. For comparison, the case of $R=2{:}1$ at a loading concentration equal to the intrinsic dissociation constant, i.e. $\theta=1.0$, was fit as a single data set. The second case is for lower signal-to-noise ratio, i.e. lower concentration, also at $\theta=1.0$, see Table 2. It is clear from an examination of the standard deviations of the fitted parameters that the parameters are very well determined in both cases in which global fitting of six cells was used. In contrast, the standard deviations of all the fitted parameters were considerably larger for the two cases in which just one data set was used. This is not a surprising result, but emphasizes the importance of combining data and fitting them globally.

The analysis shown in Table 2 also demonstrates that the abbreviated Monte Carlo method of determining parameter standard deviations is useful and that the global fitting to multiple data sets results in a significant improvement in the determination of the fitted parameters. In this fit (Table 2) all parameters were allowed to float to demonstrate the robustness of the global fits. Usually, one would have characterized both A and B in separate experiments to determine their molar masses and sedimentation coefficients and non-ideality coefficients. Prior knowledge of those parameters will leave just the sedimentation coefficients of the complexes and the equilibrium constants as the remaining global variables. The loading concentrations of A and B, which are local variables, should be floated to account for pipetting and other experimental errors in their determination. The fits in this case would be even more robust than in the case presented in Table 2.

4. Conclusions

A curve fitting method for the analysis of sedimentation velocity data has been presented. A software package called SEDANAL has been developed to fit for sedimentation coefficients, molar masses, and equilibrium constants and/or rate constants for multispecies and multi-component

Table 2
Results of Monte Carlo simulations

	θ	Correct value	Average from fits ($n=10$)	Std. dev./average
(a) For $K_{int}=10^7$, global fit to six cells				
CA1 (M)	10.0	1.0E−06	9.970E−07	4.12E−03
CB1/CA1		2.0	2.008E+00	4.55E−03
CA2	3.0	3.0E−07	2.999E−07	4.45E−03
CB2/CA2		2.0	1.997E+00	3.50E−03
CA3	1.0	1.0E−07	1.003E−07	6.59E−03
CB3/CA3		2.0	2.008E+00	4.71E−03
CA4	3.0	3.0E−07	2.999E−07	2.66E−03
CB4/CA4		4.0	4.009E+00	3.36E−03
CA5	3.0	3.0E−07	2.996E−07	1.43E−03
CB5/CA5		8.0	8.015E+00	1.39E−03
CA6	3.0	3.0E−07	3.009E−07	9.94E−03
CB6/CA6		0.5	5.089E−01	1.98E−02
K_1 (M^{-1})		2.0E+07	2.136E+07	1.13E−01
K_2 (M^{-1})		5.0E+06	5.071E+06	3.43E−02
Sc (S)		6.0	5.953	1.19E−02
Sd (S)		8.0	8.001	9.56E−04
(b) For $K_{int}=10^6$, global fit to six cells				
CA1 (M)	3.0	3.0E−06	2.998E−06	5.93E−04
CB1/CA1		2.0	2.002E+00	7.93E−04
CA2	1.0	1.0E−06	9.994E−07	1.30E−03
CB2/CA2		2.0	2.002E+00	1.78E−03
CA3	0.3	3.0E−07	3.006E−07	6.90E−03
CB3/CA3		2.0	1.995E+00	8.92E−03
CA4	1.0	1.0E−06	9.993E−07	1.41E−03
CB4/CA4		4.0	4.004E+00	1.80E−03
CA5	1.0	1.0E−06	9.998E−07	4.44E−04
CB5/CA5		8.0	8.002E+00	4.33E−04
CA6	1.0	1.0E−06	1.001E−06	2.06E−03
CB6/CA6		0.5	4.996E−01	3.26E−03
K_1 (M^{-1})		2.0E+06	1.996E+06	5.19E−03
K_2 (M^{-1})		5.0E+05	5.019E+05	4.86E−03
Sc (S)		6.03	5.997	7.73E−04
Sd (S)		8.0	7.999	1.56E−04
(c) For $K_{int}=10^7$, fit to one cell				
CA (M)	1.0	1.0E−7	9.756E−08	5.87E−02
C_B/C_A		2.0	2.082E+00	7.38E−02
K_1 (M^{-1})		2.0E+07	2.865E+07	3.57E−01
K_2 (M^{-1})		5.0E+06	6.340E+06	2.52E−01
Sc (S)		6.0	5.524	1.75E−01
Sd (S)		8.0	7.990	6.74E−02
(d) For $K_{int}=10^6$, fit to one cell				
CA (M)	1.0	1.0E−06	1.009E−06	1.00E−02
C_B/C_A		2.0	1.979E+00	1.29E−02
K_1 (M^{-1})		2.0E+06	1.919E+06	4.35E−02
K_2 (M^{-1})		5.0E+05	4.866E+05	5.04E−02
Sc (S)		6.0	6.096	1.79E−02
Sd (S)		8.0	7.958	6.79E−03

Notes: Reaction simulated was $A+B=C$; $C+B=D$ with K_1 and K_2 as the macroscopic association constants for the first and second step, respectively. Sedimentation coefficients for *A*, *B*, *C* and *D* were 3.0, 4.0, 6.0 and 8.0 S, respectively. Molecular weights of *A* and *B* were 50 000 and 90 000, respectively. $K_{int}=0.5K_1=2K_2$; $\theta=C_AK_{int}$; $R=C_B/C_A$. Noise added to the synthetic data was ± 0.004 fringes, which corresponds to $\pm 1.2\times10^{-3}$ g l^{-1}. Simulations with the same values of θ and *R* can be normalized to each other, the only difference being the signal-to-noise level (Stafford, 1994).

