

# Determination of Molecular Parameters by Fitting Sedimentation Data to Finite-Element Solutions of the Lamm Equation

Borries Demeler\* and Hashim Saber#

\*Department of Biochemistry, University of Texas Health Sciences Center at San Antonio, San Antonio, Texas 78284-7760, and

#Department of Mathematics, University of Pittsburgh at Bradford, Bradford, Pennsylvania 16701 USA

**ABSTRACT** A method for fitting experimental sedimentation velocity data to finite-element solutions of various models based on the Lamm equation is presented. The method provides initial parameter estimates and guides the user in choosing an appropriate model for the analysis by preprocessing the data with the  $G(s)$  method by van Holde and Weischet. For a mixture of multiple solutes in a sample, the method returns the concentrations, the sedimentation ( $s$ ) and diffusion coefficients ( $D$ ), and thus the molecular weights (MW) for all solutes, provided the partial specific volumes ( $\bar{v}$ ) are known. For nonideal samples displaying concentration-dependent solution behavior, concentration dependency parameters for  $s(\sigma)$  and  $D(\delta)$  can be determined. The finite-element solution of the Lamm equation used for this study provides a numerical solution to the differential equation, and does not require empirically adjusted correction terms or any assumptions such as infinitely long cells. Consequently, experimental data from samples that neither clear the meniscus nor exhibit clearly defined plateau absorbances, as well as data from approach-to-equilibrium experiments, can be analyzed with this method with enhanced accuracy when compared to other available methods. The nonlinear least-squares fitting process was accomplished by the use of an adapted version of the "Doesn't Use Derivatives" nonlinear least-squares fitting routine. The effectiveness of the approach is illustrated with experimental data obtained from protein and DNA samples. Where applicable, results are compared to methods utilizing analytical solutions of approximated Lamm equations.

## INTRODUCTION

Sedimentation velocity data from the analytical ultracentrifuge are rich in information content. In principle, sedimentation coefficients, diffusion coefficients, molecular weights, equilibrium constants, and partial concentrations of each solute in a multiple-component solution can be obtained from velocity data. Currently available computer programs for analyzing sedimentation velocity data obtained from the Beckman XL-A analytical ultracentrifuge rely on one of the following methods: 1) simple calculations, such as midpoint and second moment methods (XLABEL XL-A Data Analysis Software; Beckman Instruments, Spinco Division, Palo Alto, CA); 2) graphical transformations of experimental data (van Holde and Weischet, 1978; Demeler et al., 1997); 3) difference methods based on time (Stafford, 1992) or radius (XLABEL); or 4) nonlinear least-squares fitting methods of approximate solutions to the Lamm equation (Holladay, 1979, 1980; Philo, 1994). In the latter category, several novel computer methods have been developed recently that greatly improve on previous methods by introducing additional correction terms. Behlke and Ristau (1997) developed a program that utilizes five different model functions derived by Fujita (1962, 1975), including the Faxén and the Archibald types, which are all analytical solutions to approximations of the Lamm equation

(Lamm, 1929). The latest version of the program SVEDBERG (Philo, 1997) introduces two correction terms to the solution of the approximated Lamm equation by Fujita and MacCosham (1959) to reduce some of the limitations imposed by the assumptions. However, most currently available methods suffer in accuracy from the approximations or assumptions in one or more of the following cases: 1) experimental data points near the meniscus are included in the analysis in cases where the sample has not cleared the meniscus; 2) the sample does not produce a well-defined plateau value for each scan; 3) in cases where low-molecular-weight samples produce a concentration gradient that rapidly approaches equilibrium distributions at the speeds possible in the XL-A; or 4) experimental data from samples exhibiting concentration dependency in  $s$  or  $D$ . With the exception of the work of Behlke et al. (1997), the concentration dependency of  $s$  is not considered. The accuracy of currently available direct fitting methods is generally tested by fitting data that have been simulated with finite-element solutions of the Lamm equation as developed by Claverie and co-workers (Claverie et al., 1975; Claverie, 1976). The fitting results are then compared to the known input parameters for the simulation. The reason for selecting finite-element solutions as test functions is based on the fact that these solutions of the Lamm equation are devoid of any assumptions or correction terms. These numerical solutions discretize the variables of the equation (in this case, time and radius), and the resulting accuracy depends on the step size used in the discretization of the variables. Hence, by making the step size small, it is possible to obtain any desired degree of accuracy supported by the amount of memory available in the computer. Thus, although each of

Received for publication 9 June 1997 and in final form 25 September 1997.

Address reprint requests to Dr. Borries Demeler, Department of Biochemistry, The University of Texas Health Sciences Center at San Antonio, 7703 Floyd Curl Dr., San Antonio, TX 78284-7760. Tel.: 210-567-6592; Fax: 210-567-6595; E-mail: demeler@bioc02.uthscsa.edu.

© 1998 by the Biophysical Society

0006-3495/98/01/444/11 \$2.00

the currently available fitting methods has its own merits and can yield accurate results under appropriate conditions, the full spectrum of possible cases cannot currently be accommodated. Examples that cannot easily be analyzed include very low molecular weight molecules with large diffusion coefficients ( $\geq 1 \times 10^{-6} \text{ cm}^2/\text{s}$ ), when data near the cell boundary are included, or samples that exhibit concentration dependency in  $s$ . Examples of the latter include larger DNA molecules as well as unfolded, charged, or large proteins. In these cases, if concentration dependency is not included in the model, the fitting results obtained will be unreliable. Todd and Haschemeyer (1981) presented a method for fitting simulated data to finite-element solutions of various modified Lamm equations. In this article we present an improved method that fits digitally acquired data from the Beckman XL-A analytical ultracentrifuge directly to finite-element solutions by using the van Holde–Weischet method for initial guesses and a nonlinear least-squares approach that does not rely on derivatives to improve calculation speed. Several experimental model systems have been analyzed, including samples that 1) exhibit concentration dependence, 2) are composed of single or multiple noninteracting solutes, and 3) approach sedimentation equilibrium.

## MATERIALS AND METHODS

### Materials

All calculations were performed on a Pentium-type personal computer configured with both the LINUX (v. 2.0.30) and MS-DOS/Windows operating systems, and the UltraScan software package (version 2.97). Finite-element fitting was accomplished with the SEDVFIN module of UltraScan. A binary version of SEDVFIN for various UNIX platforms, including a technical report describing the file format, has been made available at the sedimentation data analysis software archives on the WWW server from the University of Texas Health Science Center, Biochemistry Department (<http://bioc09.uthscsa.edu/biochem/xla2.html>). Chicken egg white lysozyme and horse heart myoglobin were obtained from Sigma. DNA samples used were derived from the pPOL-208–12 plasmid, which harbors 12 tandem repeats of the sea urchin 5S ribosomal RNA gene (Georgel et al., 1993). DNA samples were obtained by digesting the plasmid with *EcoRI* and purifying the resulting 16-bp fragments on a 1% agarose gel by using Supelco GenElute agarose spin columns.

### Analytical ultracentrifugation

DNA samples were centrifuged in 10 mM Tris-HCl/1 mM EDTA (pH 8.0). Protein samples were centrifuged in 0.1 M phosphate-buffered saline at pH 6.5 (PBS). All experiments were performed using a Beckman XL-A equipped with an AN-60 Ti rotor and absorbance optics. For runs above 42,000 rpm, aluminum centerpieces were used. In all other cases, epon/charcoal centerpieces were used. Other details of run conditions are indicated in the figure legends.

### Models of the Lamm equation

The Lamm equation (Lamm, 1929), which describes the sedimentation velocity behavior of a broad range of systems in the ultracentrifuge cell, is

given in its general form by

$$\frac{\partial C_k}{\partial t} + \frac{1}{r} \frac{\partial(rJ_k)}{\partial r} = f_k; \quad J_k = s_k \omega^2 r C_k - D_k \frac{\partial C_k}{\partial r} \quad (1)$$

with boundary and initial conditions:

$$J(r_m, t) = J(r_b, t) = 0, \quad 0 \leq t \leq T; \quad C(r, 0) = C_0(r) \quad (2)$$

In this equation, the subscript  $k$  refers to the solute  $k$ ,  $C$  is the solute concentration,  $J$  is the flux,  $t$  is time,  $T$  is the elapsed time at the end of the experiment,  $r$  is the radius from the center of rotation, and  $r_m \leq r \leq r_b$ , where  $r_m$  and  $r_b$  are the radii of the meniscus and the bottom of the cell.  $\omega$  is the angular velocity, and  $s_k$  and  $D_k$  are the sedimentation and the diffusion coefficient of component  $k$ .

