Analysis of Exponential Curves by a Method of Moments, with Special Attention to Sedimentation Equilibrium and Fluorescence Decay*

Robert D. Dyson† and Irvin Isenberg

ABSTRACT: The method of moments, which was earlier developed to resolve the simultaneous fluorescence of several species in the presence of an overlapping excitation (Isenberg, I., and Dyson, R. D. (1969), *Biophys. J. 9*, 1337), has now been extended to the problem of analyzing linear sums of several exponentials. For m independent components, we proceed by finding the values for the amplitudes, α_j, and time constants, τ_j, that minimize the difference between the $2m$ moments of the decay, given

$$\mu_k = \int_0^T y(t)t^k dt \qquad k = 0, 1, \ldots, 2m - 1$$

and the corresponding set of moments calculated from the assumption that the decay itself may be represented as

$$y(t) = \sum_{j=1}^{m} \alpha_j \exp(-t/\tau_j) \qquad 0 \leq t \leq T$$

In addition, a smoothing functional, which we refer to as the *mean displaced ratio* (MDR), is presented as a way of reducing the effects of random deviations without introducing systematic error. When the data consist of $N + 1$ observations, equally spaced in time such that $t_0 = 0$ and $t_{i+l} = t_i + t_l$, the MDR is obtained from

$$Y_l = \frac{1}{N - L + 1} \sum_{i=0}^{N-L} y_{i+l}/y_i \qquad l = 0, 1, \ldots, L$$

It is shown that this new curve, Y *vs.* t, has the same time constants as the original curve, amplitudes that are simply related to those of the original decay process, but considerably less noise. These procedures are tested by successfully resolving the sedimentation equilibrium pattern obtained from a known mixture of two proteins (bovine serum albumin and ovalbumin) and by resolving the combined fluorescence of two long-lived chromophores. Extensions to more general systems are also discussed.

The analysis of exponential curves is a recurring problem in both the physical and biological sciences. Examples include the impulse response of linear systems in network analysis (*cf.* McDonough and Huggins, 1968; Marzollo, 1969), transfer rates in biological systems, as determined by compartmental analysis (Robertson, 1957), the distribution of solute at sedimentation equilibrium in the ultracentrifuge (Svedberg and Pederson, 1940), phosphorescence and fluorescence decay (Jablonski, 1935; Wahl and Lami, 1967; Tao, 1969), the kinetics of isotope exchange (Hvidt and Neilsen, 1966), and a host of other relaxation and time-decay phenomena (*e.g.*, Friess *et al.*, 1963; Caldin, 1964; Amdur and Hammes, 1966). Each of these situations may often be represented with simple sums of exponentials. However, the techniques used for analyzing such systems—that is, for determining the amplitude and time constant for each of the exponential terms—are notoriously unreliable.

Most attempts at resolving exponential curves have relied on local linearization with iterative improvements (as in the Newton–Raphson technique), or on the application of Prony's linear operator. Both are extremely sensitive to even small errors in the experimental data (Hildebrand, 1956; Lanczos, 1956), even when the least-squares criterion can be used on an overdetermined system. A graphical or analytical peeling of the curve to remove the contribution of each component one at a time has also been widely used (*e.g.*, Smith and Morales, 1944; Perl, 1960; Bennett *et al.*, 1965; Van Liew, 1967) but suffers from a cumulative buildup of error.

The pitfalls of relying on these procedures in various biological applications have been discussed previously. In particular, Worsley and Lax (1962) and Worsley (1964) have examined the iterative least-squares techniques in general terms, while Glass and de Garreta (1967) have attempted to define their limits when applied specifically to transfer rates in clinical studies where data are sparse. In addition, the problems and prospects of least-squares fitting to hydrogen-exchange data were pointed out by Laiken and Printz (1970). Myhill (1965) has discussed the applicability of Prony's linear operator to the analysis of biological tracer data, while Haschemeyer and Bowers (1970) discussed the value of this technique and that of the least-squares procedures in the analysis of sedimentation equilibrium patterns derived from paucidisperse systems.

Another, but less widely employed technique for the analysis of exponential curves, has been presented by Gardner and coworkers (1959). They state the problem in terms of the Laplace integral equation, and then invert it with Fourier transforms. Like the other procedures just mentioned, their approach seems to be satisfactory when many half-lives of decay can be observed, but is seriously limited when more restricted observational ranges must be used.

In an earlier paper (Isenberg and Dyson, 1969) we extended

* From the Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331. *Received March 22, 1971.* This work was supported in part by the National Institutes of Health, U. S. Public Health Service, through Research Grants GM-15715 and CA-10872. A preliminary report of a portion of this work was presented to the joint conference of the American Chemical Society and the Chemical Institute of Canada, Toronto, Ont., May 25, 1970.

† To whom to address correspondence.

Bay's work on the method of moments (Bay, 1950), to develop a formalism for obtaining the decay parameters from fluorescence data in which the exciting lamp flash and observed fluorescence overlap. Our procedures have recently been used (Schuyler and Isenberg, 1971) for the analysis of a variety of data. In the present paper we present a method for the analysis of data that can be represented as a simple sum of exponentials. In addition, we introduce a smoothing functional that can be used to improve the signal-to-noise ratio of exponential data without changing the form of the curve or introducing systematic errors. This improvement permits one to analyze data covering a more restricted range of values. Following a discussion of the method, we present the analysis of two sets of such data, one obtained from the sedimentation equilibrium of a mixture of two proteins, and the other taken from the combined fluorescence of two long-lived chromophores.

