

A Method for Directly Fitting the Time Derivative of Sedimentation Velocity Data and an Alternative Algorithm for Calculating Sedimentation Coefficient Distribution Functions

John S. Philo

Alliance Protein Laboratories, 3957 Corte Cancion, Thousand Oaks, CA 91360; jphilo@mailway.com

Received August 27, 1999

The time derivative method for deriving the sedimentation coefficient distribution, $g(s^*)$, from sedimentation velocity data that was developed by Walter Stafford has many advantages and is now widely used. By fitting Gaussian functions to the $g(s^*)$ distribution both sedimentation and diffusion coefficients (and therefore molecular masses) for individual species can be obtained. However, some of the approximations used in these procedures limit the accuracy of the results. An alternative approach is proposed in which the dc/dt data are fitted rather than $g(s^*)$. This new approach gives improved accuracy, extends the range to sedimentation coefficients below 1 S, and enhances resolution of multiple species. For both approaches the peaks from individual species are broadened when the data cover too wide a time span, and this effect is explored and quantified.

An alternative algorithm for calculating $\hat{g}(s^*)$ from the dc/dt curves is presented and discussed. Rather than first averaging the dc/dt data for individual scan pairs and then calculating $\hat{g}(s^*)$ from that average, the $\hat{g}(s^*)$ distributions are calculated for every scan pair and then subsequently averaged. This alternative procedure yields smaller error bars for $g(s^*)$ and somewhat greater accuracy for fitted hydrodynamic properties when the time span becomes large.

Key Words: sedimentation velocity; analytical ultracentrifugation; time derivative analysis; sedimentation coefficient; diffusion coefficient; numerical methods; least-squares fitting.

In recent years a new method of sedimentation velocity analysis pioneered by Walter Stafford at Boston Biomedical Research Institute has proven very useful and is now widely used. This time-derivative analysis or “ dc/dt ” methodology allows derivation of the sedimentation coefficient distribution function, $\hat{g}(s^*)$, by subtracting pairs of scans of concentration versus radius in the centrifuge cell, rather than from the dc/dr data provided by a Schlieren optical system or

dc/dr computed by numerically differentiating c vs. r data (1, 2) A great virtue of this method is that systematic, time-independent noise in the data, such as baseline distortions from the windows of the centrifuge cell, is completely removed. This fact, plus the ability to average together the results from many pairs of scans, provides a tremendous improvement in signal/noise and thus sensitivity. Indeed, sedimentation velocity studies on proteins are now routinely run at concentrations below 100 $\mu\text{g/ml}$ instead of the 2-10 mg/ml common in the past, thus allowing many more applications and avoiding complications in interpretation due to thermodynamic and hydrodynamic non-ideality at high concentrations.

The $\hat{g}(s^*)$ distribution represents the concentration distribution at the time of the analysis, when different components will have undergone radial dilution to a different extent. Thus for samples containing a single species or a mixture of independently-sedimenting components $\hat{g}(s^*)$ is usually corrected to $g(s^*)$, the apparent distribution when the sample was loaded. Another important feature of the time derivative method is that it is possible to fit individual peaks in the $g(s^*)$ distribution to Gaussian functions and thereby derive both the sedimentation coefficient (from the center position) and the diffusion coefficient (from the width of the Gaussian) for individual species. (3, 4) Further, with both the sedimentation and diffusion coefficients known, one can obtain the molecular mass using the Svedberg equation.

While tremendously useful, these methodologies have their shortcomings, and the purpose of this paper is to propose some revised algorithms to address those problems. As noted by Stafford, (5) there are difficulties in applying the dc/dt method to peptides and proteins in the 5-20 kDa range due to their high diffusion coefficients. Indeed, in an attempt to study a ~4.5 kDa peptide in our laboratory it was found that although the peptide gave a distinct peak around 0.8 S in the dc/dt curve (Figure 1A), after transformation there was no peak in the $g(s^*)$ distribution (Fig. 1B), and it was impossible to obtain either the

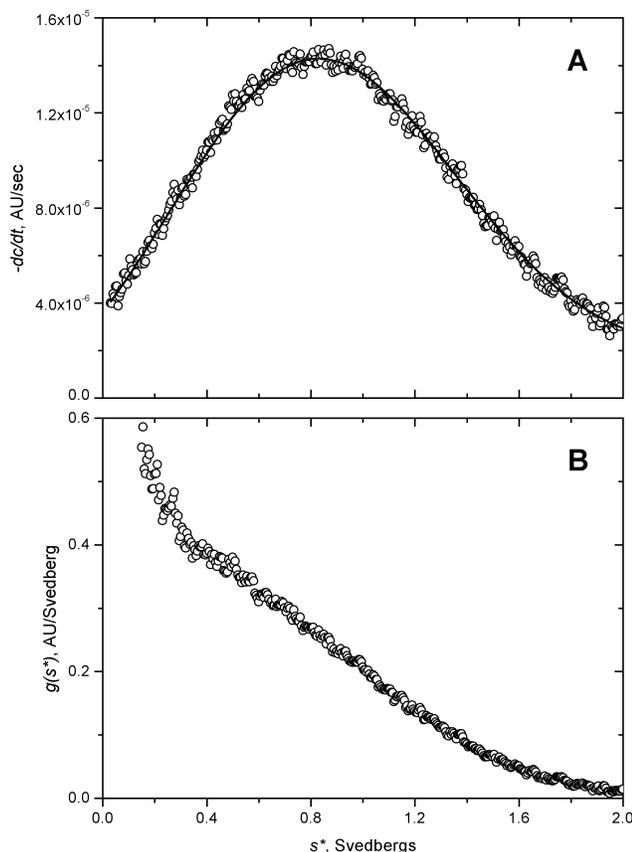


Fig. 1 An example of applying time-derivative analysis to a small peptide (~4.5 kDa) run at 60000 rpm. The average dc/dt data from 10 scans at 10 min intervals are shown in panel A. Although a clear peak at $s^* \sim 0.8$ is present in dc/dt , no corresponding peak can be discerned in the $g(s^*)$ distribution from these data, shown in panel B (and this is true even if the dc/dt data are truncated before transformation anywhere within the region from 0.05 to 0.5 S). The solid curve in Panel A is the result of a least-squares fit to the dc/dt data (see *Results*).

sedimentation or diffusion coefficient from that distribution. Further, other studies in our laboratory suggested that the accuracy of the diffusion coefficients and masses derived by fitting $g(s^*)$ distributions for single proteins of $> \sim 100$ kDa can give errors of 8% or more, somewhat larger than the 2-5% accuracy reported by Stafford (4), and significantly worse than can be obtained by directly fitting the c vs. r data using either approximate (6, 7) or finite-element numerical solutions (8) of the Lamm equation.

It is certainly true that errors in mass of this magnitude are trivial if one merely wants to ascertain whether the protein is a monomer or a dimer in solution. However, an important application for sedimentation velocity analysis is to provide tests of sample homogeneity. In the past the criterion for

homogeneity was largely limited to the absence of resolved boundaries from minor species. Today the modern data analysis methods that allow derivation of apparent diffusion coefficients from the widths of the boundaries, when combined with an accurate value for the monomer mass from the amino acid sequence or from mass spectrometry, permit a much more rigorous test of homogeneity. Given that heterogeneity arising either from contaminants, irreversible aggregates, self-association, and/or variations in conformation will always produce broader boundaries, if the mass obtained from the ratio of sedimentation coefficient to apparent diffusion coefficient accurately agrees with the monomer mass (or an integer multiple of that mass when the native state is an oligomer), then the sample must be highly homogeneous. This type of test has proven quite useful, for example, in characterizing proteins used as pharmaceuticals. However, this test begs the question “how much error in mass is tolerable before concluding that heterogeneity is present?” Therefore before using the dc/dt methodology for such tests it is critical to understand the magnitudes and sources of error. These considerations led to attempts to reduce the magnitude of these errors and to developing the new procedure described herein where the dc/dt data rather than the $g(s^*)$ data are fitted.