In the general case, concentration-dependent nonideal solution behavior has to be considered, and  $s_k$  and  $D_k$  are not constant, and are functions of concentration. As described by Claverie (1976), this can be represented in the form of Eqs. 3 and 4, where  $\sigma_k$  and  $\delta_k$  are constant, dimensionless parameters that describe the variation of  $s_k$  and  $D_k$  from their value at infinite dilution,  $s_{k0}$  and  $D_{k0}$ :

$$s_k = s_{k0}(1 - \sigma_k C_k) \quad (3)$$

$\sigma = \delta = 0$  corresponds to the ideal case where the sedimentation and diffusion coefficient are

$$D_k = D_{k0}(1 + \delta_k C_k) \quad (4)$$

independent of  $C$ . The latter is generally true for small, globular macromolecules at low concentration. The models for concentration dependence of  $s$  and  $D$  as shown in Eqs. 3 and 4 represent the simplest case where interactions between different components are neglected.

Multiple noninteracting components were modeled by summing the partial concentrations,  $C_k$ , of each solute  $k$ :

$$C_{\text{total}} = \sum_{k=1}^i C_k \quad (5)$$

where  $i$  is the total number of solutes in the system.

No analytical solution of Eq. 1 has yet been presented. Approximate solutions have been developed (Faxén, 1929; Fujita and MacCosham, 1959) and have been employed to estimate  $D$ ,  $s$ , and  $s/D$  (Philo, 1994, 1997; Behlke and Ristau, 1997; Holladay, 1979, 1980). A number of numerical solutions have also been employed (for a discussion see Cox, 1965a–c; Dishon et al., 1969). For the studies presented here, the finite-element method was chosen to obtain solutions to Eq. 1. A detailed derivation of each finite-element solution has been described previously by Claverie et al. (Claverie et al., 1975; Claverie, 1976). Detailed information on the finite-element method and its application to solutions of partial differential equations can be found in Zienkiewicz (1971).

### Curve fitting of sedimentation data

The process of fitting experimental data consists of four distinct steps:

1. *Initialization*: In the initialization step, experimental data are first analyzed by the van Holde–Weischet analysis method (van Holde and Weischet, 1978) in order to determine the appropriate model and initial parameter estimates for the finite element fit.

2. *Simulation*: By using the initial estimates and the model obtained in step 1, experimental data are simulated from the appropriate finite-element solution.

3. *Minimization*: During minimization, the DUD nonlinear least-squares minimization algorithm (Ralston and Jennrich, 1978) is used to determine a new set of parameter estimates. In an iterative process, steps 2 and 3 are

repeated until the sum of the differences between experimental and simulated data has been minimized.

4. *Refinement*: The fit is refined by repeating steps 2 and 3, with either additional parameters being allowed to float, and/or by adjusting fitting step sizes in DUD and discretization step sizes in the finite-element solution until an optimal solution has been found. Each of these steps is discussed in detail below.

### Initialization

To perform nonlinear least-squares minimization on the experimental data, the finite-element simulation needs to be initialized with estimates for each parameter. As in most nonlinear least-squares minimizations, the closer the initial parameter estimates are to the final result, the fewer iterations are required to achieve convergence. Because functional evaluations of the finite-element solution are computationally expensive, it is highly desirable to be as close to the final solution as possible when starting the minimization. We have found that preprocessing the data with the van Holde–Weischet analysis method (van Holde and Weischet, 1978) provides all of the required estimates with sufficient accuracy to ensure rapid convergence. In addition, this analysis is well suited to identifying the appropriate model for the analysis for a wide range of experimental conditions—a very important requirement. A detailed description of the diagnostics available from the van Holde–Weischet method and how this method can be used to determine which model is appropriate can be found in Demeler et al. (1997). The experimental data from the XL-A are analyzed with the UltraScan software package (version 2.97) to determine meniscus position, temperature, loading concentration, baseline absorbance, and time corrections accounting for the acceleration period of the rotor. The UltraScan software package is then used to create a van Holde–Weischet extrapolation plot and an integral distribution plot. By visual inspection of the extrapolation plot, the number of components in the system and the required finite-element model are determined by the user. Parameter estimates are obtained as follows. The number of converging lines in the van Holde–Weischet plot are used to calculate partial concentrations of components (see Fig. 5,  $C$  and  $D$ ). Sedimentation coefficients are determined by averaging the intercepts of respective components in the van Holde–Weischet plot. Diffusion coefficients are estimated as described by equation 10 in van Holde et al. (1978). From the van Holde–Weischet analysis we have

$$s_w^* = s - \frac{2D^{1/2}}{r_0\omega^2} \Phi^{-1}(1 - 2w)t^{-1/2} \quad (6)$$

where  $w$  is the partial concentration  $C_i/C_p$ ,  $s_w^*$  is the apparent sedimentation coefficient at that concentration,  $s$  is the extrapolated sedimentation coefficient,  $D$  is the diffusion coefficient,  $r_m$  is the meniscus position,  $\omega$  is the radial velocity,  $t$  is the time, and  $\Phi$  is the error function. This is the equation of a straight line, and solving for  $D$ , one obtains

$$D = \left[ \frac{mr_0\omega^2}{2\Phi^{-1}(1 - 2w)} \right]^2 \quad (7)$$

The slopes can be obtained from the van Holde–Weischet plot, and, by using a numerical evaluation of the error function  $\Phi$ , diffusion coefficient estimates can be calculated by averaging the  $D$ -value obtained for each  $w$  for each component  $k$ . Because this method sometimes can result in considerable error (see van Holde et al., 1978), an alternative method is provided: if an estimate for the molecular weight of the sample is known a priori, the initial value for  $D_k$  can also be calculated from the molecular weight of each component  $k$  and the estimate of  $s_k$  according to the Svedberg relationship:

$$D_k = \frac{s_k RT}{M_k(1 - \bar{v}_k \rho)} \quad (8)$$

where  $M_k$  is the molecular weight of component  $k$ ,  $R$  is the gas constant,  $T$  is the temperature in Kelvin,  $\rho$  is the density of the buffer, and  $\bar{v}_k$  is the partial specific volume of the component. For the case of concentration dependency, integral distribution plots are used to estimate a sedimentation coefficient at the most dilute concentration (see Fig. 3). Initial estimates for the values of  $\sigma_k$  are determined by the following relationship, suggested by the model in Eq. 3:

$$\sigma(k) = \left( 1 - \frac{s(C_{0,k})}{s(C=0)} \right) * (C_{0,k})^{-1} \quad (9)$$

Here the values for the sedimentation coefficient for each component are read off the integral distribution plot. The  $s_{C=0}$  values are estimated by extrapolating the distribution plot to  $C = 0$  (Demeler, 1992). The baseline absorbance is estimated by averaging a few points in the depleted region near the meniscus in the last scan of each experiment. Because no method for the initialization of the concentration dependency of the diffusion coefficient is available, the diffusion coefficient was always initialized at zero and held constant until refinement, when it was initialized with a small number ( $\sim 1.0 \times 10^{-4}$ ) and allowed to float, if concentration dependence seemed to be present and an improved fit could be obtained by including the parameter.

Finally, because the finite-element solution incorporates the radial position at the bottom of the cell, speed-dependent rotor stretching must be taken into account. Data measured by Beckman Instruments (D. McRorie, Beckman Instruments, personal communications) for rotor stretching of the AN 50 Ti (eight-position) and AN 60 Ti (four-position) rotors were fitted to a fourth-degree polynomial (see Table 1). Because charcoal epon and aluminum centerpieces are cut from different dies, a provision is made in the software to enter the appropriate position at the bottom of the cell at rest. The polynomial was incorporated into the finite-element solution to automatically correct for rotor speed-dependent position changes at the bottom of the cell.

