Biophysical Chemistry 11 (1980) 303–308
© North-Holland Publishing Company

SIMULTANEOUS RAPID ESTIMATION OF SEDIMENTATION COEFFICIENT AND MOLECULAR WEIGHT

Leslie A. HOLLADAY
Department of Biochemistry, Vanderbilt University, Nashville, Tennessee 37232, USA

Received 29 October 1979

Data obtained from the early portion of sedimentation velocity experiments may be analyzed to simultaneously estimate both s and s/D. The C versus r data obtained are analyzed using a nonlinear least squares algorithm and an approximate solution to the Lamm equation. This procedure was tested both with simulated noisy data and with experimental data obtained using ribonuclease, ovalbumin, and somatostatin•dodecylsulfate. The procedure assumes that both s and D are independent of concentration. The results suggest that optimal estimation of both s and s/D is obtained at values of $\epsilon(= 2D/(s\cdot\omega^2\cdot r_a^2))$ in the range of 0.002 to 0.01 and values of $\tau(= 2\, s\omega^2 t)$ less than 0.04. Appropriate selection of rotor speed allows the estimation of both s and s/D for nearly globular macromolecules in the range of 10^4 to 10^6 daltons with data obtained during the first 3000–5000 seconds of a sedimentation velocity experiment.

1. Introduction

Some fifty years after the development of analytical ultracentrifugation by Svedberg and his collaborators, there remain only three types of experiments which are in wide use. The most notable aspect of current ultracentrifugal practice is that it does not use the great majority of the data which may be generated during the time course of the ultracentrifugal experiment. The conventional low speed or meniscus depletion experiments using 3 mm columns continue the experiment until equilibrium is closely approached, and the data are analyzed using the slope of the $\ln C$ versus r^2 plot to yield estimates of molecular weight. The data generated during approach-to-sedimentation equilibrium have been analyzed using extrapolations to the ends of the solution column [1], although as Fujita has pointed out, there is no generally agreed upon procedure for the required extrapolations [2]. Sedimentation velocity data are usually analyzed using the midpoint of the boundary curve, r^*, and plotting $\ln r^*$ versus time to estimate the sedimentation coefficient.

As a means of macromolecular characterization, each of these methods, by itself, is not entirely satisfactory. The sedimentation equilibrium method is not useful for macromolecules which are unstable and requires long experimental times for very large macromolecules. In addition only the parameter s/D is available from the data obtained at equilibrium. The sedimentation velocity procedure currently used yields only a value for s, although procedures are available to estimate D separately [2]. The Archibald procedure requires extrapolation of the experimental data to yield $(dc/dr)_a$ and yields only a value for s/D. The diffusion coefficient may be estimated from the same data only if the quantity ΔC_{eq} is available [3].

The use of Fujita's approximate Archibald-type solution to the Lamm equation [4] has been used to provide reliable estimates of s/D for data obtained during the early portion of approach-to-sedimentation equilibrium [5]. The estimates for the sedimentation coefficient obtained with this method were not reliable, differing as much as 30% from their correct value. It seemed possible that an experiment using higher rotor speeds might provide data which could be used to reliably estimate both s and s/D simultaneously. Unfortunately, the Fujita solution is in the form of an infinite series which converges at very slow rates for values of $\epsilon\cdot\tau$ smaller than 0.01, where $\epsilon = 2D/$

$(s\omega^2 r_a^2)$ and $\tau = 2s\omega^2 t$. A typical sedimentation velocity run has $\epsilon \cdot \tau = 0.0003$ at the midpoint of the experiment. An approximate solution to the Lamm equation based on the semi-infinite cell approximation and Fujita's approximation that $(r/r_a)^2 = ((r_b/r_a)^2 + 1)/2$ (r_a is the radial distance to the solution meniscus) has been derived [6]. The approximate solution used accounts for the effect of radial dilution and assumes concentration independence of s and D. This study deals with the use of this approximate solution with a nonlinear least squares algorithm using sedimentation velocity data to rapidly estimate both s and s/D.