Another difficulty that often arises in applying the dc/dt methodology, particularly for studies at low concentrations, is that in order to obtain sufficient signal/noise a large number of scans must be used. As the time span covered by these scans increases, eventually the boundary movement becomes too large and the approximation that $\Delta c/\Delta t \cong dc/dt$ breaks down, resulting in broadening of the peaks in the dc/dt and $g(s^*)$ curves. Walter Stafford has proposed a “rule of thumb” for estimating the maximum time span that can be used before the peak broadening becomes excessive (<http://www.bbri.org/dcdt/Rule.pdf>), and here we explore in more detail the size of the errors in diffusion coefficients which are induced by this mechanism for both the old method of fitting to $g(s^*)$ curves and the new method of fitting to dc/dt .

What, then, is limiting the accuracy? For the low masses the dominant error arises from the high diffusion coefficient, and occurs during the mathematical transformation from dc/dt to $g(s^*)$. This transformation requires dividing each value of dc/dt by the corresponding s value. However, because of diffusion the values of dc/dt from a single species are non-zero over a broad range of sedimentation coefficients, and consequently the dc/dt values are often being divided by s values that differ significantly from the true value. This effect tends to shift the peak

TABLE 1
Comparisons of different fitting methods on simulated data for single species^a

Properties of Simulated Species and Rotor Speed	Results from fitting to $g(s^*)$	Results from fitting to dc/dt	Results from direct boundary fitting using SVEDBERG
$s = 1.2$ S; $D = 13.5$ F ^b (~8 kDa) 60000 rpm	$s = 1.079$ S (-10.1%) $D = 14.16$ F (+4.9%) $c_o = 1.145$ (+14.5%) $M/M_t^c = 0.857$ (-14.3%)	$s = 1.191$ S (-0.8%) $D = 13.54$ F (+0.3%) $c_o = 1.003$ (+0.3%) $M/M_t = 0.990$ (-1.0%)	$s = 1.199$ S (-0.1%) $D = 13.81$ F (+2.3%) $c_o = 1.010$ (+1.0%) $M/M_t = 0.977$ (-2.3%)
$s = 2$ S; $D = 10$ F (~18 kDa) 60000 rpm	$s = 1.92$ S (-4.0%) $D = 9.59$ F (-4.1%) $c_o = 1.052$ (+5.2%) $M/M_t = 1.001$ (+0.1%)	$s = 1.991$ S (-0.5%) $D = 10.11$ F (+1.1%) $c_o = 1.004$ (+0.4%) $M/M_t = 0.985$ (-1.5%)	$s = 1.993$ S (-0.4%) $D = 9.97$ F (-0.3%) $c_o = 0.999$ (-0.1%) $M/M_t = 0.999$ (-0.1%)
$s = 3$ S; $D = 7$ F (~39 kDa) 60000 rpm	$s = 2.947$ S (-1.8%) $D = 6.60$ F (-5.7%) $c_o = 1.023$ (+2.3%) $M/M_t = 1.042$ (+4.2%)	$s = 2.995$ S (-0.2%) $D = 7.05$ F (+0.7%) $c_o = 1.003$ (-0.3%) $M/M_t = 0.991$ (-0.9%)	$s = 2.995$ S (-0.2%) $D = 7.01$ F (+0.1%) $c_o = 1.000$ (+0.0%) $M/M_t = 0.997$ (-0.3%)
$s = 5$ S; $D = 6$ F (~75 kDa) 60000 rpm	$s = 4.965$ S (-0.7%) $D = 5.56$ F (-7.3%) $c_o = 1.009$ (+0.9%) $M/M_t = 1.072$ (+7.2%)	$s = 4.996$ S (-0.1%) $D = 6.04$ F (+0.7%) $c_o = 1.000$ (+0.0%) $M/M_t = 0.993$ (-0.7%)	$s = 4.995$ S (-0.1%) $D = 6.03$ F (+0.5%) $c_o = 1.000$ (+0.0%) $M/M_t = 0.994$ (-0.6%)
$s = 5$ S; $D = 6$ F (~75 kDa) 40000 rpm	$s = 4.910$ S (-1.8%) $D = 5.62$ F (-6.3%) $c_o = 1.024$ (+0.9%) $M/M_t = 1.048$ (+4.8%)	$s = 4.991$ S (-0.2%) $D = 6.00$ F (+0.0%) $c_o = 1.000$ (+0.0%) $M/M_t = 0.998$ (-0.2%)	$s = 4.990$ S (-0.2%) $D = 6.00$ F (+0.0%) $c_o = 1.000$ (+0.0%) $M/M_t = 0.998$ (-0.2%)
$s = 6.2$ S; $D = 3.9$ F (~144 kDa) 45000 rpm	$s = 6.160$ S (-0.6%) $D = 3.60$ F (-7.7%) $c_o = 1.009$ (+0.9%) $M/M_t = 1.076$ (+7.6%)	$s = 6.195$ S (-0.1%) $D = 3.92$ F (+0.5%) $c_o = 1.000$ (+0.0%) $M/M_t = 0.994$ (-0.6%)	$s = 6.195$ S (-0.1%) $D = 3.91$ F (+0.3%) $c_o = 1.000$ (+0.0%) $M/M_t = 0.996$ (-0.4%)
$s = 9.0$ S; $D = 5.5$ F (~148 kDa) 45000 rpm	$s = 8.945$ S (-0.8%) $D = 5.06$ F (-8.0%) $c_o = 1.008$ (+0.8%) $M/M_t = 1.080$ (+8.0%)	$s = 8.994$ S (-0.1%) $D = 5.50$ F (+0.0%) $c_o = 1.000$ (+0.0%) $M/M_t = 0.999$ (-0.1%)	$s = 8.993$ S (-0.1%) $D = 5.50$ F (+0.0%) $c_o = 1.000$ (+0.0%) $M/M_t = 0.999$ (-0.1%)

^aIn all cases the simulated data (without added random noise) were evaluated at a time when the boundary had advanced to the midpoint of the cell, and the simulations were for a loading concentration, c_o , of 1.

^b1 F (Fick) equals 10^{-7} cm²s⁻¹

^cthe mass ratio M/M_t is the ratio of the apparent mass to the true value that went into the simulation (based on s/D)

in $g(s^*)$ to lower sedimentation coefficients, and to distort the shape increasingly as s approaches zero. Since this effect is inherent in the transformation, it appeared that the only way to avoid this problem and obtain higher accuracy for low mass species would be to work directly with the untransformed dc/dt data.

For the larger masses this distortion of the data is sharply reduced, so the fact that the errors in D again grow larger as mass is increased suggests a second source of error. At higher masses the errors in D appear to arise from the use of a Gaussian curve to fit the data,

which is based on the Faxén approximate solution to the Lamm equation (4). This Faxén approximate solution neglects the effects of restricted diffusion near the meniscus early in the run, and has previously been shown to underestimate D for higher mass proteins when used for whole-boundary analysis (7).

Thus in an attempt to avoid both these problems it was decided to attempt to fit directly to the dc/dt data, thus retaining all the noise reducing advantages of the dc/dt computation, but using a fitting function that would more accurately describe the theoretical dc/dt

curves for individual species over a wide range of masses. One such approximate solution of the Lamm equation, the so-called modified Fujita-MacCosham function (7), was previously developed in our laboratory for use in whole-boundary analysis, and has been implemented and used in the velocity analysis program SVEDBERG. This analytical function describes the concentration distribution across the cell at any time in the run, and is valid for all regions except near the base of the cell where solutes accumulate, at times after the meniscus is > 50% depleted. By taking its analytical time derivative (Eq. 3 in *Methods*), one should therefore get an analytical form for dc/dt that should be valid under the same conditions. A Visual Basic program to compute dc/dt in this way was written and combined with a non-linear least square fitting routine to estimate s and D for up to five individual components in a mixture. In some cases it may be appropriate or necessary to deliberately use a time span that will cause peak broadening in order to obtain better signal/noise. This situation may arise, for example, in studying samples containing a broad distribution of species, because in order to obtain detectable boundary movement for the slowly sedimenting components the movement of the fast components is likely to be too large. In theory, for scans at constant time intervals increasing the number of scans used in the analysis, N , will improve the signal/noise of the $g(s^*)$ distribution approximately as $N^{3/2}$. (The 3/2 exponent arises because the concentration difference Δc grows approximately linearly with time, in addition to the statistical improvement by $N^{1/2}$). However, in practice we have found that the error bars on the $g(s^*)$ distribution often do not improve with N nearly as much as expected, and may even grow larger as more scans are added (particularly in regions near the peaks of individual species). Further investigation has now revealed that this effect is, in part, a consequence of the fact that as the time span grows longer, radial dilution significantly reduces the amplitude of the dc/dt curves for individual scan pairs, leading to systematic variations among the curves being averaged, and this systematic effect can become quite significant compared to the random noise. To address this problem, an alternative method of averaging together the results from the individual scan pairs was developed, and this new algorithm is presented here and compared to the original procedure developed by Stafford.