### Simulation

A brief discussion of the finite-element solution is included here; a more complete description can be found in Claverie et al. (Claverie et al., 1975; Claverie, 1976). Simulation is achieved by deriving the finite-element solution of the system of equations given in Eq. 1. The finite-element solution uses a variational formulation of Eq. 1, which is given by

$$\int_{\Omega} \frac{\partial C_k}{\partial t} \cdot v \cdot r \cdot dr + \int_{\Omega} \frac{\partial(rJ_k)}{\partial r} \cdot v \cdot dr = \int_{\Omega} f_k \cdot v \cdot r \cdot dr \quad (10)$$

where  $\Omega$  is the  $[r_m, r_b]$  domain and  $v$  is a test function. By integrating by parts, inserting the explicit expression of  $J_k$  (see Eq. 1), and using the linear forms of  $s_k$  and  $D_k$  given in Eqs. 3 and 4, an equation similar to equation 10 of Claverie et al. (1975) is obtained. To discretize this equation in  $r$ , the distance between meniscus and the bottom of the cell is divided into  $N + 1$  equally spaced increments of length  $h = (r_b - r_m)/N$ , where  $N$  is the

**TABLE 1** Coefficients for a fourth-degree polynomial describing rotor stretching in the AN 60 Ti and AN 50 Ti rotors for the XL-A

$Y = A_0 + A_1X + A_2X^2 + A_3X^3 + A_4X^4$ ( $Y$ in cm, $X$ in rpm) (valid range: 0–60,000 rpm)					
	$A_0$	$A_1$	$A_2$	$A_3$	$A_4$
AN 60 Ti	3.128E-5	−6.620E-9	7.264E-12	−6.152E-17	5.760E-22
SD	3.868E-6	8.973E-10	6.107E-14	1.532E-18	1.265E-23
AN 50 Ti	7.754E-5	−1.546E-8	9.601E-12	−5.800E-17	6.948E-22
SD	2.671E-6	7.531E-10	6.139E-14	1.849E-18	1.836E-23

number of elements in the space discretization. At this point it should be mentioned that the data from the XL-A do not provide uniform radial increments. We have observed that for a 0.001-cm acquisition setting in the XL-A, radial increments ranging from  $5 \times 10^{-4}$  cm to  $5 \times 10^{-3}$  cm can be obtained, with an average of  $\sim 1.8 \times 10^{-3}$  cm step size. For convenience in computation, we chose to use a constant value of  $h$  of  $1.0 \times 10^{-3}$  cm, which can be adjusted according to experimental needs (see Discussion). For the purpose of residual calculation, we adjust the experimental data by interpolating missing points and omitting points that have a smaller radial increment than 0.001 cm. For the time discretization, the finite difference technique was used: The time coordinate  $t$  is divided into intervals of length  $\Delta t$ . It was found that a  $\Delta t$  value of 25 s provides sufficient accuracy for initial minimizations, improving speed of computation. During refinement, the  $\Delta t$  value is adjusted to 1 s in the final fit. This leads to a system of  $N + 1$  linear equations in matrix form for each solute (explicitly shown in equation 5 of Claverie et al., 1975), which is given by

$$AC_{n+1} = B_n \quad (11)$$

where  $n$  is an integer index for the time discretization. The matrix  $A$  is tridiagonal and can be solved with Gaussian elimination. For each solute, the concentration  $C$  is represented by a vector  $C_n$ , the elements of which are the values of the concentration  $C$  at radii  $r_m, r_m + h, \dots, r_b$  at time  $t_0 + n\Delta t$ . Starting with  $n = 0$  and the initial conditions ( $C_0$  for each solute), together with an estimate for the unknown parameters  $s_k$  and  $D_k$ , the vector  $C_{n+1}$  is recursively calculated from  $C_n$  by solving Eq. 11. For the case of strong concentration dependency, we found that inclusion of second-order Taylor expansion terms in equations 1–3 and 16–18 of Claverie (1976) can improve the fit. The equations for the concentration dependency of  $s$  and  $D$ , including the second-order terms, are given by

$$s \approx s_0(1 - \sigma C + \sigma^2 C^2) \quad (12)$$

$$D \approx D_0(1 + \delta_1 C + \delta_2 C^2) \quad (13)$$

Using the same notation as Claverie (1976), we can rewrite the matrix equations of the finite-element solution as

$$\begin{aligned} (B + \Delta t D_0 A^1 - \Delta t s_0 \omega^2 A^2) C_{n+1} \\ = BC_n + \Delta t D_0 (UC^{U1} + VC^{V1} + WC^{W1}) \\ + \Delta t D_0 (UC^{U2} + VC^{V2} + WC^{W2}) \\ - \Delta t s_0 \omega^2 A^2 C^{A1} - \Delta t s_0 \omega^2 A^2 C^{A2} \end{aligned} \quad (14)$$

where

$$\begin{aligned} C_i^{U1} &= (\delta_n)_i (C_n)_i (C_n)_{i+1} & C_i^{U2} &= (\delta_n^2)_i (C_n)_i^2 (C_n)_{i+1} \\ C_i^{V1} &= (\delta_n)_i (C_n)_i (C_n)_1 & C_i^{V2} &= (\delta_n^2)_i (C_n)_i^2 (C_n)_1 \\ C_i^{W1} &= (\delta_n)_{i+1} (C_n)_{i+1} (C_n)_{i+1} & C_i^{W2} &= (\delta_n^2)_{i+1} (C_n)_{i+1}^2 (C_n)_i \\ C_i^{A1} &= (\sigma_n)_i (C_n)_i (C_n)_i & C_i^{A2} &= (\sigma_n^2)_i (C_n)_i^2 (C_n)_i \end{aligned} \quad (15)$$

### Minimization

The nonlinear least-squares fitting method used in this study is a derivative-free algorithm based on the Gauss-Newton method, termed “Doesn’t use Derivatives” (DUD) (Ralston and Jennrich, 1978). The general procedure can be summarized as follows. Let  $\mathbf{P}$  be a vector, the elements of which are estimates of the unknown parameters, and let the experimental data be expressed as a vector  $\mathbf{Y}$ , the elements of which are the experimental data values at various scan times and radial positions. A corresponding vector of simulated data,  $Z(\mathbf{P})$ , can be calculated by using the finite-element solution to an appropriate model based on the Lamm equation (Eq. 1). The goal of the nonlinear least-square algorithm is to find  $\hat{\mathbf{P}}$ , the value

of  $\mathbf{P}$  that minimizes the residual sum of squares  $\mathbf{Q}(\mathbf{P})$ :

$$\mathbf{Q}(\mathbf{P}) = \sum_{j=1}^l (Y_j - Z_j(\mathbf{P}))^2 \quad (16)$$

where  $l$  is the length of  $\mathbf{Y}$ , which corresponds to the total number of experimental observations included in the fit. A number of algorithms for finding  $\hat{\mathbf{P}}$  are available, all of which must be initially supplied with estimates of the unknown parameters. As stated earlier, the initial estimates are supplied by the van Holde–Weischet analysis of the experimental data. Global parameters of the run that can be adjusted (floated) during the fit include baseline absorbance, meniscus position, the position at the bottom of the cell, and a linear correction term proportional to the total absorbance:

$$C_{i,j,\text{corrected}} = (1 + \epsilon) (C_{i,j} + C_{\text{baseline}}) \quad (17)$$

where  $\epsilon$  is a small, dimensionless, positive or negative constant. This correction is necessary in some cases where an apparent slope proportional to the total absorbance is superimposed on the experimental data. We have not been able to definitively identify the cause of the sloping, and do not observe it in each experiment, but we suspect that this error is caused by a fault in the optical system or by stray light. For each component, the partial initial concentration, the sedimentation and diffusion coefficients, and the concentration dependency parameter  $\sigma$  can be allowed to float. Then the parameters are iteratively corrected until an optimal solution is found, i.e.,  $\mathbf{P}$  converges to  $\hat{\mathbf{P}}$ .

Most nonlinear minimization algorithms utilize steepest gradient descend methods requiring either an analytical or a numerically calculated partial derivative for each parameter (for a review, see Johnson and Faunt, 1992). In our case, such an approach poses a disadvantage, because 1) it is difficult to provide an analytical derivative for this function and 2) because it is computationally expensive to provide a numerically evaluated partial derivative for each parameter for this function in each iteration. In the general case, there will be  $5k + 4$  parameters given by the vector  $\mathbf{P}$ , where  $k$  is the number of solutes present in the system. Therefore, we chose to implement the DUD algorithm (Ralston and Jennrich, 1978). Instead of derivatives, this algorithm utilizes a derivative-free Gauss-Newton approach, providing substantial computational cost savings for this problem. For the convenience of the reader, we summarize here the  $j$ th iteration of the DUD algorithm for a system with  $p$  fitted parameters: let  $P_1^j, P_2^j, P_3^j, \dots, P_{p+1}^j$  be a set of estimated parameters for the solution computed in the previous iteration (numbered by age, where  $P_1^j$  is the oldest). Approximate the function  $Z(\mathbf{P})$  by a linear function  $\Theta^j(\mathbf{P})$ , which is equal to  $Z(\mathbf{P})$  at  $p + 1$  points ( $P_i^j; i = 1, \dots, p + 1$ ). Find  $P_{\text{new}}^j$ , the point that minimizes the distance between  $\Theta^j(\mathbf{P})$  and  $\mathbf{Y}$ . Replace  $P_1^j$  with  $P_{\text{new}}^j$  to obtain the updated parameter set. The  $p + 1$  initial starting values required by DUD are generated from one user-supplied  $\mathbf{P}_{p+1}$ . For  $l = 1, \dots, p$ ,  $\mathbf{P}_i$  is computed from  $\mathbf{P}_{p+1}$  by displacing its  $i$ th component by a user-definable singularity factor times the corresponding component of  $\mathbf{P}_{p+1}$ . The algorithm can then be summarized as follows.