2. Materials and methods

2.1. Nonlinear least squares algorithm

The algorithm estimates three parameters: s, D/s, and C_0. The method is based on Marquardt's algorithm [7] as implemented in the ZXSSQ routine of the International Mathematical and Statistical Library [8]. The ZXSSQ routine uses a finite difference method for estimation of the required partial derivatives [9]. The algorithm seeks to minimize the residual sum of squares ($= \Sigma(C_{exp} - C_{estimated})^2$) by iteratively improving the starting estimates for s, D/s, and C_0. The value of the Marquardt parameter λ is adjusted each iteration such that a reduction in the residual sum of squares was obtained. It was found that in some instances when the iteration process failed due to λ exceeding the preset limit (500,000) that convergence could be obtained when a different scaling factor was used. The initial value of λ was 0.01 and the scaling factor was in the range of from 2 to 5. Convergence was considered attained when the residual sum of squares for two successive iterations had a relative difference of less than 1×10^{-4} or when all three parameter estimates were constant to four digits. In the analysis of simulated data the initial guesses for s, D/s, and C_0 were 6.5×10^{-13} s, 1.01×10^6 cm^2/s^2, and 0.95. For the ovalbumin data the initial guesses for s and D/s were 4.0×10^{-13} s and 2.0×10^6 cm^2/s^2. The initial guess for C_0 was computed from the plateau concentration at about 200–500 s. For ribonuclease the initial guesses for s and D/s were 2×10^{-13} s and 0.5×10^6 cm^2/s^2. For somatostatin·dodecyl sulfate the initial guesses for s and D/s were 1×10^{-13} s and 0.7×10^6 cm^2/s^2. The algorithm required from 5 to 15 iterations to reach convergence with these initial estimates. Standard errors of the parameter estimates were computed using the assumption [10] that the parameter variance-covariance matrix $V_\theta = 2\sigma^2 N^{-1}$. The quantity $OD_{res} = (\Sigma_{i=1}^{n}(OD_{exp} - OD_{calc})^2)/(n-3))^{1/2}$, where n is the number of data points.

2.2. Proteins and reagents

Three times recrystallized bovine ribonuclease (code R-OG-B) and two times recrystallized ovalbumin (code 38S928) were purchased from Worthington Biochemical Corporation. Sematostatin was purchased from Pierce Chemical Company. Sodium dodecyl sulfate was purchased from Schwarz/Mann (Lauryl sodium sulfate grade). All salts were reagent grade.

2.3. Ultracentrifugation

Data were obtained with a Model E analytical ultracentrifuge equipped with electronic speed control and photoelectric scanner. The monochromator was set at 280 nm, the monochromator slit at 1.0–2.0 mm, and the photomultiplier slit at 0.1 mm. Scans were taken at instrument settings which gave a recording of about 322 mm length between the two reference edges. An AN-H rotor with 12 mm Kel-F centerpiece was used; 0.430 ml of solution and 0.440 ml of solvent were loaded in the cell. The solvent for somastostatin was 0.5 g/dl sodium dodecyl sulfate, 50 mM sodium phosphate, pH 7.0. The solvent for ribonuclease and ovalbumin was 0.15 M potassium chloride, 20 mM potassium phosphate pH 7.0. Times for the scans of C versus r were corrected for lower angular velocities during acceleration [3]. The rotor was allowed to reach 20°C for at least one hour prior to acceleration at a 12 amp current until the desired speed was reached. Ten pairs of OD_{280}, r were used for each time point.

2.4. Simulated noisy data

Synthetic data were generated using the approximate solution. Ten time points were used, equally spaced from $\tau = 0.004$ to $\tau = 0.04$. At each time point

the value of C was calculated at 10 equally spaced intervals between $r = 12\,e^{(\tau/2)} - 6$ and r_a. The data calculated covered the boundary portion of the C versus r curve. Normally distributed random numbers (mean zero) with standard deviation 0.010 or 0.020 were added to C. The initial concentration C_0 was taken to be 1.0.

2.5. Molecular weight calculations

The partial specific volume for ribonuclease was taken to be 0.695 cm^3/g [11]. The value for the partial specific volume of ovalbumin was taken to be 0.748 cm^3/g [12]. Densities of solvents were calculated from available data [13].