METHODS

Sedimentation velocity data were obtained using absorbance scans at 280 nm in a Beckman Optima XL-A analytical centrifuge. Numerical simulations of sedimentation velocity experiments were done using a Claverie finite-element routine, as described previously (7). In order to obtain high accuracy these simulations used 800-1600 radial points and time steps of 0.2-0.5 seconds. The calculation of dc/dt and $g(s^*)$ distributions was done using a Visual Basic program named DCDT+ that implements the algorithms described by Walter Stafford (2) as well as the new algorithm described below.¹ This implementation was compared to the DOS program DCDT supplied by the National Analytical Ultracentrifugation Facility and gives equivalent results. Non-linear least squares fitting employed a modified Gauss-Newton method, as described previously (6). Weight-, z -, and $z+1$ -average sedimentation coefficients were calculated from the $g(s^*)$ distributions using the definitions

$$s_w = \frac{\int s \cdot g(s) \cdot ds}{\int g(s) \cdot ds}; s_z = \frac{\int s^2 \cdot g(s) \cdot ds}{\int s \cdot g(s) \cdot ds}; s_{z+1} = \frac{\int s^3 \cdot g(s) \cdot ds}{\int s^2 \cdot g(s) \cdot ds} \quad (1)$$

The uncertainties σ_s in these quantities were calculated from the standard deviations of the points in the $g(s^*)$ distribution, σ_g , through standard propagation of error formulae, giving

$$\sigma_s^2 = \frac{\int s^{2(n+1)} \cdot \sigma_g^2 \cdot ds}{\left(\int s^n \cdot g(s) \cdot ds\right)^2} + \frac{\left(\int s^{n+1} \cdot g(s) \cdot ds\right)^2 \int s^{2n} \cdot \sigma_g^2 \cdot ds}{\left(\int s^n \cdot g(s) \cdot ds\right)^4} \quad (2)$$

where $n = 0, 1, 2$ correspond to s_w , s_z , and s_{z+1} , respectively.

An analytical expression for dc/dt that can be used in fitting

The function used in fitting the dc/dt vs. s^* data is the time derivative of the “modified Fujita-MacCosham function” (7), which is given by Eq. 3 (next page).

An alternative algorithm to calculate $\hat{g}(s^)$ and the average dc/dt curve*

For calculating $\hat{g}(s^*)$ the alternative algorithm first calculates a $\hat{g}(s^*)$ curve from the dc/dt curve from each pair of scans using a procedure identical to that in the Stafford algorithm (2), including three rounds of iterative correction for the contributions of the plateau

¹ The program description and instructions for downloading can be found at <http://www.jphilo.mailway.com/dcdt+.htm>.

$$\frac{dc}{dt}(s^*, t) = \frac{c_o e^{-2s\omega^2 t}}{2} \left\{ \begin{aligned} & -2s\omega^2 \left[1 - \operatorname{erf}\left(\frac{2s\omega^2 t - z}{4\sqrt{D}}(1 + 2\alpha s\omega^2 t)\right) - \frac{2}{\sqrt{\pi}}\left(\frac{s\omega^2 t}{\sqrt{D}}\right) \exp\left(\frac{-(2s\omega^2 t - z)^2}{16D}(1 + 4\beta D)\right) \right] \\ & + \left(1 + \frac{2s^2\omega^4 t^2}{D} + \frac{s\omega^2 z}{D}\right) \left(1 - \operatorname{erf}\left(\frac{s\omega^2 t}{\sqrt{D}} + \frac{z}{4\sqrt{D}}\right)\right) \exp\left(\frac{s\omega^2 tz}{2D}\right) \\ & - \frac{2}{\sqrt{\pi}} \exp\left(\frac{-(2s\omega^2 t - z)^2}{16D}(1 + 2\alpha s\omega^2 t)^2\right) \left(\frac{s\omega^2}{2\sqrt{D}} + \frac{2s^2\omega^4 t}{\sqrt{D}}\right) \\ & - \frac{2}{t\sqrt{\pi}} \left(\frac{s\omega^2 t}{\sqrt{D}}\right) \exp\left(\frac{-(2s\omega^2 t - z)^2}{16D}(1 + 4\beta D)\right) \left(1 - (1 + 4\beta D)\left(\frac{2s\omega^2 t - z}{4\sqrt{D}}\right)\left(\frac{s\omega^2 t}{\sqrt{D}}\right)\right) \\ & + \frac{s\omega^2}{\sqrt{D}} \left[\frac{2s\omega^2 t}{\sqrt{D}} \left(1 + \frac{s^*}{s}\right) \left(1 - \operatorname{erf}\left(\frac{s\omega^2 t}{\sqrt{D}} + \frac{z}{4\sqrt{D}}\right)\right) \exp\left(\frac{s\omega^2 tz}{2D}\right) \right. \\ & \left. - \frac{1}{\sqrt{\pi}} \left(1 + \frac{2s^2\omega^4 t^2}{D} + \frac{s\omega^2 z}{D}\right) \exp\left(-\left(\frac{s\omega^2 t}{\sqrt{D}} + \frac{z}{4\sqrt{D}}\right)^2\right) \right. \\ & \left. + \frac{s\omega^2 t}{s^* \sqrt{D}} \left(1 + \frac{2s^2\omega^4 t^2}{D} + \frac{s\omega^2 z}{D}\right) \left(1 - \operatorname{erf}\left(\frac{s\omega^2 t}{\sqrt{D}} + \frac{z}{4\sqrt{D}}\right)\right) \exp\left(\frac{s\omega^2 tz}{2D}\right) \right] \end{aligned} \right\} \quad (3)$$

Equation 3, giving dc/dt for a single species with sedimentation coefficient s , diffusion coefficient D , loading concentration c_o at elapsed time t during a run at angular velocity ω , where $\operatorname{erf}()$ is the error function, z is defined as $2s^*\omega^2 t$, and α and β are empirically-defined numerical constants ($\alpha = .2487$, $\beta = 2$).

region to dc/dt . Each of these individual $\hat{g}(s^*)$ curves is then averaged together to compute the final average $\hat{g}(s^*)$ distribution, with error bars calculated from the deviations among the curves used in computing that average.

This final $\hat{g}(s^*)$ distribution is then used to calculate a corresponding average dc/dt distribution. Because in general parts of the sedimentation coefficient range of the dc/dt data are excluded in calculating $\hat{g}(s^*)$, in reversing the usual procedure some way of supplying the full range of the dc/dt data is needed. Hence the first step is to calculate a preliminary average dc/dt in the conventional way by interpolating the dc/dt curve from each pair of scans onto a uniform grid of sedimentation coefficients. Next, those data points in this average dc/dt curve beginning at s_0 , the lowest sedimentation coefficient covered by the $\hat{g}(s^*)$ curve, and ending at the highest sedimentation coefficient included in the $\hat{g}(s^*)$ curve, are replaced by back-calculating them from $\hat{g}(s^*)$ using the formula:

$$\frac{dc}{dt}(s) = \frac{s \cdot \hat{g}(s)}{t_m} + 2\omega^2 \int_{s_0}^s s' \cdot \hat{g}(s') \cdot ds' \quad (4)$$

where t_m is the harmonic mean of the effective run times of all the scans and ω is the rotor angular

frequency. The integral on the right hand side of Eq. 4 adds back the contributions to dc/dt from the plateau region for all species having sedimentation coefficients less than the value being computed. The error bar for each point that is computed by Eq. 4 is calculated by multiplying the error bar in $\hat{g}(s^*)$ by s/t_m , neglecting the very small contribution from the uncertainty of the integral term in Eq. 4.