1. Generate an  $l \times p$  matrix  $\Delta F$  whose  $i$ th column is given by

$$\Delta F_i = Z(\mathbf{P}_i) - Z(\mathbf{P}_{p+1}) \quad (18)$$

where each vector  $\mathbf{Z}$  is a vector of size  $l$ , generated from the simulation algorithm, where  $l$  stands for the total number of experimental observations.

2. Generate a  $p$ -vector  $\alpha$  using

$$\alpha = (\Delta F^T \Delta F)^{-1} (\Delta F^T (\mathbf{Y} - Z(\mathbf{P}_{p+1}))) \quad (19)$$

3. Compute a new parameter vector  $\mathbf{P}_{\text{new}}$ :

$$\mathbf{P}_{\text{new}} = \mathbf{P}_{p+1} + \Delta \mathbf{P} \alpha \quad (20)$$

where  $\Delta \mathbf{P}$  is the  $p \times p$  matrix whose  $i$ th column is given by

$$\Delta \mathbf{P}_i = \mathbf{P}_i - \mathbf{P}_{p+1} \quad (21)$$



4. The convergence criterion is given by

$$Q(P_{\text{new}}) - Q(P_{\text{old}}) = 0 \quad (22)$$

5. If the convergence criterion is not met, update the set of starting values  $P_1, P_2, \dots, P_{p+1}$ . Normally, the new estimate  $P_{\text{new}}$  replaces  $P_1$  (the oldest number in the set), and the new starting set is formed by rearranging the set  $P_1, P_2, \dots, P_{p+1}$ , so that  $P_1$  is the oldest.

6. For some data sets, convergence will not occur without using a step-shortening procedure that modifies  $P_{\text{new}}$ . We found the following procedure to work quite well for the cases that have been examined to date:

$$P_{\text{new}} = d * P_{\text{new}} + (1 - d)P_{p+1} \quad (23)$$

where  $d$  is the first member of the sequence:

$$d_i = -(-10^{-3})^i \quad i = 1, 2, \dots, 5 \quad (24)$$

### Refinement

The process of refinement of a fit involves repeating the simulation and minimization steps by allowing additional parameters to be floated that may be required, and repeating the fit to converge on a better solution. Moreover, step size and singularity factor adjustments often improve the convergence and help to overcome local minima in the solution space.

## RESULTS

### Single ideal species, long-column sedimentation velocity

The sedimentation pattern and overlaid finite-element fit by a single ideal species model of horse heart myoglobin in a long-column sedimentation velocity experiment are shown in Fig. 1 *A*. Residuals of the fit are shown in Fig. 1 *B*. To be able to compare results with those obtained from approximate solutions to the Lamm equation from SVEDBERG

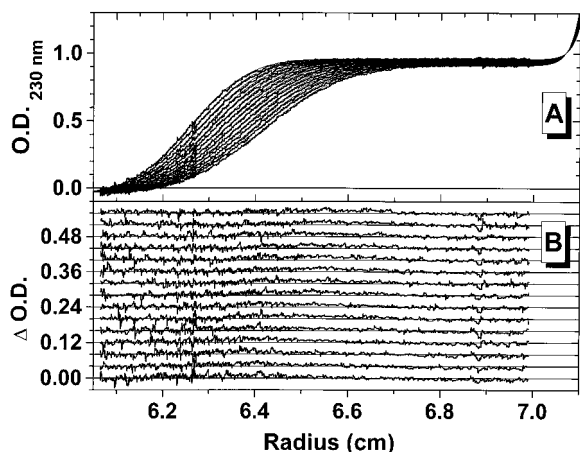


FIGURE 1 Finite-element fit of a long column ( $\sim 1.3$  cm) sedimentation velocity experiment of horse heart myoglobin in PBS buffer. Run conditions were as follows: Rotor speed: 60 krpm; wavelength: 230 nm; temperature: 20.2°C. (*A*) Velocity data and overlaid finite-element solution for a single, ideal component system. Only scans that display a well-defined plateau region and have cleared the meniscus are included in the fit. (*B*) Residuals of finite-element fit in *A*. Residuals are shown with 0.03 absorption unit offsets, with the residuals for the first scan shown on the bottom (0.0), and residuals of the last scan shown at the top of the plot (0.54).

(Philo, 1994), we restricted the data included in this fit by the following criteria: 1) the meniscus was depleted; 2) no data from the bottom of the cell were included; 3) a region of well-defined plateau absorbance was available. Both Fujita and Fujita-McCosham solutions provided results essentially identical to those from the finite-element fit. The calculated molecular weight for all three models agreed with the calculated molecular weight derived from the amino acid sequence to within 1% or less (shown in Table 2).

### Single ideal species, short column, and approach to equilibrium

The sedimentation pattern of lysozyme and overlaid finite-element fit of the data by a single ideal species model in a short-column, approach-to-equilibrium sedimentation velocity experiment is shown in Fig. 2 *A*. In this case, neither depletion of absorbance near the meniscus nor a stable plateau absorbance can be obtained. This is caused by significant diffusion and the relatively small sedimentation coefficients common in small molecules. The solute will almost immediately build a gradient through back-diffusion into the cell, obliterating any stable plateau absorbance, and, during later scans, a gradient approaching equilibrium is created that neither depletes the meniscus nor provides any discernible plateau concentration. As Behlke and Ristau point out (Behlke and Ristau, 1997), such data will produce skewed results because of errors introduced by the solution near the cell boundary when solutions to approximated Lamm equations are used. Therefore, comparisons to other methods are not appropriate and were not attempted. The residuals of the fit are shown in Fig. 2 *B*. Sedimentation and diffusion coefficients agree well with previously published data (Behlke and Ristau, 1997). As shown in Fig. 2 *B*, the residuals increase in magnitude for data near the bottom of the cell for the later scans. Those scans have an increased absorbance near the bottom of the cell. This effect is caused by the reduced signal-to-noise ratio at higher optical densities. Therefore, some of the data near the bottom of the cell were excluded from the analysis. The results are summarized in Table 2.

### Two-component system and concentration dependency of DNA

The sedimentation data and the finite-element fit of the sedimentation profile of a 196-bp DNA fragment in low-salt conditions are shown in Fig. 3 *A*. Residuals of the fit are shown in Fig. 3 *B*. The van Holde-Weischet extrapolation plot of the data (Fig. 3 *C*) indicates the presence of a small amount of other components. As can be deduced from the analyzed boundary fraction,  $\sim 12\%$  of the material sediments slower than the major component, whereas a small amount ( $\sim 4\%$ ) sediments faster than the major component ( $\sim 84\%$ ; fractions are available from the integral distribution

**TABLE 2** Results for velocity fits for ideal, single-component model

	Horse heart myoglobin	Horse heart myoglobin	Horse heart myoglobin	Lysozyme short column
Program	Finite-element	Svedberg*	Svedberg <sup>#</sup>	Finite-element
Baseline OD	-0.03258	-0.04709	-0.04663	0.02729
$C_k$ (OD)	1.066	1.084	1.0829	0.62205
Slope corr.	0.0	0.0	0.0	0.0
Meniscus	6.038 cm	6.038	6.038	6.633
Bottom <sup>§</sup>	7.1749 cm (A)	N/A	N/A	7.205 (E)
$s_{20,w}$ (s)	$1.882 \times 10^{-13}$	$1.913 \times 10^{-13}$	$1.907 \times 10^{-13}$	$1.849 \times 10^{-13}$
$D_{20,w}$ (cm <sup>2</sup> /s)	$1.051 \times 10^{-6}$	$1.081 \times 10^{-6}$	$1.080 \times 10^{-6}$	$1.042 \times 10^{-6}$
$\bar{v}$ (cm <sup>3</sup> /g)	0.745 <sup>¶</sup>	0.745 <sup>¶</sup>	0.745 <sup>¶</sup>	0.703 <sup>  </sup>
MW (calc.)**	$1.7012 \times 10^4$	$1.6858 \times 10^4$	$1.6765 \times 10^4$	$1.450 \times 10^4$
MW (theor.) <sup>###</sup>	$1.6956 \times 10^4$	$1.6956 \times 10^4$	$1.6956 \times 10^4$	$1.444 \times 10^4$
MW ( $\Delta\%$ )	0.33%	-0.58%	-1.13%	0.38%

NA, Not applicable.