3. Results

The effect of varying rotor speed on the precision of the parameter estimates is shown in table 1. Simulated noisy data ($\sigma = 0.020$) were generated using $s = 5 \times 10^{-13}$ s and $D = 8.2 \times 10^{-7}$ cm^2/s. These data suggest that lower rotor speeds only marginally improve the precision of the estimate for D/s, but that the precision of the estimate for s becomes appreciably poorer at lower rotor speeds.

The usefulness of the method in the study of detergent-protein interactions is shown in table 2. Sedimentation equilibrium results [14] for the somatostatin–dodecyl sulfate complex yielded an estimate for $M(1 - \phi_2^* \rho)$ of 3874 (ϕ_2^* is the effective partial specific volume). From the data of table 2 an estimate for $M(1 - \phi_2^* \rho)$ of 3937 ± 140 is obtained, in excellent agreement with previous results. Fig. 1 displays experimental and least squares boundary curves.

Data obtained for ovalbumin are shown in table 3.

Table 2
Somatostatin·dodecyl sulfate: 56,000 rpm, 20°C: τ_{final} = 0.038, ϵ = 0.010

t (s)	OD_{res}	s (× 10^{13})	D/s (× 10^{-6})
510	0.004	1.23	4.96
1020	0.006	1.15	6.51
1535	0.006	1.02	6.49
2000	0.005	1.19	6.17
2500	0.004	1.11	6.39
3000	0.005	1.11	6.17
3500	0.003	1.10	6.10
4012	0.006	1.07	6.13
4500	0.003	1.13	6.31
5000	0.006	1.11	6.34
Mean result:	–	1.12 ± 0.06	6.16 ± 0.45
Composite result:	0.009	1.11 ± 0.03	6.19 ± 0.23

C_0 = 0.65 OD at 280 nm (1.2 cm path).

At 34,000 rpm the molecular weight is estimated to be 42,900 ± 1,400 and the estimate for $s_{20,\omega}$ is 3.21 Svedbergs. Data obtained for ribonuclease are shown in table 4. The molecular weight is estimated to be 12,900 ± 400. The estimate for $s_{20,\omega}$ is 1.77 Svedbergs. Values for s in the tables are not corrected to water values.

An important value which is required in the data analysis is the radial position of the solution meniscus. This value occurs throughout the approximate solution used in this analysis. Table 5 presents data which show that it is necessary to carefully measure this quantity if reliable results are to be obtained. The scanner records obtained in this study covered about 322 mm between the reference hole edges. It appears possible to routinely measure r_a to within about 0.2 mm on the chart. The position of the

Table 1
Effect of ϵ on precision of parameter estimation

RPM	ϵ	OD_{res}	s (× 10^{13})	σ_s (× 10^{13})	D/s (× 10^{-6})	$\sigma_{D/s}$ (× 10^{-6})
60,000	0.0023	0.023	5.024	0.076	1.581	0.079
50,000	0.0033	0.019	5.020	0.090	1.650	0.079
40,000	0.0052	0.021	4.926	0.155	1.661	0.082
30,000	0.0092	0.018	4.940	0.218	1.626	0.063
20,000	0.0207	0.018	4.933	0.416	1.651	0.069