RESULTS

As discussed earlier, it appeared experimentally that the accuracy of hydrodynamic parameters derived by fitting $g(s^*)$ curves obtained by the dc/dt method is compromised both when the mass is very low and the boundaries are very broad, and when the mass is moderately high and the boundaries are fairly narrow. To confirm this behavior noise-free data sets were produced by the Claverie finite-element method to simulate five globular proteins ($f/f_o < 1.2$) with masses ranging from ~ 8 to ~ 150 kDa, plus a 145 kDa asymmetric protein ($f/f_o \sim 1.6$) with properties similar to those of immunoglobulins. Ten scans spaced closely in time (to avoid any boundary broadening) from these simulated data were analyzed by the dc/dt method and the resulting $g(s^*)$ distributions fitted to a single

Gaussian function. The results tabulated in Table 1 confirm that the diffusion coefficients, D , are in error by +4.9% to -8.0%, the sedimentation coefficients, s , by -10.1% to -0.7%, and the loading concentrations, c_o , by +14.5% to +0.8% over this range of proteins.

The errors in both sedimentation and diffusion coefficients often have the same sign, and thus generally there is a compensation effect in computing the apparent mass, M , from s/D , except for the lowest mass species. Overall the errors in M range from -14.3% to +8.0%. For the ~75 kDa species simulations were run at two different rotor speeds, and this confirmed that the error in D and M is worse for the relatively sharp boundaries produced at 60000 rpm. Comparing the two differently shaped proteins with mass ~146 kDa shows that they give essentially equivalent errors; *i.e.*, the errors appear to be a function of mass alone. (Note, however, that the error size is fairly sensitive to the position of the boundary within the cell.)

The results of applying this function to fit directly to the dc/dt data from the same simulated data sets are presented in Table 1. For purposes of comparison, these same scans were also fitted by whole boundary analysis using the program SVEDBERG (7), and these results are also shown in Table 1. Fitting to dc/dt in this manner produces a 4 to 10-fold improvement in accuracy for s , D , and c_o relative to fitting $g(s^*)$ with Gaussians over the entire group of simulations, and the absolute accuracy of s , D , and M is always better than 2%. Fitting to dc/dt gives accuracy comparable to, but slightly lower than, direct analysis using SVEDBERG.

This new approach was also tested on the ~4.5 kDa peptide data shown in Fig. 1 for which the $g(s^*)$ data could not be analyzed. Fitting the dc/dt data gave the continuous curve shown in Fig. 1A and returned $s = 0.678$ S [95% confidence 0.674 to 0.681] and $D = 13.89$ [13.72 to 14.06] F.² These values imply a mass that matches the sequence mass well within the ~4% uncertainty in mass arising from the ~1% uncertainty in the calculated partial specific volume. It is also worth noting that for such slowly sedimenting species the position of the maximum in the dc/dt curve does not give a reliable estimate of the true sedimentation coefficient.

Use for multi-species fitting

While this new approach appears to work very well for single species, there is potentially a serious problem in applying it to multi-species analysis. In the $g(s^*)$

distribution each species makes a contribution only over a finite range of sedimentation coefficients. In contrast, each species produces a dc/dt curve with a peak at the corresponding sedimentation coefficient, but also having a flat “tail” or “shelf” on the right side of this peak that remains constant and non-zero to $s = \infty$. This tail arises from the contribution this species makes to the decrease in plateau concentration. This property of the dc/dt curves might make it rather difficult to resolve minor components in the presence of a more slowly sedimenting major component.

To test the new approach several simulations were done of mixtures of a monomer with a small fraction of (non-interacting) dimer, for monomers having properties approximately like those of ovalbumin or immunoglobulin. The dc/dt and $g(s^*)$ curves were computed from closely spaced scans at a time in the run just prior to the disappearance of the plateau region (a time which gives optimal resolution) and then fitted. For these simulations a small amount of random noise was added to the simulated data, comparable to the intrinsic noise in the interference optical system. This amount of noise is sufficient to allow a comparison of

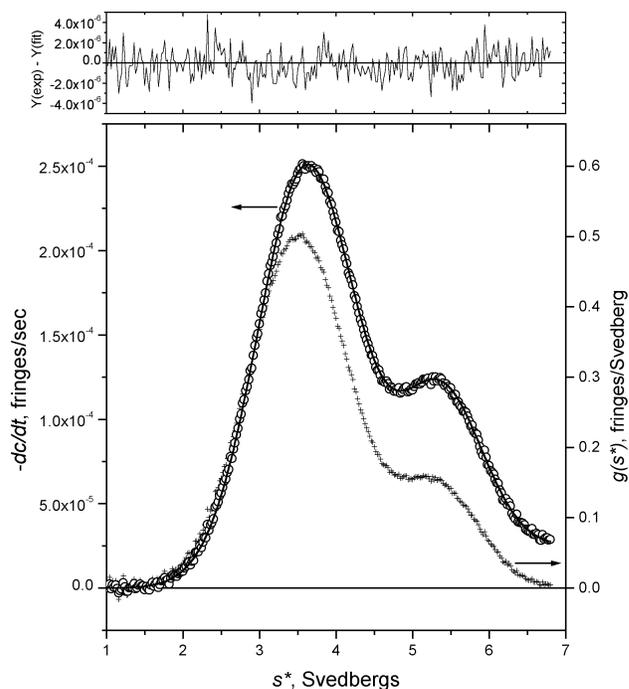


Fig. 2. Fitting of the dc/dt data from a simulation for an ovalbumin sample containing 20% dimer (see Table 2). The lower panel shows the dc/dt data (circles) and the fitted curve (solid line). The $g(s^*)$ data from the same simulation are also shown (+) to illustrate the difference in the relative amplitude of the dimer signal. For the sake of clarity only every fourth data point is shown. The upper panel shows the residuals from the fit.

² Abbreviations used: F, Fick, 1 F = 10^{-7} cm²s⁻¹

TABLE 2
Comparison of multi-species analysis of simulated data by fitting either $g(s^*)$ or dc/dt data^a