\* Fujita solution, one species model.

<sup>#</sup> Fujita-McCosham solution, one species model.

<sup>§</sup> The bottom of the cell position listed refers to the rotor-stretching corrected value. (A), Aluminum centerpiece; (E), epon centerpieces.

<sup>¶</sup> Partial specific volumes were calculated based on the partial amino acid composition according to Laue (1992).

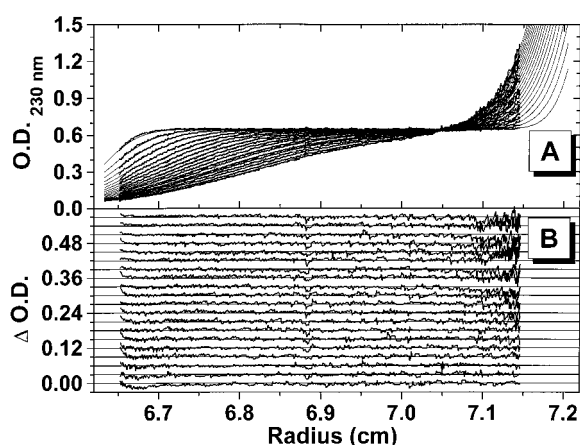
<sup>||</sup> Partial specific volume as reported in Sober (1968).

\*\* Molecular weight based on fitted parameters according to Eq. 8.

<sup>###</sup> Molecular weight based on sequence (Swiss Protein Database).

plot in Fig. 3 D). (Because of experimental noise in the data, 5% of the total boundary near the baseline was excluded from the analysis and was not included in the distribution and extrapolation plots. This portion of the data would have caused stray analysis results, if it had been included. The partial concentration calculated for the minor component near the baseline therefore must be increased by an additional 5%.) These minor components probably result from small amounts of restriction enzyme or small-molecular-

weight digestion products that copurified during the isolation of the major component, the DNA fragment. In addition, the data are further complicated by the presence of concentration dependency in the major component. As indicated by the crossing over of the extrapolation lines in the van Holde–Weischet extrapolation plot (Fig. 3 C) near the y axis at infinite time, the DNA fragment exhibits moderate concentration dependency in  $s$  (see Demeler et al., 1997). To account for the slower sedimenting species, a second component was introduced in the fit. The fast sedimenting impurities (4% of the total absorbance) were ignored in the fit, resulting in a small systematic error in the fit (Fig. 3 B). The major component was fitted by allowing the parameters for concentration dependency of both  $s$  and  $D$  to float. The data for this fit are summarized in Table 3.



**FIGURE 2** Finite-element fit of a short column (~6 mm) approach-to-equilibrium experiment of chicken eggwhite lysozyme in PBS buffer. Run conditions were as follows: Rotor speed: 40 krpm; wavelength: 230 nm; temperature: 20.1°C. (A) Velocity data and overlaid finite-element solution for a single, ideal component system. Scans that have not cleared the meniscus and show no well-defined plateau absorbance are included in the fit. (B) Residuals of finite-element fit in A. Residuals are shown with 0.03 absorption unit offsets, with the residuals for the first scan shown on the bottom (0.0), and residuals of the last scan shown at the top of the plot (0.57). Note that residuals at the bottom of the cell reach into higher absorbance ranges and therefore show increasing noise levels.

### Partial boundary fitting

In cases where small amounts of minor components are present (see previous example, shown in Fig. 3), it is possible to fit only the portion of the boundary corresponding to the solute in question (Fig. 4). This allows the user to reduce the number of floating parameters, which often improves the convergence properties for the fit. Sedimentation velocity data and the overlaid finite-element fit are shown in Fig. 4 A, and the residuals of the fit are shown in Fig. 4 B. To obtain accurate results during partial boundary fitting, it is essential that both the baseline and the partial concentration of the solute in question are known, so that these parameters can be fixed during the fit. Fortunately, this information is directly available from the van Holde–Weischet analysis of the data. As shown in Fig. 3 C, the portion of the boundary corresponding to the major solute comprises ~86% of the total concentration, with an offset of 12% from the baseline

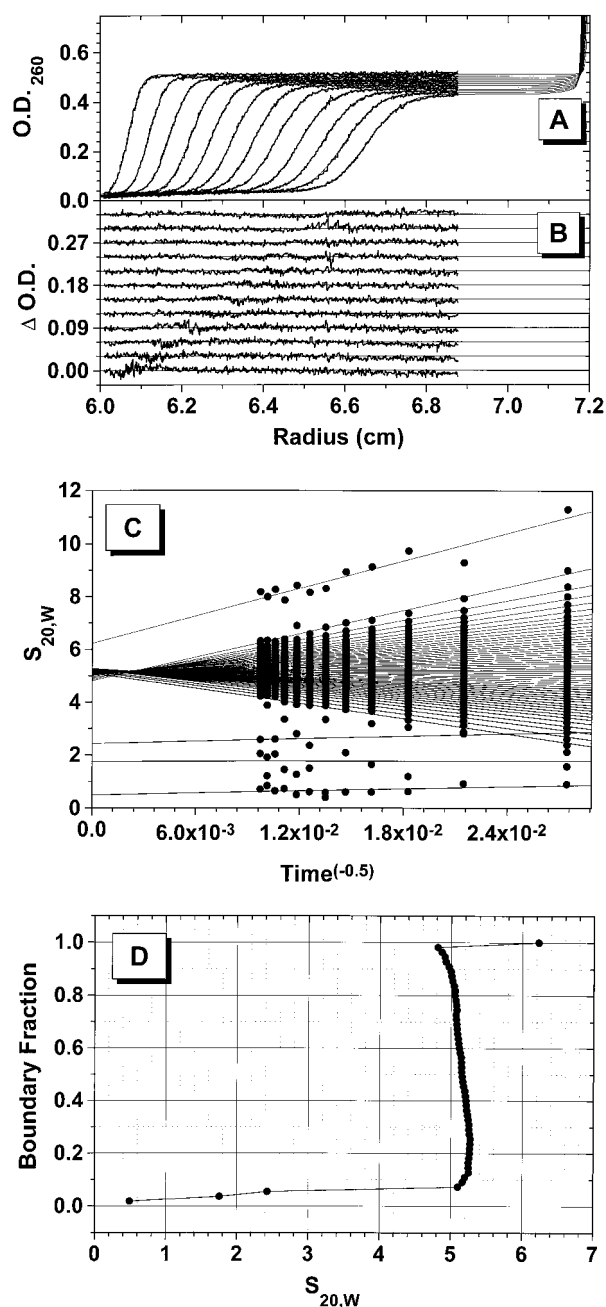


FIGURE 3 Finite-element fit of sedimentation velocity data of a 196-bp DNA fragment from the 5S ribosomal RNA gene of sea urchin in TE buffer. Run conditions were as follows: Rotor speed: 42 krpm; wavelength: 260 nm; temperature: 20.0°C. (A) Velocity data and overlaid finite-element solution for a double, nonideal component system. (B) Residuals of finite-element fit in A. Residuals are shown with 0.03 absorption unit offsets, with the residuals for the first scan shown on the bottom (0.0), and residuals of the last scan shown at the top of the plot (0.33). (C) van Holde-Weischet extrapolation plot for data shown in A. Note that several impurities are present; some are sedimenting faster, some slower than the DNA. (D) Integral distribution plot of C.