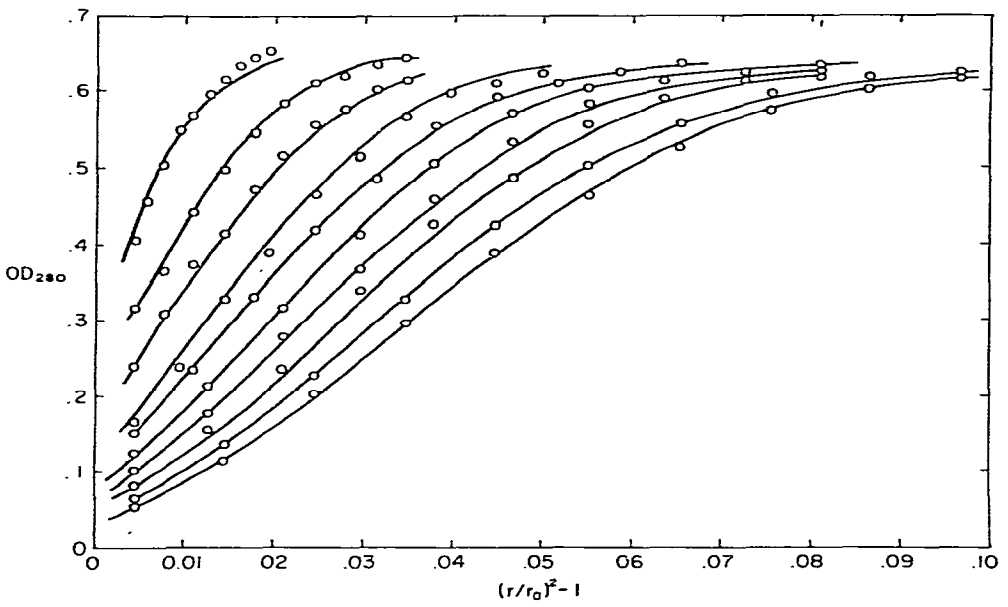

Fig. 1. Plot of experimental (open circles) and calculated (lines) boundary shapes for somatostatin dodecyl sulfate at 56,000 rpm and 20°C. The boundary shapes were calculated using the least squares estimates of $s = 1.11 \times 10^{-13}$ s and $D/s = 6.19 \times 10^{6}$ cm^2/s^2.

meniscus was taken as the radial distance to the top of the meniscus spike. This corresponds to a probable error in r_a of about 0.001 cm. A two and a half fold greater error (0.0025 cm) alters s and D/s by about 5 per cent.

The effect of a higher molecular weight contaminant on the method is shown in table 6. Simulated noisy data ($\sigma = 0.010$) were calculated for a mixture for which 90% of the material had $s = 5 \times 10^{-13}$ s, $D = 8.2 \times 10^{-7}$ cm^2/s, and 10% had $s = 7.94 \times 10^{-13}$ s with $D = 6.51 \times 10^{-6}$ cm^2/s. The resulting estimate for s is very close to the weight-average s value for the mixture. The resulting estimate for molecular weight is close to the molecular weight of the principal component.

Table 3
Ovalbumin: 34,000 rpm, 20°C: $r_{final} = 0.031$, $\epsilon = 0.011$

t (s)	OD_{res}	s ($\times 10^{13}$)	D/s ($\times 10^{-6}$)
1060	0.002	3.26	2.06
1540	0.003	3.07	2.34
2020	0.005	3.08	2.27
2500	0.002	3.16	2.29
2980	0.004	3.20	2.37
3460	0.003	3.14	2.38
3940	0.004	3.13	2.51
Mean result:	–	3.15 ± 0.07	2.32 ± 0.14
Composite result:	0.004	3.12 ± 0.11	2.31 ± 0.08

$C_0 = 0.37$ OD at 280 nm (1.2 cm path).

Table 4
Ribonuclease: 56,000 rpm, 20°C: $r_{final} = 0.025$, $\epsilon = 0.010$

t (s)	OD_{res}	s ($\times 10^{13}$)	D/s ($\times 10^{-6}$)
425	0.007	1.62	5.59
665	0.004	1.71	5.44
905	0.008	1.65	6.05
1145	0.006	1.68	6.18
1385	0.004	1.72	6.03
1625	0.008	1.76	6.45
1865	0.008	1.74	6.30
2105	0.012	1.70	7.02
Mean result:	–	1.70 ± 0.05	6.13 ± 0.49
Composite result:	0.010	1.73 ± 0.04	6.32 ± 0.22

$C_0 = 0.88$ OD at 280 nm (1.2 cm path).

Table 5
Effect of error in r_a on parameter estimates

r_a	OD_{res}	s ($\times 10^{13}$)	D/s ($\times 10^{-6}$)
5.8980 a)	0.0042	3.12	2.31
5.8881	0.0086	3.46	1.92
5.9005	0.0040	3.03	2.42
5.9079	0.0061	2.87	2.95

a) Correct value.