Mixture Simulated	Species Composition	Results from Fitting $g(s^*)$ Data	Results from Fitting dc/dt Data
ovalbumin + 20% dimer; 60,000 rpm; 30 scans	$s = 3.55$ S $D = 7.89$ F $c_o = 0.80$	$s = 3.507$ [3.506 to 3.508] S (-1.2%) $D = 7.47$ [7.43 to 7.51] F (-5.3%) $c_o = 0.822$ [.820 to .824] (+2.2%)	$s = 3.546$ [3.545 to 3.547] S (-0.1%) $D = 7.94$ [7.89 to 7.98] F (+0.6%) $c_o = 0.801$ [.799 to .803] (+0.1%)
	$s = 5.325$ S $D = 5.918$ F $c_o = 0.20$	$s = 5.326$ [5.321 to 5.330] S (+0.0%) $D = 5.03$ [4.96 to 5.10] F (-15.0%) $c_o = 0.193$ [.192 to .194] (-0.7%)	$s = 5.324$ [5.319 to 5.328] S (-0.0%) $D = 5.93$ [5.85 to 6.02] F (+0.2%) $c_o = 0.199$ [.197 to .201] (-0.1%)
ovalbumin + 13.33% dimer; 60,000 rpm; 30 scans	$s = 3.55$ S $D = 7.89$ F $c_o = 0.8667$	$s = 3.508$ [3.506 to 3.511] S (-1.2%) $D = 7.49$ [7.43 to 7.55] F (-5.1%) $c_o = 0.8901$ [.887 to .893] (+2.7%)	$s = 3.547$ [3.545 to 3.549] S (-0.1%) $D = 7.92$ [7.86 to 7.97] F (+0.4%) $c_o = 0.8671$ [.865 to .870] (+0.1%)
	$s = 5.325$ S $D = 5.918$ F $c_o = 0.1333$	$s = 5.343$ [5.330 to 5.355] S (+0.3%) $D = 4.82$ [4.59 to 5.06] F (-18.6%) $c_o = 0.1248$ [.122 to .128] (-0.9%)	$s = 5.327$ [5.317 to 5.335] S (-0.0%) $D = 5.90$ [5.72 to 6.09] F (-0.3%) $c_o = 0.1322$ [.130 to .135] (-0.1%)
immunoglobulin + 20% dimer; 45,000 rpm; 16 scans	$s = 6.20$ S $D = 3.9$ F $c_o = 0.80$	$s = 6.164$ [6.160 to 6.167] S (-0.6%) $D = 3.62$ [3.59 to 3.67] F (-7.2%) $c_o = 0.810$ [.807 to .812] (+1.0%)	$s = 6.197$ [6.194 to 6.200] S (-0.1%) $D = 3.92$ [3.88 to 3.95] F (+0.5%) $c_o = 0.802$ [.799 to .804] (+0.3%)
	$s = 8.80$ S $D = 2.77$ F $c_o = 0.20$	$s = 8.799$ [8.793 to 8.805] S (-0.0%) $D = 2.38$ [2.33 to 2.44] F (-14.1%) $c_o = 0.198$ [.196 to .200] (-0.2%)	$s = 8.804$ [8.798 to 8.810] S (+0.1%) $D = 2.75$ [2.69 to 2.82] F (-0.7%) $c_o = 0.200$ [.198 to .202] (+0.0%)

^aThe data were evaluated at a time in the run just before the plateau region disappears, as the fastest species nears the cell base. Random noise of 0.001 r.m.s. amplitude was added to the simulated data. Values in square brackets are 95% confidence intervals; their absolute range reflects only the noise levels assumed, but their relative size is a measure of the resolution of the fitting methods. Values in parentheses indicate the percentage error (percentage of the total concentration for the concentration terms).

the relative precision of the parameters obtained, but produces an uncertainty small compared to the differences between the two methods.

The results of the fits to these simulated data are presented in Table 2. The simulation of ovalbumin + 20% dimer (Fig. 2) demonstrates that the ‘tail’ of the monomer peak in the dc/dt data does not interfere with resolving the dimer. Indeed, fitting to dc/dt gives results more than an order of magnitude more accurate than fitting to $g(s^*)$, and even the properties of the minor component can be resolved with better than 1% accuracy by fitting to dc/dt . It could be argued, however, that the superior results from fitting to dc/dt are a trivial consequence of the fact that because the dimer s value is 50% greater than the monomer value, the relative amplitude of the dimer signal is 50% stronger relative to monomer in the dc/dt data than in the $g(s^*)$ data, as can also be seen in Fig. 2. (Note, however, that the error bar on dc/dt at the dimer s value

is also 50% greater so the signal/noise ratio for the dimer is identical in both sets of data.)

Therefore a second simulation was done with the dimer fraction reduced to 13.33%, which makes the relative amplitude of the dimer signal in the dc/dt data correspond to the dimer amplitude in $g(s^*)$ for the 20% dimer simulation. Despite the lower amplitude, the accuracy of the dimer hydrodynamic parameters from fitting dc/dt is excellent, and still more than an order of magnitude better than the results from fitting $g(s^*)$ for the 20% dimer sample. Although the accuracy of the results from fitting dc/dt is much better, the precision of the two methods (the size of the error bars on fitted parameters) is quite similar (as expected since the signal/noise ratio of the dc/dt and $g(s^*)$ data is the same).

As a further test of the ability to resolve minor components when the physical separation is poor, a simulation was done for a mixture of immunoglobulin with 20% dimer. Studies in our laboratory have shown

that immunoglobulins often form dimers with quite low sedimentation coefficients for their size (presumably because they have highly extended structures in solution), and consequently the sedimentation coefficient of such dimers is only ~40% greater than that of monomer. Again by fitting to dc/dt the results for both components are accurate to better than 1%, which is more than an order of magnitude better than is possible by fitting $g(s^*)$. It is also worth noting that for both fitting methods the presence of the dimer component has not significantly degraded the accuracy of values for the monomer, as can be seen by comparing these results to the corresponding noise-free simulation for monomer in Table 1.

Therefore it appears that this new methodology works quite well for multi-species analysis. It should not, however, be concluded that potential problems arising from the “tails” on the dc/dt curves are non-existent, and that fitting to dc/dt will be superior in all situations. A consequence of the fact that the dc/dt value at any particular s^* value, s_1 , contains contributions from all species with $s^* < s_1$ is that a poor fit for any low s component will propagate to all species with higher s values. Thus if we are trying to study some faster sedimenting minor components, but the fit to the major, slower component is poor (perhaps due to unresolved heterogeneity), then the results for

the minor components could be significantly compromised by uncertainty about the correct level of the “tail” contribution from the major component. In such a situation the $g(s^*)$ data may give superior results, because the contributions of the “tails” will have been removed in a way that is basically model-independent rather than the model-dependent way inherent in fitting dc/dt to individual components.

An alternative procedure for calculating $g(s^)$ and the average dc/dt*

When studying samples containing both fast- and slowly-sedimenting species it is often the case in practice that in order to obtain good signal/noise for the slow species the time span of the scans will cause broadening of $g(s^*)$ peaks for the faster, higher mass species. (Exactly how the time span affects this broadening will be addressed further below). In principle this boundary broadening does not pose problems in these cases, since for such heterogeneous systems there usually is no attempt to derive properties for individual species, and it is only the changes in species distributions and/or the overall weight-average sedimentation coefficient that are of interest. Surprisingly, however, in such situations the error bars on the $g(s^*)$ distribution for the faster species are quite large, and may actually increase as more scans are

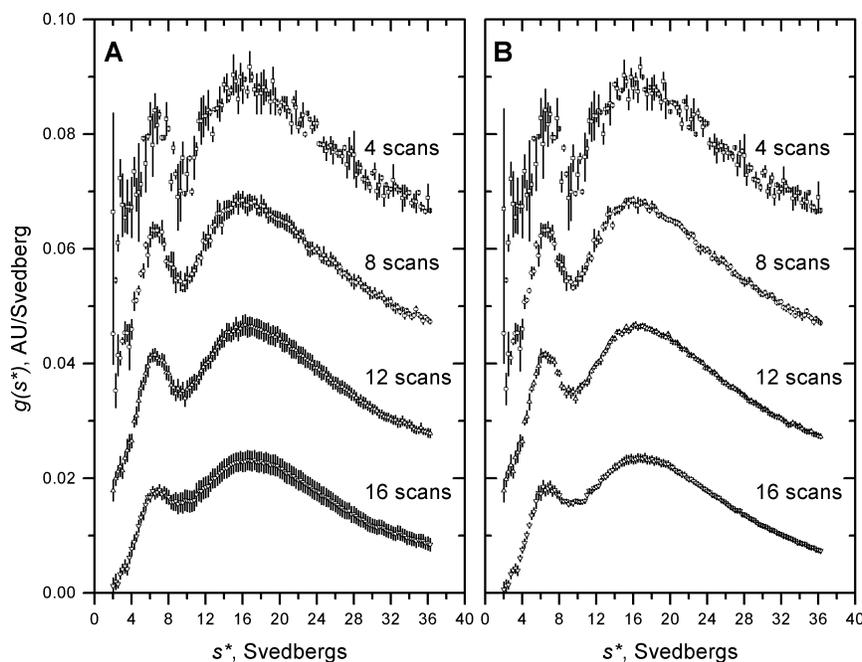


Fig. 3. Variation in the noise level of $g(s^*)$ distributions with the number of scans used in the analysis for a highly polydisperse sample. For the sake of clarity the curves for 4, 8, and 12 scans have been successively shifted upward by 0.02 AU/Svedberg, and only every fifth data point from the distributions is shown. Panel A shows the results using the original Stafford algorithm, while panel B uses the new algorithm described in the text.

used.