(fractions are available from the integral distribution plot in Fig. 3 D). If an initial scan has been prepared, the total concentration is also known, and the baseline and plateau concentrations can easily be calculated for the solute in

TABLE 3 Results for DNA velocity experiment, full and partial boundary fits

Sample Parameter	DNA (Component 1)	Impurities (Component 2)	DNA (partial boundary)
Baseline OD	$9.315 \times 10^{-3}$	N/A	$3.302 \times 10^{-2}$
$C_k$ (OD)	$4.94 \times 10^{-1}$	$2.37 \times 10^{-2}$	$4.91 \times 10^{-1}$
Slope corr.	$1.10 \times 10^{-3}$	N/A	0.0
Meniscus	5.991 cm	N/A	5.991 cm
Bottom*	7.2079 cm	N/A	7.2079 cm
$s_{20,w}$ (s)	$5.326 \times 10^{-13}$	$8.662 \times 10^{-14}$	$5.296 \times 10^{-13}$
$D_{20,w}$ (cm <sup>2</sup> /s)	$2.384 \times 10^{-7}$	$6.578 \times 10^{-7}$	$2.406 \times 10^{-7}$
$\sigma_k$	$4.471 \times 10^{-2}$	0.0	$4.470 \times 10^{-2}$
$\delta_k$	$1.215 \times 10^{-3}$	0.0	$7.727 \times 10^{-4}$
$\bar{v}$ (cm <sup>3</sup> /g)	0.55	Not known	0.55
MW (calc.) <sup>#</sup>	$1.2070 \times 10^5$	Not known	$1.1892 \times 10^5$
MW (theor.) <sup>§</sup>	$1.2074 \times 10^5$	Not known	$1.2074 \times 10^5$
MW ( $\Delta\%$ )	-0.033%	N/A	-1.5%

NA, Not applicable.

\* The bottom of the cell position listed refers to the rotor-stretching corrected value and epon centerpieces.

<sup>#</sup> Molecular weight based on fitted parameters according to Eq. 8.

<sup>§</sup> Molecular weight based on nucleic acid sequence.

question. As shown in Fig. 4 C, exclusion of the bottom 12% and the top 4% results in a single-component system, exhibiting concentration dependence in  $s$ . The concentration dependency parameter for  $s$ ,  $\sigma$ , can be estimated from the integral distribution plot (Fig. 4 D and Eq. 9), yielding a value of 0.076, which is a good estimate of the final fitted value of 0.045. Results are in good agreement with the expected molecular weight and the results from the fit in Fig. 3 (data are shown in Table 3).

## Two-ideal-component system

Velocity data from a mixture of a 196-bp DNA fragment and lysozyme were fitted to a two-component ideal model. The velocity data and the overlaid finite-element fit are shown in Fig. 5 A, and the residuals of the fit are shown in Fig. 5 B. The van Holde-Weischet extrapolation plot clearly shows the presence of two components (Fig. 5 C), present with a ratio of 43% (1.8S):57% (5.1S), as shown in the integral distribution plot (Fig. 5 D). Because of the reduced concentration of the DNA and the higher ionic strength of 0.2 M NaCl (compared to the experiment in Figs. 3 and 4), concentration dependence of  $s$  is not apparent, and therefore both components could be fitted to an ideal model (see also the distribution plot in Fig. 5 D). The increased ionic strength may also help to reduce nonspecific interactions between lysozyme and DNA by shielding the negative charges on the DNA molecules with Na<sup>+</sup> counterions. A fit with the SVEDBERG program (Philo, 1997) did not converge to produce a reasonable set of residuals for the Fujita solution model; however, the Fujita-McCosham solution produced residuals and a variance comparable to the finite-element residuals (data not shown). The calculated molecular weight of the DNA component is slightly higher than expected from the sequence of the DNA. The calculation of

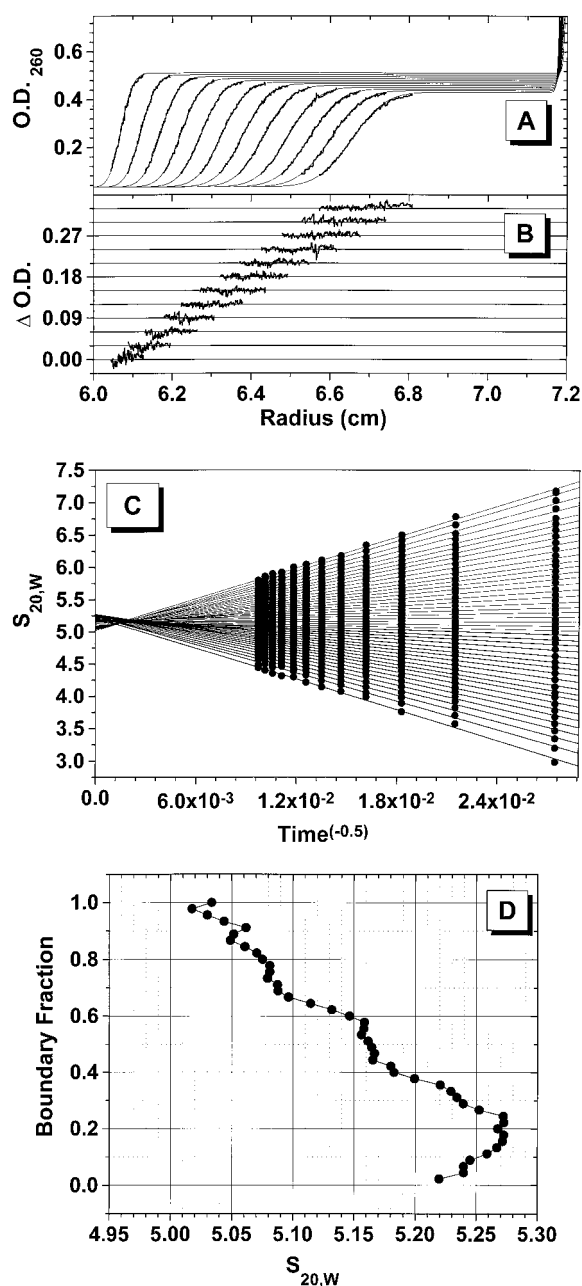


FIGURE 4 Finite-element fit of partial sedimentation velocity data of a 196-bp DNA fragment from the 5S ribosomal RNA gene of sea urchin in TE buffer. In this experiment only 86% of the boundary was fitted, with a 12% offset from the baseline. Run conditions were as follows: Rotor speed: 42 krpm; wavelength: 260 nm; temperature: 20.0°C. (A) Velocity data and overlaid finite-element solution for a double, ideal component system. (B) Residuals of finite-element fit in A. Residuals are shown with 0.03 absorption unit offsets, with the residuals for the first scan shown on the bottom (0.0), and residuals of the last scan shown at the top of the plot (0.33). (C) van Holde-Weischet extrapolation plot of data shown in A. Data display typical “cross-over,” indicative of concentration dependency in  $S$ . (D) Integral distribution plot of van Holde-Weischet extrapolation data shown in C.

the molecular weight of the lysozyme component resulted in values significantly different from those expected from the sequence-based value in fits by both models, with a differ-

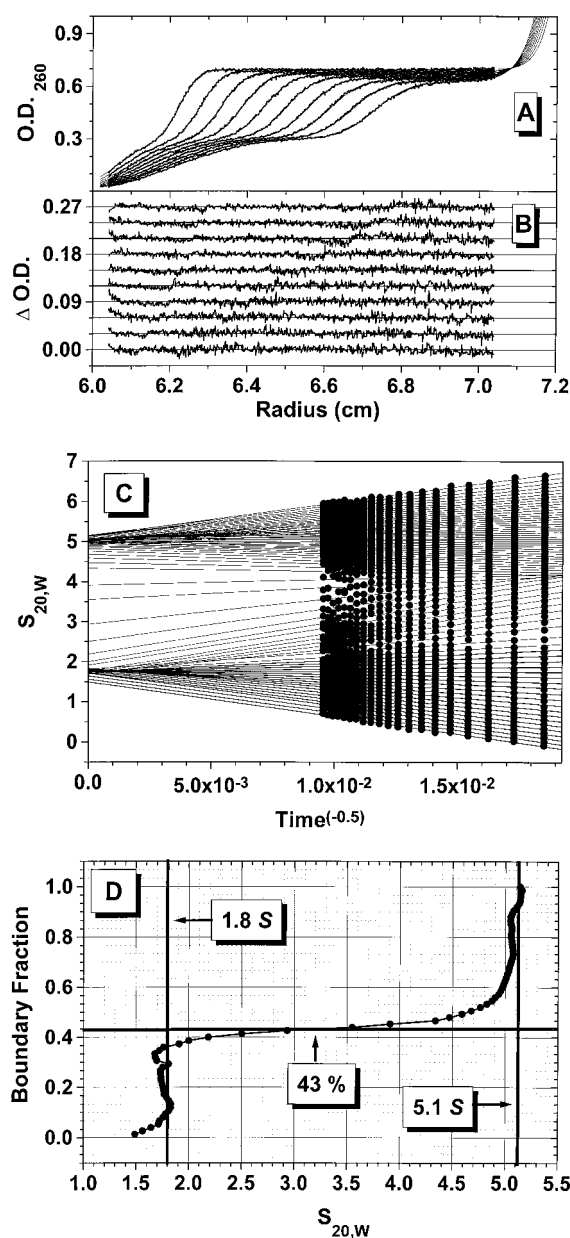


FIGURE 5 Finite-element fit of sedimentation velocity data of a 196-bp DNA fragment from the 5S ribosomal RNA gene of sea urchin mixed with chicken egg white lysozyme in TE/0.2 M NaCl. Run conditions were as follows: Rotor speed: 42 krpm; wavelength: 260 nm; temperature: 20.0°C. (A) Velocity data and overlaid finite-element solution for a double, ideal component system. (B) Residuals of finite-element fit in A. Residuals are shown with 0.03 absorption unit offsets, with the residuals for the first scan shown on the bottom (0.0), and residuals of the last scan shown at the top of the plot (0.27). (C) van Holde-Weischet extrapolation plot of data shown in A. Data clearly show the presence of two components, the “spread” of the fan plot is proportional to the diffusion coefficient, indicating a larger diffusion coefficient for the smaller sample. (D) Integral distribution plot of data shown in A. Solid vertical lines indicate the  $S$ -value estimates from the van Holde-Weischet analysis; the horizontal solid line indicates the concentration boundary between the two components.