The effect of a lower molecular weight contaminant on the method is shown in table 7. Simulated noisy data ($\sigma = 0.010$) were calculated for a mixture for which 90% of the material had $s = 5 \times 10^{-13}$ s, $D = 8.2 \times 10^{-7}$ cm^2/s, and 10% had $s = 3.15 \times 10^{-13}$ s with $D = 10.3 \times 10^{-6}$ cm^2/s. The resulting composite estimate for s is close to the weight average s, while the resulting estimate for molecular weight is close to the weight average molecular weight.

4. Discussion

In the development of this method the assumption is made that the sedimentation and diffusion coefficients are independent of concentration. A recently developed theory [15] of the concentration dependence of transport processes relates K_s in the equation $s_c = s_0(1 - K_s c)$ to the intrinsic viscosity of a macromolecule $[\eta]$ and a newly derived shape factor R. For spherical macromolecules the sedimentation coefficient at a concentration of 0.5 mg/ml is predicted to be 0.998 s_0. For macromolecules which are prolate ellipsoids of axial ratio ten, the sedimentation coefficient at a concentration of 0.5 mg/ml is predicted to be 0.995 s_0. For macromolecules which are prolate ellipsoids of axial ratio 100, the sedimentation coefficient is predicted to be 0.94 s_0. These calculations suggest that for concentrations suitable for use with scanner optics that the concentration dependence of the sedimentation coefficient will be negligible except for markedly asymmetric macromolecules. The concentration dependence of D will in general be much smaller than that for s [15].

Table 6
Effect of 10% higher molecular weight contaminant on parameter estimates

t (s)	OD_{res}	s ($\times 10^{13}$)	D/s ($\times 10^{-6}$)
146	0.012	4.81	1.38
292	0.006	5.38	1.65
438	0.012	5.22	1.58
584	0.006	5.33	1.78
730	0.011	5.32	1.77
875	0.010	5.23	1.58
1021 a)	0.006	5.25	1.73
1167	0.008	5.21	1.69
1313	0.014	5.31	1.96
1459 a)	0.013	5.27	1.76
Mean result:	–	5.23 ± 0.16	1.69 ± 0.15
Composite result:	0.011	5.23 ± 0.06	1.65 ± 0.05

a) Convergence was not obtained for these two time points because the Marquardt parameter exceeded the set limit of 500,000 and iteration was terminated.

Table 7
Effect of 10% lower molecular weight contaminant on parameter estimates

t (s)	OD_{res}	s ($\times 10^{13}$)	D/s ($\times 10^{-6}$)
146	0.015	4.63	1.63
292	0.009	4.82	1.77
438	0.008	4.79	1.74
584	0.009	4.77	1.63
730	0.010	4.75	1.75
875	0.008	4.88	1.94
1021	0.010	4.74	1.73
1167	0.011	4.85	1.81
1313	0.010	4.83	1.88
1459	0.009	4.90	1.75
Mean result:	–	4.80 ± 0.08	1.76 ± 0.10
Composite result:	0.010	4.82 ± 0.05	1.76 ± 0.04

Fig. 2 displays suitable values of rotor speed as a function of molecular weight. For the upper curve ($t_{final} = 3000$ s) values of ϵ range from 0.0124 to 0.0031. For the lower curve ϵ ranges from 0.0164 to 0.0052. Since it is in reality the reduced dimensionless parameters τ and ϵ which determine the boundary position and shape, once optimal values of τ and ϵ are determined then it is only a matter of adjusting the rotor speed and time during which data are accumulated such that $\tau < 0.04$ and $0.002 \leqslant \epsilon \leqslant 0.02$. Fig. 2 shows that for this method the size of the molecule only marginally affects the time over which data should be accumulated. Thus, this method