This phenomenon is illustrated in Fig. 3, some data for a monoclonal antibody sample that became highly aggregated after exposure to high temperatures. In theory the average error bars should be successively reduced by factors of 2.83, 1.84, and 1.54 in going from 4 to 8, 8 to 12, and 12 to 16 scans, respectively. Clearly the actual reduction is far less than this in going from 8 to 12 scans, particularly over the region from ~ 14 to 24 S. In going from 12 to 16 scans the error bars actually slightly increase (although the data become smoother). This same phenomenon can also be seen in the standard errors of the weight-, z-, and z+1-average sedimentation coefficients, s_w , s_z , and s_{z+1} , computed from these distributions (Table 3), which increase rather than decrease as the number of scans grows beyond 8.

Why doesn't the uncertainty in the $g(s^*)$ distribution decrease in the predicted manner? The standard error for each point in the average dc/dt curve is computed as the standard error of the mean of the dc/dt values from each pair of scans, and then the SE of the corresponding point in the $g(s^*)$ distribution is calculated by the usual propagation of error procedures. Thus the underlying problem must be that deviations among the individual dc/dt curves are increasing as the time span increases. In fact this is true, and this can most easily be seen by again considering noise-free synthetic data. Fig. 4A illustrates the systematic changes in amplitude and shape of the individual dc/dt curves for a time span that would produce moderate broadening of the corresponding peak in $g(s^*)$. Also shown is the average

curve along with its substantial error bars arising entirely from the systematic variation rather than intrinsic noise.

These systematic changes in dc/dt are the result of two different phenomena: (1) an overall loss of amplitude with time due to dilution of the sample as sedimentation progresses, and (2) a narrowing of the curves with time as the relative influence of diffusion decreases. Is there a way to reduce these systematic effects? Eliminating the changes due to diffusion would require knowing the diffusion coefficients *a priori* and cannot be done in a model-independent way. There is, however, a fairly straightforward way to minimize the consequences of the different radial dilutions for different scan pairs, since compensation for the run time is incorporated into the normalization process in transforming from dc/dt to $g(s^*)$. The standard approach (2) is to first average the dc/dt curves, then transform to $\hat{g}(s^*)$, and finally, if desired, to correct $\hat{g}(s^*)$ to $g(s^*)$. A possible alternative approach is to transform the dc/dt curve from each scan pair into $\hat{g}(s^*)$ and then average those data to produce an average $\hat{g}(s^*)$ distribution. Fundamentally the difference between these two approaches is that the new algorithm uses the mean time for each scan pair during the transformations, rather than using an overall mean time for the whole group (which becomes a worse approximation as the time span becomes larger).

Does this alternative algorithm actually help? Fig. 4B shows the $\hat{g}(s^*)$ curves from each scan pair as well as the average of those curves. With the alternative procedure the systematic differences consist of changes

TABLE 3

Variation in s_w , s_z and s_{z+1} with the number of scans used, for the data from Fig. 3

Number of scans	averages from the standard algorithm \pm SE (Svedbergs)	averages from the new algorithm \pm SE (Svedbergs)
4	$s_w = 18.365 \pm 0.138$ $s_z = 22.023 \pm 0.130$ $s_{z+1} = 24.694 \pm 0.168$	$s_w = 18.354 \pm 0.134$ $s_z = 22.020 \pm 0.119$ $s_{z+1} = 24.694 \pm 0.155$
8	$s_w = 18.377 \pm 0.067$ $s_z = 22.051 \pm 0.073$ $s_{z+1} = 24.752 \pm 0.089$	$s_w = 18.369 \pm 0.044$ $s_z = 22.035 \pm 0.040$ $s_{z+1} = 24.730 \pm 0.053$
12	$s_w = 18.468 \pm 0.071$ $s_z = 22.189 \pm 0.089$ $s_{z+1} = 24.911 \pm 0.109$	$s_w = 18.424 \pm 0.025$ $s_z = 22.123 \pm 0.022$ $s_{z+1} = 24.830 \pm 0.027$
16	$s_w = 18.705 \pm 0.095$ $s_z = 22.537 \pm 0.129$ $s_{z+1} = 25.303 \pm 0.167$	$s_w = 18.542 \pm 0.027$ $s_z = 22.314 \pm 0.026$ $s_{z+1} = 25.052 \pm 0.028$

in peak width, resulting in good congruence among the curves along the sides of the peak, but substantial variation at the peak itself and near the base. Although this procedure certainly still leaves substantial systematic variations, nonetheless the net result is an overall reduction in variance, and the standard error of s_w is reduced from ± 0.019 down to ± 0.010 by the new algorithm. Fig. 3B illustrates the significant improvement in the uncertainties of the $g(s^*)$ distribution obtained from a real experiment. The actual $g(s^*)$ distributions obtained from the two algorithms are very similar; the main effect of the new algorithm is to provide a better estimate of the error bars when the time span is long. The only real penalty from using the new algorithm is that it is more

computationally intensive. The new algorithm requires roughly $N/2$ times more computational time than the standard computation, but with current desktop computers this extra computation is not at all prohibitive.

This alternative algorithm can also provide an alternative means of calculating the average dc/dt curve and its error bars, for use in fitting to dc/dt . Once the average $\hat{g}(s^*)$ distribution has been calculated via the new algorithm, it can be used to calculate a corresponding dc/dt curve by applying the reverse of the normal transform from dc/dt to $\hat{g}(s^*)$, as detailed in *Methods*. As shown in Fig. 4A, the main effect is again to reduce the magnitude of the error bars, and for single species that effect is strongest along the steep portions of the dc/dt curves.

As would be expected, the reduction in error bars by the new algorithm leads to lower uncertainties in fitted parameters for fits to either the dc/dt or $g(s^*)$ curves. The largest improvement is in defining the sedimentation coefficients and concentrations, with little effect on the diffusion coefficients. While this improvement in precision of s and c_o is 3-4 fold for the noise-free data shown in Fig. 4, for real experiments the improvement would probably be fairly small.

Since weight average (and possibly z - or $z+1$ -average) sedimentation coefficients can be very useful in characterizing both ligand-induced assembly and self-associating systems (9-11), the quantitative effects of the new algorithm on these quantities for the real sample shown in Fig. 3 are detailed in Table 3. A systematic shift of s_w , s_z , and s_{z+1} toward higher values as more scans are used can be seen in Table 3. To confirm this behavior and assess the maximum accuracy attainable for these average quantities, a simulation was done for a mixture containing 8 species ranging from 6.2 to 30 S. This simulation was intended to mimic optimal signal/noise conditions for absorbance scans, and scanning of only a single sample at the maximum scan rate. As shown in Fig. 5, this simulation confirms that systematic errors are an inherent property of the values derived by time derivative techniques, and as the time span increases the systematic errors can significantly exceed the apparent precision of the averages. Over the range of scan numbers shown in Fig. 5 the systematic shifts in s_w are quite small compared to those for s_z and s_{z+1} , but over still longer time spans s_w also shifts to higher values in a systematic fashion. Under these optimal conditions the new algorithm provides no real advantages, but simulations confirm that when signal/noise is poor and long time spans are needed,