ence of ~43% for the Fujita-McCosham solution of SVEDBERG (Philo, 1997), and a difference of ~20% for the finite-element solution. The difference in the molecular weight



between the sequence-derived and the measured values for the DNA component was around 10% for both models. From our analysis it is not clear what factors contribute to the difference between expected and calculated molecular weight for the DNA component. The higher sedimentation coefficient observed in this experiment when compared to the experiment shown in Fig. 4 may be due to conformational changes in the DNA when it is subjected to increased ionic strength. Because the two methods produce similar deviations in the expected DNA molecular weight, it is also possible that the result is a reflection of the quality of the data and that the data do not contain enough signal to allow accurate determination of either  $s$  or  $D$ . Possible reasons for the relatively large discrepancy in molecular weight between the expected and observed values for lysozyme include the fact that the sedimentation signal for this component was not very large. Because of the relatively slow centrifugation speed, the smaller sample did not sediment sufficiently fast to provide an adequate signal for a correct sedimentation coefficient determination. As pointed out by Demeler et al. (1997), faster centrifugation speeds generally result in higher accuracy in sedimentation coefficient determination. In a two-component system, however, a compromise must be made that allows for both components to be measured simultaneously. If the components sediment with significantly different rates, neither component can be measured at optimal speed. Consequently, the sedimentation coefficients for lysozyme derived from this experiment differ by  $\sim 15\%$  from independent measurements (experiment shown in Fig. 2, and others). The results are summarized in Table 4.

## DISCUSSION

### Technical and experimental advantages

We have looked at a variety of experimental model systems and assessed the suitability of the finite-element analysis method for sedimentation velocity data analysis. We have found that accurate results can be obtained for a large range

of experimental systems. The results from finite-element fits compare well to those obtained from other applicable direct-fitting methods, and reproduce the expected molecular weights derived from sequence information equally well or better than other methods. In situations where other approximate solutions fail, the finite-element analysis can prove useful by providing reliable information for systems that exhibit concentration dependency in  $s$  or  $D$  or for samples that are very small and can only be studied under equilibrium or approach-to-equilibrium conditions. The van Holde–Weischet method proved to be a useful preprocessor of the data by assisting the user in the selection of a proper model for the fit. Initial values provided by the van Holde–Weischet method generally were close to the final, fitted solutions, improving the speed of convergence substantially for cases where initial values are not known a priori. The use of partial boundaries for fitting allows the user to effectively improve the confidence in a subcomponent of a complex system by eliminating the requirement of simultaneously fitting multiple parameters and thus reducing the overall confidence in all parameters obtained. In addition, by eliminating additional parameters, the convergence properties of the nonlinear least-squares fit are improved. In comparison to other derivative-based nonlinear least-squares fitting algorithms, the DUD routine performs adequately, significantly reducing computing time, because fewer functional evaluations are required. It should also be mentioned that the finite-element method is ideally suited for data analysis of experimental data obtained from interference optics. The discretization of time shown in Eq. 11 requires that the solutions of the equation have to be calculated for closely spaced time intervals, regardless of the number of scans included in the experimental data. The interference optics hardware on the XL-A permits acquisition of scans closely spaced in time, thus allowing the user to take direct advantage of all intermediate calculations required by the finite-element simulation, and greatly en-

**TABLE 4 Results for finite-element fits of DNA mixed with lysozyme**

Sample Parameter	Finite-element two-component fit (component 1)	Finite-element two-component fit (component 2)	Svedberg two-component fit* (component 1)	Svedberg two-component fit* (component 2)
Baseline OD	$-1.435 \times 10^{-3}$	N/A	$-2.855 \times 10^{-2}$	N/A
$C_k$ (OD)	0.4055	0.3257	0.4009	0.3564
Meniscus	6.019 cm	N/A	6.019 cm	N/A
Bottom <sup>#</sup>	7.208 cm	N/A	N/A	N/A
$s_{20,w}$ (s)	$5.47 \times 10^{-13}$	$1.64 \times 10^{-13}$	$5.48 \times 10^{-13}$	$1.620 \times 10^{-13}$
$D_{20,w}$ (cm <sup>2</sup> /s)	$2.12 \times 10^{-7}$	$1.137 \times 10^{-6}$	$2.25 \times 10^{-7}$	$1.316 \times 10^{-6}$
$\bar{v}$ (cm <sup>3</sup> /g)	0.55	0.703 <sup>§</sup>	0.55	0.703 <sup>§</sup>
MW (calc.) <sup>¶</sup>	$1.3569 \times 10^5$	$1.2012 \times 10^4$	$1.317 \times 10^5$	$1.006 \times 10^4$
MW (theor.) <sup>  </sup>	$1.2074 \times 10^5$	$1.444 \times 10^4$	$1.2074 \times 10^5$	$1.444 \times 10^4$
MW ( $\Delta\%$ )	11.0%	-20.2%	9.1%	-43.5%

NA, Not applicable.

\* Svedberg Program (Philo, 1997) Fujita-McCosham solution, two-species model.

<sup>#</sup> The bottom of the cell position listed refers to the rotor-stretching corrected value and epon centerpieces.

<sup>§</sup> Partial specific volume as reported by Sober (1968).

<sup>¶</sup> Molecular weight based on fitted parameters according to Eq. 8.

<sup>||</sup> Molecular weight based on nucleic acid sequence (component 1) and protein sequence (Swiss Protein Database) (component 2).

hancing the usefulness of all additional data made available through the interference optics.

### Technical limitations of the method

A significant drawback of the finite-element method is the large amount of computing time required for fitting experimental data. Most of the CPU cycles required are used for the functional evaluations. For all but the fastest personal desktop computers equipped with a Unix operating system, this analysis may not yet be feasible until faster processors become available. We have done all fitting on a 200-MHz Pentium Pro personal computer, equipped with the Linux/Unix operating system, or on a DEC ALPHA. As a result, statistical analysis of the results has not been feasible, because of the large amount of computing time required for such modeling. Instead of exhaustive searches of the error domain to determine confidence intervals, we suggest the possibility of a bootstrap method, as discussed by Efron (1982). We plan to investigate this possibility in future work. Although the DUD fitting routine requires fewer functional evaluations for fitting than other least-squares methods, the algorithm tends to show signs of instability when a large number of parameters are floated simultaneously, or when the solution becomes locked in a local minimum. In such a case it often helps to fix some of the parameters and to repeat the fit in additional refinement cycles. We have not explored other optimization algorithms at this time, but as faster personal computers become available, it may be possible to use more robust routines based on gradient descent, which utilize derivatives for nonlinear least-squares optimization. For the case of very large or asymmetrical molecules with small diffusion coefficients ( $<1 \times 10^{-7} \text{ cm}^2/\text{s}$ ), the finite-element solution tends to become unstable near the bottom of the cell. This problem is amplified by the second-order term of the concentration dependency corrections, which include higher-order concentration terms. The error is caused by the almost discontinuous concentration change at the bottom of the cell. In such cases, an oscillation in the concentration gradient near the bottom of the cell appears in the early scans. It is intensified with each iteration in  $t$ , and propagates backward to the top of the cell. As shown in Fig. 6, a decrease in the radial discretization step size can effectively overcome this limitation. Because the smaller step size is not necessary for calculation throughout the cell, it is possible to simulate the bottom third of the cell with a step size small enough to avoid the oscillation effect, while conserving computing time (and memory) by using a larger radial step size for the rest of the cell. For small-molecular-weight samples, the method is applicable to any system that forms a boundary, even for molecules so small that under experimental conditions possible with the XL-A a velocity gradient cannot be formed, and only equilibrium gradients can be obtained. However, in such a case it is not practical to use the finite-element analysis because of the excessive computing