macromolecular interacting systems. What distinguishes this approach from other nonlinear curve fitting algorithms using numerical solutions to the Lamm equation is that time-difference curves are employed to eliminate completely the time-independent systematic errors in the data. It should be pointed out that time-difference data are also used in fitting interference data by John Philo's program SVEDBERG. Moreover, time-dependent, radius-independent irregular fluctuations are nearly completely removed from the data by aligning the scans in the air–air region of the cell. This alignment can be improved considerably by placing interference window holders on the top of the cells (personal communications, Jeff Lary). The resulting time-difference curves usually contain only stochastic noise with a possible small amount of time-dependent instrumental drift caused mainly by changes in refractive properties of the optics arising from temperature variations. SEDANAL allows global fitting to multiple data sets obtained over a wide concentration range with multiple optical systems. The method allows fitting for either equilibrium constants or kinetic rate constants for each reaction step. With the currently available analytical ultracentrifuges, data from both the interference and absorbance optics can be combined in the global fit.

It is important to stress that the successful determination of equilibrium constants by this or any other method requires that all species represented in the model be significantly populated over the concentration range being studied. The best results will be achieved if the concentration range spans the same magnitude as the dissociation constants. Therefore, the range of applicability will be determined by the range of concentrations that can be accurately measured. Because of the high degree of effective averaging inherent in the least-squares approach, the lowest concentrations at which useable parameter estimates may be obtained is on the order of magnitude of the noise in the system. At the high concentration end of the scale, the upper limit will be determined by limitations in the optics like the amount of light getting through the optical system and other sources of systematic error related to high concentrations. One possible source of error in sedimentation velocity experiments is boundary instability that may lead to convection. These effects would be most serious at low protein concentrations at low salt and buffer concentrations. Redistribution of buffer components during sedimentation will produce gradients that can stabilize the boundary helping to prevent convection.

The approach to the analysis of heterogeneous interacting systems by sedimentation velocity that is presented in this paper, and embodied in SEDANAL, is fundamentally new and unique. It has been shown that with the proper choice of concentration range and component ratios, the method is robust with respect to the estimation of equilibrium constants and sedimentation coefficients. One advantage it has over equilibrium analysis is that shape information about the complexes can be obtained from their sedimentation coefficients. SEDANAL can also fit directly for kinetic rate constants.

SEDANAL can be obtained directly from one of the authors (WFS) or downloaded from the RASMB ftp archives at ftp://rasmb.bbri.org/rasmb/spin/ms_dos/sedanal-stafford/. An extensive user's manual and help file is included on the ftp site.

Acknowledgments

We acknowledge generous support from the National Science Foundation USA (NSF BIR-9513060 to WFS). This paper is dedicated to David Yphantis on the occasion of his retirement from the University of Connecticut. Without his example and encouragement over the years, this work could not have been accomplished.

References

[1] J.M. Claverie, H. Dreux, R. Cohen, Sedimentation of generalized systems of interacting particles. I. Solution of systems of complete Lamm equations, Biopolymers 14 (1975) 1685–1700.

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

[3] G.P. Todd, R.H. Haschemeyer, General solution to the inverse problem of the differential equation of the

ultracentrifuge, Proc. Natl. Acad. Sci. USA 78 (1981) 6739–6743.
[4] B. Demeler, H. Saber, Determination of molecular parameters by fitting sedimentation data to finite-element solutions of the Lamm equation, Biophys. J. 74 (1998) 444–454.
[5] P. Schuck, Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation, Biophys. J. 75 (1998) 1503–1512.
[6] W.F. Stafford, Time difference sedimentation velocity analysis of rapidly reversible interacting systems: determination of equilibrium constants by non-linear curve fitting procedures, Biophys. J. 74 (1998) A301.
[7] G. Rivas, W.F. Stafford, A.P. Minton, Characterization of Heterologous Protein–Protein Interaction via Analytical Ultracentrifugation. Methods Oct. 19 (2) (1999) 194–212.
[8] P. Schuck, B. Demeler, Direct sedimentation analysis of interference optical data in analytical ultracentrifugation, Biophys. J. 76 (1999) 2288–2296.
[9] W.F. Stafford, Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile, Anal. Biochem. 203 (1992) 295–301.
[10] J.S. Philo, An improved function for fitting sedimentation velocity data for low-molecular-weight solutes, Biophys. J. 72 (1997) 435–444.
[11] C.A. Sontag, W.F. Stafford, J.J. Correia, A comparison of weight average and direct boundary fitting of sedimentation velocity data for indefinite polymerizing systems, Biophys. Chem. 1108 (2004) 215–230.
[12] O. Lamm, Die Differentialgleichung der Ultrazentrifugierung, Arkiv Math. Astron. Fysik 21B (2) (1929) 1–4.
[13] S.E. Harding, P. Johnson, The concentration-dependence of macromolecular parameters, Biochem. J. 231 (1985) 543–547.
[14] J.A. Nelder, R. Mead, A simplex method for function minimization, Comput. J. 7 (1965) 308–313.
[15] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Richardson Extrapolation and the Bulirsch-Stoer Method. Numerical Recipes in FORTRAN: The Art of Scientific Computing, second ed., Cambridge University Press, Cambridge, England, 1992, pp. 718–725.
[16] G.W. Snedecor, W.G. Cochran, Statistical Methods, Iowa State University Press, 1967.