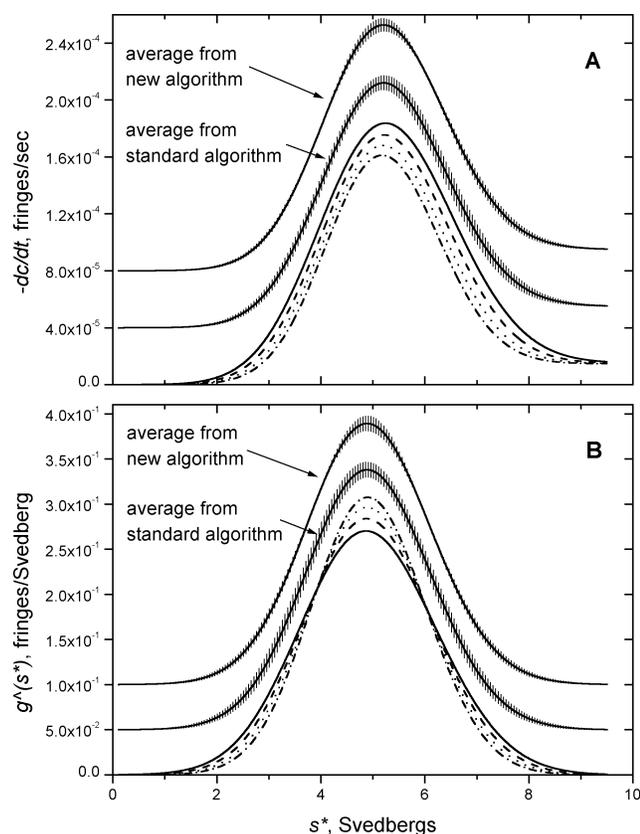


Fig. 4. (panel A) Systematic deviations in the dc/dt curves from individual scan pairs when the time span of the data becomes moderately excessive. These were derived from a simulation of a species with $s = 5$ S and $D = 6$ F (~ 75 kDa) run at 40000 rpm using 8 scans collected every 12.5 min. Also shown are the average \pm SE, calculated by both the standard algorithm and the new algorithm presented here; both have been shifted upward for the sake of clarity. (panel B) curves for the same four scan pairs. Also shown are the average \pm SE from both the standard and new algorithm, which have again been shifted upward for the sake of clarity.

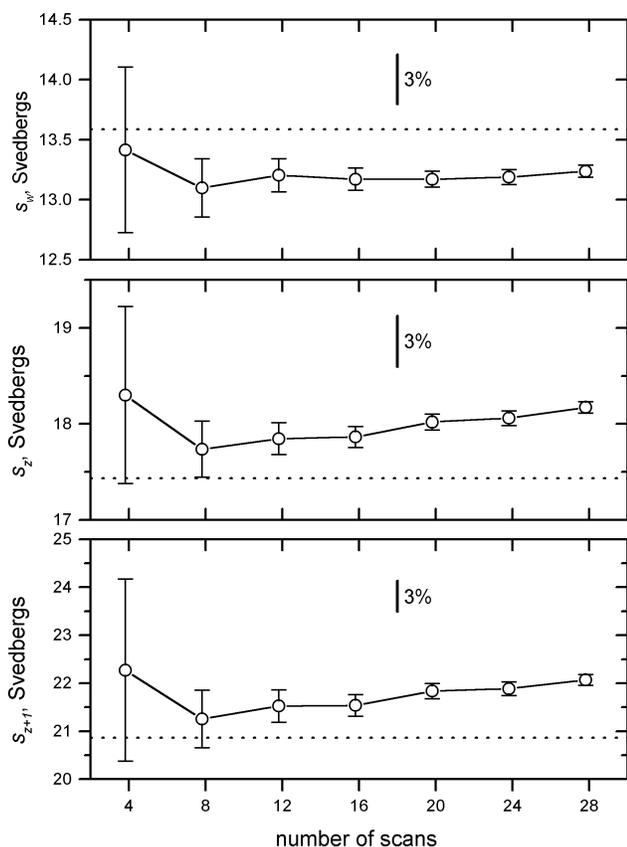


Fig. 5. Average sedimentation coefficients versus the number of scans used in the analysis, obtained from a simulation of a mixture of 8 species ranging from 6.2 to 30 S. The true values for the weight-, z-, and z+1-average sedimentation coefficients are indicated by the dotted horizontal lines in each panel. The vertical bars indicate the size of a 3% deviation from the true value. The data points are from $g(s^*)$ distributions calculated by the standard algorithm; over this time span the results from the new algorithm are equivalent. The simulation had a total loading concentration of 0.56 AU, scans every 90 s, and 0.006 r.m.s. of added random noise. To increase the total scans both earlier and later scans were added symmetrically to keep the midpoint of the group unchanged.

this algorithm does give lower error bars and a small reduction in the systematic errors for the various averages.

Effects of using too many scans (peak broadening)

As discussed in the introduction, in practice the use of time-derivative analysis generally requires a trade-off between using more scans to improve signal/noise and a loss of accuracy due to peak broadening. Now that the intrinsic accuracy of these methods has been quantified, it is worth examining at what point this

accuracy is compromised, whether this problem affects the results from fitting to dc/dt or $g(s^*)$ equivalently, and whether the accuracy differs when the distributions are computed using the conventional or alternative algorithms. Walter Stafford has proposed a “rule of thumb” (<http://www.bbri.org/dcdt/Rule.pdf>) to calculate the maximum time span between the first and last scan used in the analysis, Δt_{\max} , before substantial broadening occurs:

$$\Delta t_{\max} = \frac{160 \cdot t}{\sqrt{M} \cdot (\text{RPM}/1000)} \quad (5)$$

where t is the mean time between first and last scans in seconds, M is the mass in kDa, and RPM is the rotor speed. This formula assumes that the boundary is near the midpoint of the cell, a partial specific volume, \bar{v} , of 0.725 ml/g, and a solvent density, ρ , of 1 g/ml.

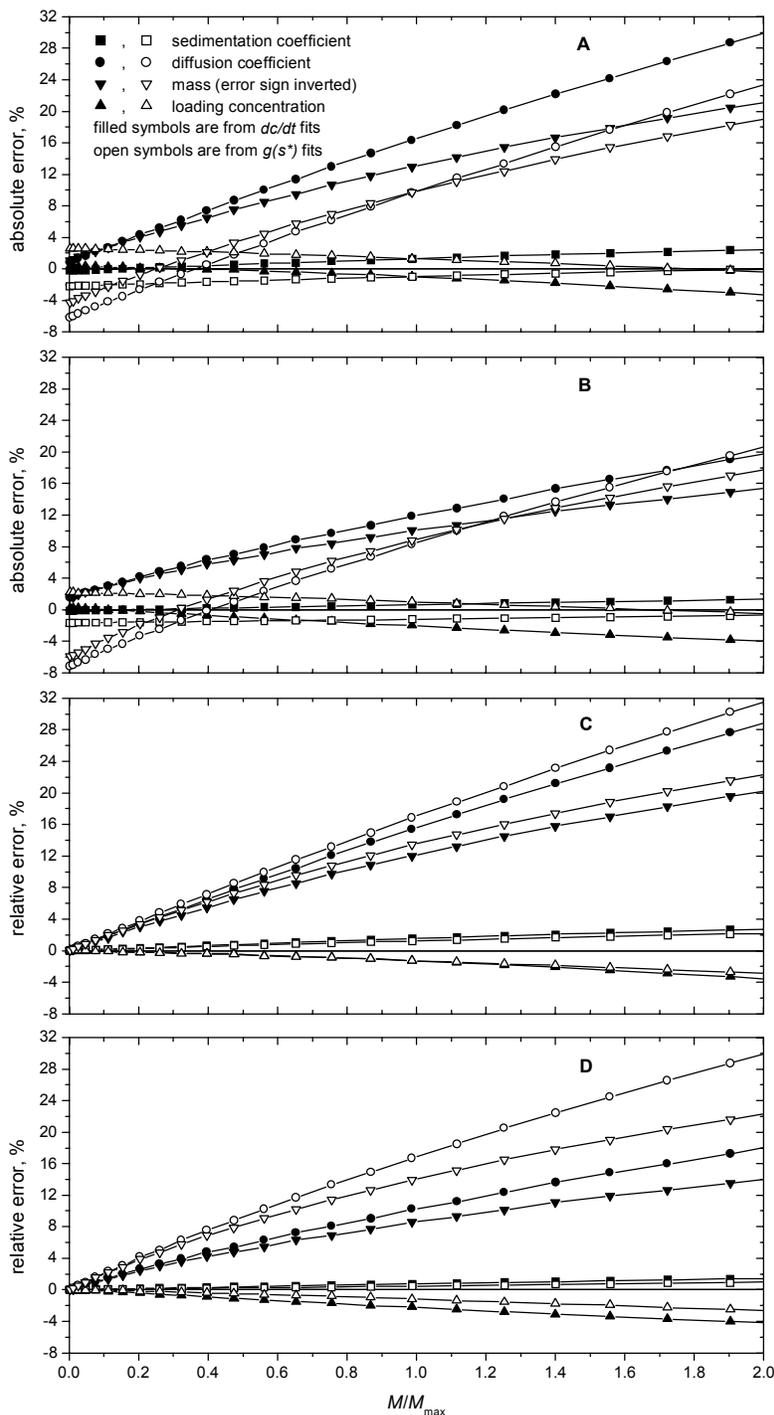
Since this maximum time span is directly related to sample mass, for a given set of scans this rule can also be used to calculate the maximum mass that can be present in the sample without significant broadening, M_{\max} :

$$M_{\max} = \left(\frac{160 \cdot t}{\Delta t \cdot (\text{RPM}/1000)} \right)^2 \frac{0.375}{(1 - \bar{v} \cdot \rho)} \quad (6)$$