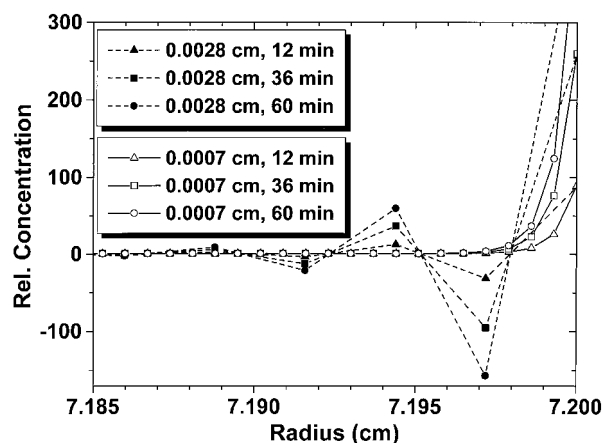


FIGURE 6 Shown here are simulated data to demonstrate the effect of the radial discretization step size on the stability of the finite-element solution near the bottom of the cell for cases of small diffusion coefficients. Simulation parameters were as follows: Model: single, ideal species with  $S = 11 \times 10^{-13} \text{ s}$ ,  $D = 5 \times 10^{-8} \text{ cm}^2/\text{s}$  and  $C_0 = 1$  (relative concentration); rotor speed: 30 krpm;  $\Delta t$ : 1 s. For clarity, only data at the bottom of the cell are shown. Open symbols refer to simulations with a step size of  $7 \times 10^{-4} \text{ cm}$ , filled symbols to a radial step size 4 times larger ( $2.8 \times 10^{-3} \text{ cm}$ ). Radial increments that are too large cause oscillations near the transition region that propagate back into the cell with each iteration.

time required to simulate the length of an equilibrium experiment, and fitting of experimental data with conventional methods is preferable. However, collecting scans before equilibrium is reached and analyzing those scans is feasible, because approach-to-equilibrium experiments are well suited to finite-element analysis of any molecule. However, in such a case the van Holde–Weischet method will not provide reliable estimates, and initial parameter guesses must be made by other means.

### Extensions of the method

Although the examples presented here focus on the use of the finite-element analysis for single and multiple (ideal or nonideal) component systems, the method is easily extended to include models such as reversibly associating monomer-polymer (Todd and Haschemeyer, 1981) and isomerization equilibria for the computation of equilibrium constants, or active enzyme sedimentation (Cohen and Claverie, 1975). The finite-element method provides the convenient feature of allowing easy modification by simply adding additional right-side terms in the discretized matrix equation (Eq. 14). Hence any new external force term can easily be included in the solution to model a great variety of systems.

### CONCLUSION

We have shown that the finite-element solution for various models of the Lamm equation in conjunction with the van Holde–Weischet analysis method and the DUD nonlinear least-squares fitting algorithm provides the researcher with

a novel tool for analyzing sedimentation velocity data for cases that have previously been inaccessible to methods that directly fit sedimentation velocity boundaries. This method can provide enhanced accuracy for cases where a concentration dependency of  $s$  or  $D$  is present, and where approach-to-equilibrium conditions are inevitable (i.e., small molecules with very small  $s$  values and large diffusion coefficients). The finite-element solution to the Lamm equation does not incorporate any assumptions or approximation terms, and hence provides a rigorous mathematical model of the sedimentation process. For cases that do not fall into these two categories, other direct fitting methods (Philo, 1997; Behlke and Ristau, 1997) will provide equivalent results and are preferable because of the convenience of reduced computational requirements.

We thank Kensal E. van Holde and Jeffrey C. Hansen for helpful discussions.

This work was supported in part by National Institutes of Health grant GM45916.

## REFERENCES

- Behlke, J., and O. Ristau. 1997. Molecular mass determination by sedimentation velocity experiments and direct fitting of the concentration profiles. *Biophys. J.* 72:428–434.
- 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. Solutions of systems of complete Lamm equations. *Biopolymers*. 14:1685–1700.
- Cohen, R., and J. M. Claverie. 1975. Sedimentation of generalized systems of interacting particles. II. Active enzyme centrifugation—theory and extensions of its validity range. *Biopolymers*. 14:1701–1716.
- Cox, D. J. 1965a. Computer simulation of sedimentation in the ultracentrifuge. I. Diffusion. *Arch. Biochem. Biophys.* 112:249–258.
- Cox, D. J. 1965b. Computer Simulation of sedimentation in the ultracentrifuge. II. Concentration independent sedimentation. *Arch. Biochem. Biophys.* 112:259–266.
- Cox, D. J. 1965c. Computer simulation of sedimentation in the ultracentrifuge. III. Concentration dependent sedimentation. *Arch. Biochem. Biophys.* 112:230–239.
- Demeler, B. 1992. New methods for sedimentation and diffusion analysis of macromolecular structure. Ph.D. thesis. Oregon State University, Corvallis, OR.
- Demeler, B., H. Saber, and J. 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. 1969. Numerical solutions of the Lamm equation. V. Band centrifugation. *Ann. N.Y. Acad. Sci.* 164:33–51.
- Efron, B. 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. Regional Conference Series in Applied Mathematics. SIAM, Philadelphia.
- Georgel, P., B. Demeler, C. Terpening, M. R. Paule, and K. E. van Holde. 1993. Binding of the RNA polymerase I transcription complex to its promoter can modify positioning of downstream nucleosomes assembled in vitro. *J. Biol. Chem.* 268:1947–1954.
- Faxén, H. 1929. Über eine Differentialgleichung aus der physikalischen Chemie. *Ark. Mat. Astr. Fys.* 21B:1–6.
- Fujita, H. 1962. Mathematical Theory of Sedimentation Analysis. Academic Press, New York.
- Fujita, H. 1975. Foundations of Ultracentrifugal Analysis. John Wiley & Sons, Inc., New York.
- Fujita, H., and V. J. MacCosham. 1959. Extension of sedimentation velocity theory to molecules of intermediate sizes. *J. Chem. Phys.* 30:291–298.
- Holladay, L. A. 1979. An approximate solution to the Lamm equation. *Biophys. Chem.* 10:187–190.
- Holladay, L. A. 1980. Simultaneous rapid estimation of sedimentation coefficient and molecular weight. *Biophys. Chem.* 11:303–308.
- Johnson, M. L., and L. M. Faunt. 1992. Parameter estimation by least-squares methods. *Methods Enzymol.* 210:1–37.
- Lamm, O. 1929. Die Differentialgleichung der Ultrazentrifugierung. *Ark. Mat. Astron. Fys.* 21B:1–4.
- Laue, T. M., B. D. Shah, T. M. Ridgeway, and S. L. Pelletier. 1992. Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. S. E. Harding, A. J. Rowe, and J. C. Horton, editors. Royal Society of Chemistry, Cambridge, England. 90–125.
- Philo, J. S. 1994. Measuring sedimentation, diffusion, molecular weight of small molecules by direct fitting of sedimentation velocity concentration profiles. In *Modern Analytical Ultracentrifugation*. T. M. Schuster and T. M. Laue, editors. Birkhäuser, Boston. 156–170.
- Philo, J. S. 1997. An improved function for fitting sedimentation velocity data for low-molecular-weight solutes. *Biophys. J.* 72:435–444.
- Ralston, M. L., and R. I. Jennrich. 1978. DUD, a derivative-free algorithm for nonlinear least squares. *Technometrics*. 20:7–14.
- Sober, H. 1968. The Handbook of Biochemistry and Molecular Biology. Chemical Rubber Co., Cleveland, OH.
- Stafford, W. 1992. Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Anal. Biochem.* 203:295–301.
- Todd, G. P., and R. H. Haschemeyer. 1981. General solution to the inverse problem of the differential equation of the ultracentrifuge. *Proc. Natl. Acad. Sci. USA*. 78–11:6739–6743.
- van Holde, K. E., and W. O. Weischet. 1978. Boundary analysis of sedimentation velocity experiments with monodisperse and paucidisperse solutes. *Biopolymers*. 17:1387–1403.
- Zienkiewicz, O. C. 1971. The Finite Element Method in Engineering Science. McGraw-Hill, London.