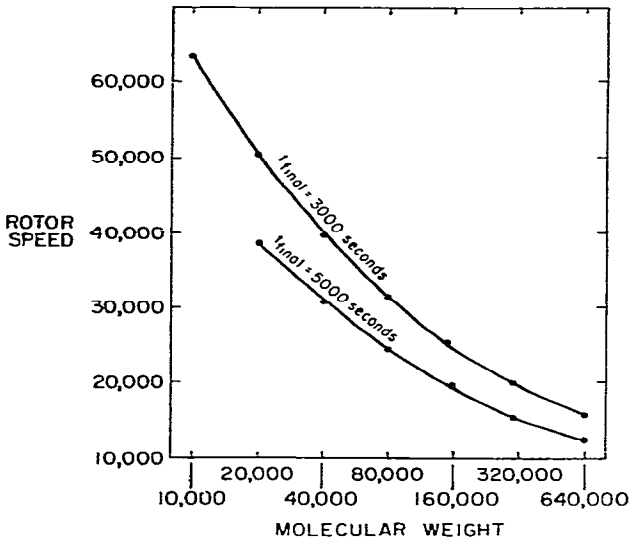

Fig. 2. Plot of suggested rotor speed versus molecular weight for two data collection times.

appears suitable for any nearly-globular macromolecule. Even for molecules with molecular weights in excess of 5×10^5, estimates for molecular weight may be obtained within a few hours.

The restriction that $\tau < 0.04$ arises from the fact that as τ becomes larger and larger the assumption that $(r/r_a)^2$ is constant becomes untenable. The close similarity between the two approximate solutions available for the conventional sedimentation velocity experiment [2,6] for $\tau < 0.03$ suggests that at early times the approximate solutions are sufficiently close to the exact solution.

In preliminary experiments it was found that if insufficient time was allowed for complete thermal equilibrium to be reached within the rotor and cell that convection occurred during the experiment, resulting in boundary shapes which could not be fit to the approximate solution with the expected degree of precision. In order for correct results to be obtained it appears necessary for the rotor to give a constant RTIC reading for at least one hour prior to acceleration. If the cell is carefully oriented within the rotor and if thermal equilibrium has been attained, the resulting data can be fitted to the approximate solution to the expected degree of accuracy (about 0.005 – 0.010 OD units). Excellent initial estimates for s and C_0 will in general be available. An initial estimate for s is obtainable by plotting $\ln r^*$ versus time, where r^* is the point at which $C = 0.5\,C_0$. An initial estimate for C_0 is obtainable by determining the plateau concentration at very early times. For spherical macromolecules with $(1 - \overline{V}\rho) \simeq 0.25$, an estimate for D/s may be obtained using the equation $\ln(D/s) = -28.33 - 1.505 \ln(s)$.

In summary, this study presents conditions and methods by which rapid simultaneous estimation of sedimentation coefficient and molecular weight may be obtained for nearly globular macromolecules.

References

[1] W.J. Archibald, J. Phys. Colloid Chem. 51 (1947) 1204.
[2] H. Fujita, Foundations of ultracentrifugal analysis (Wiley, New York, 1975).
[3] K.E. Van Holde and R.L. Baldwin, J. Phys. Chem. 62 (1958) 734.
[4] H. Fujita, Mathematical theory of sedimentation analysis (Academic Press, New York, 1962).
[5] L. Holladay, Biophys. Chem. 10 (1979) 184.
[6] L. Holladay, Biophys. Chem. 10 (1979) 187.
[7] D.W. Marquardt, J. Soc. Ind. Appl. Math. 11 (1963) 431.
[8] T.J. Aird and N.L. Darilek, International mathematics and statistics library (IMSL, Houston, 1975).
[9] K.M. Brown and J.E. Dennis Jr., Numerische Mathematik 18 (1972) 289.
[10] Y. Bard, Nonlinear parameter estimation (Academic Press, New York, 1974).
[11] W.F. Harrington and J.A. Schellman, Compt. rend. trav. Lab. Carlsberg Series Chim. 30 (1956) 21.
[12] M.O. Dayhoff, G.E. Perlman and D.A. McInnes, J. Am. Chem. Soc. 74 (1952) 2515.
[13] R.C. Weast, ed., Handbook of chemistry and physics (CRC Publishing Co., Cleveland, 1975).
[14] L.A. Holladay and P. Wilder, Biochim. Biophys. Acta (in press).
[15] A.J. Rowe, Biopolymers 16 (1977) 2595.