The question then becomes a quantitative one of how large are the errors as the sample mass approaches M_{\max} ? To investigate this question noise-free simulations were done to generate a large group of scans separated by constant time intervals, covering the period before and after the boundary reaches the midpoint of the cell. The number of scans used in the analysis was then increased upward from the minimum of four to systematically decrease M_{\max} ; the additional scans were added to both the beginning and end of the group to keep the mean time constant.

Fig. 6 summarizes the results of such an analysis of the errors in s , D , M , and c_o from simulations for a ~ 75 kDa protein with $s = 5$ S and $D = 6$ F at 40000 rpm, and from fitting to either $g(s^*)$ or dc/dt . Panels A and B show the absolute errors in percent (relative to the true values that went into the simulation), using the Stafford algorithm to calculate dc/dt and $g(s^*)$ in panel A and the new algorithm in panel B. Panels C and D shows the errors relative to the values obtained when using a very narrow time span, which separates the effects of the time span from the issue of the absolute accuracy of these methods. These results show that the percentage error in s , D , M and c_o grows approximately linearly with the ratio M/M_{\max} up to a ratio of ~ 1 (the ratio which corresponds to just satisfying the “rule of thumb”). This approximate linearity means that for uniformly spaced scans the errors from broadening

Fig. 6. Variation in parameter errors with the time span (number of scans) used in the analysis (plotted versus the ratio of true mass to M_{\max} as calculated using Eq. [6]). These results are from noise-free simulations of an $s = 5$ S, $D = 6$ F species (~ 75 kDa) at 40000 rpm with scans recorded every 75 seconds. Analyses were done starting with 4 scans and adding scans in groups of 4 (2 earlier, 2 later), up to a total of 92 scans. Panel A shows the absolute percentage errors (relative to the true parameters used in the simulation) from fitting to either dc/dt (open symbols) or $g(s^*)$ (closed symbols) with data derived from the standard algorithm; panel B shows the absolute errors for the same scans when using the new algorithm. Panels C and D show the errors relative to the values obtained when using only 4 scans for the standard and alternative algorithms, respectively. To reduce the vertical size of the graphs the sign on the errors in mass (s/D ratio) was inverted.



grow proportionally to the square of the number of scans. As M/M_{\max} approaches or exceeds 1 the errors grow somewhat less rapidly, particularly for the error

in M . The effects of a large time span are fairly small on s and c_o , producing errors of only a few percent when M_{\max} is comparable to M . In contrast, the errors

in D (and consequently M) are much larger, with the apparent D increasing by 14-17% and M decreasing by 12-15% at $M/M_{\max} = 1$ when using the Stafford algorithm. The new algorithm gives only a very small improvement in the errors from fitting to $g(s^*)$, but reduces the errors in D from fitting to dc/dt by about 1/3.

Panels A and B also illustrate that although the absolute accuracy of the dc/dt fits is much better for small time spans, because fitting to $g(s^*)$ tends to underestimate the true D there is a compensation effect at moderately large time spans and the absolute errors in D and M from $g(s^*)$ fits are consequently quite small for $M/M_{\max} \sim 0.4$.

The results in Fig. 6 strictly speaking only apply to this specific test case, and certainly the magnitudes of the absolute errors will vary substantially from case to case as was shown in Table 1. However, the relative errors in Fig. 6 should have similar magnitudes for other experiments, and the overall patterns of absolute and relative errors should also be similar.

DISCUSSION

Tests on both real and simulated data indicate that the new methods described here can provide some significant improvements, at least for certain situations and experiments. These approaches are best viewed as alternatives or supplements to, rather than replacements for, the methods pioneered by Walter Stafford. While fitting to dc/dt rather than $g(s^*)$ can theoretically improve accuracy by an order of magnitude, the analysis in Fig. 6 clearly shows that this improved accuracy requires strongly limiting the time span of the data. In practice this may mean that except for slowly sedimenting solutes ($s < 2.5$) such improved accuracy can only be obtained using interference scans. Another implication of Fig. 6 is that some caution is needed in applying time derivative methods to answer critical questions such as whether a pharmaceutical material is heterogeneous.

Probably the most important lesson to be learned from Fig. 6 is that if absolute accuracy of D or M values from time derivative analysis is important, the best approach is to do a simulation that matches the actual experimental conditions in order to quantify the systematic errors and thereby obtain an appropriate correction factor. Since the whole boundary methods of analysis are not compromised by large time spans (and indeed are actually improved by them), another viable approach is to first use the time derivative methods to quickly provide an overview of how many species are present and to obtain good estimates of

their properties, and then switch to whole boundary methods for the rigorous quantitative analysis.

This issue of quantitative accuracy when fitting to individual species should not be allowed to obscure or overshadow the importance of time derivative methods for providing the overall distribution functions and average sedimentation coefficients in a model-independent way. In studying complex interacting systems where individual species cannot be resolved, the average properties of the distribution may be the only available criterion for testing and evaluating assembly models, and changes in the overall shape of the distributions with concentration or added ligands can provide important insights into mechanisms. The alternative algorithm discussed here will hopefully provide some modest improvement in these applications. Recently there has been renewed interest in using z - or $z+1$ -average sedimentation coefficients to characterize differences among related samples (11) and to help improve fitting of binding and assembly models. (9) The simulations done here (Fig. 5) suggest that an accuracy and precision of about 3% is achievable for s_w , s_z , and s_{z+1} , but that when the primary purpose of the experiment is to define these average sedimentation coefficients the use of a large number of scans in an attempt to improve signal/noise is likely to be self-defeating. It should also be noted that for data from the interference optical system there is often some error or ambiguity in defining the zero level for the $g(s^*)$ distribution, and any overall shift upward or downward of the distribution will have a particularly large impact on the values for s_z and s_{z+1} .

REFERENCES

- Stafford, W. F., III (1992) *Anal. Biochem.* **203**, 295-301.
- Stafford, W. F., III (1994) *Methods Enzymol.* **240**, 478-501.
- Stafford, W. F., III (1996) *Biophys. J.* **70**, MP452-MP452.
- Stafford, W. F., III (1997) *Curr. Opin. Biotechnol.* **8**, 14-24.
- Stafford, W. F., III (2000) *Methods Enzymol.* in press.
- Philo, J. S. (1994) in *Modern analytical ultracentrifugation* (Schuster, T. M. and Laue, T. M., Eds.), pp. 156-170, Birkhauser, Boston.
- Philo, J. S. (1997) *Biophys. J.* **72**, 435-444.
- Schuck, P. (1998) *Biophys. J.* **75**, 1503-1512.
- Correia, J. J. (2000) *Methods Enzymol.* in press.
- Lobert, S., Boyd, C. A., and Correia, J. J. (1997) *Biophys. J.* **72**, 416-427.

11. Toedt, J. M., Braswell, E. H., Schuster, T. M., Yphantis, D. A., Taraporewala, Z. F., and Culver, J. N. (1999) *Protein Sci.* **8**, 261-270.