Theory

The problem, in general terms, is to find the number of exponential components, m, and the amplitude and time constant (α_j and τ_j, respectively) of each component, when the total amplitude can be represented as

$$y(t) = \sum_{j=1}^{m} \alpha_j \exp(-t/\tau_j) \qquad 0 \leq t \leq T \tag{1a}$$

We are particularly concerned with the case where one is given $N + 1$ discrete observations (y_i, t_i), $i = 0, 1, \ldots, N$.

For the specific case of ultracentrifugation, the problem is one of finding the buoyant molecular weight, $M_j(1 - \bar{v}_j\rho)$, of each component at sedimentation equilibrium, along with its concentration, $C_j(0)$, at an arbitrary radial position r_0. These parameters must be obtained from the total concentration, C, measured as a function of radial position. For ideal solutions, the concentration distribution may be described as a simple sum of exponentials

$$C(z) = \sum_{j=1}^{m} C_j(0) \exp(-AM_j^* z) \qquad 0 \leq z \leq Z \tag{1b}$$

where $M_j^* = M_j(1 - \bar{v}_j\rho)$, $A = \omega^2/2RT$, and $z = r_0^2 - r^2$ ($r \leq r_0$). Here we use M_j to represent molecular weight, $\bar{v}_j$ is a partial specific volume, ρ is the solution density, ω is the angular velocity of the rotor, and R and T are the gas constant and absolute temperature, respectively. It is obvious that solving eq 1a and 1b presents the same problem. Therefore, to simplify the subsequent discussions, we will adopt the nomenclature of eq 1a.

Data Smoothing through the Mean Displaced Ratio (MDR). If an exponential curve could be obtained without noise, any of the computational methods mentioned earlier, with the possible exception of curve peeling, might be adequate to provide the individual time constants and amplitudes. The analysis of real data, however, is decisively influenced by noise and the limitations it imposes. Thus one may seek to minimize random errors by repetitive measurements or damping. Where that is not feasible or adequate, one may turn to the use of numerical filters.

The simplest numerical filtering procedure consists of fitting an overdetermined polynomial of modest degree to consecutive or overlapping sets of points, and replacing one or more of the points at each fit with the value of the polynomial at the same coordinate position. But if the analytical form of the smoothing function cannot satisfactorily represent the actual data, serious systematic errors may result. The use of polynomial filters is especially hazardous with exponential data, since such data often cannot be adequately represented by any small polynomial, except over extremely narrow spans.

In the present case, we make the assumption that the observed amplitude is a linear sum of exponentials, although the number of components is unknown. This allows us to replace the original curve, $y(t)$, with a new curve, $Y(q)$, that has a much improved signal-to-noise ratio, but has the same time constants as the original curve. For continuous data, we define $Y(q)$ by

$$Y(q) = \frac{1}{T - Q}\int_0^{T-Q} [y(t + q)/y(t)]\mathrm{d}t \qquad 0 \leq q \leq Q \text{ and } Q < T \tag{2}$$

Note that where the original observations are collected between $t = 0$ and $t = T$, this new parameter, $Y(q)$, covers a somewhat smaller range, from $q = 0$ to $q = Q$. From eq 1a

$$Y(q) = \sum_{j=1}^{m}\left[\frac{\alpha_j}{T - Q}\int_0^{T-Q} \exp(-t/\tau_j)/y(t)\mathrm{d}t\right] \exp(-q/\tau_j)$$

or

$$Y(q) = \sum_{j=1}^{m} \beta_j \exp(-q/\tau_j) \qquad 0 \leq q \leq Q \tag{3}$$

where

$$\beta_j = \frac{\alpha_j}{T - Q}\int_0^{T-Q} \exp(-t/\tau_j)/y(t)\mathrm{d}t \tag{4}$$

and

$$Y(0) = \sum_{j=1}^{m} \beta_j \equiv 1 \tag{5}$$

Thus, the transformation merely involves finding the mean ratio of two sections of the same curve. One section is the range between 0 and $T - Q$ (which, it may be noted, will be the most precise portion of the data in most instances); the second section depends on the value of the time displacement parameter, q. We call $Y(q)$ the *mean displaced ratio* (MDR) of the data.

When the experimental situation yields $N + 1$ discrete observations, made at equal increments of time, Δt, so that $t_i = i\Delta t$, we define the mean displaced ratio as

$$Y_l = \frac{1}{N - L + 1}\sum_{i=0}^{N-L} y_{i+l}/y_i \qquad l = 0, 1, \ldots, L \tag{6}$$

so that

$$Y_0 \equiv 1$$

$$Y_1 = \frac{1}{N - L + 1}\left(\frac{y_1}{y_0} + \frac{y_2}{y_1} + \frac{y_3}{y_2} + \cdots + \frac{y_{N-L+1}}{y_{N-L}}\right)$$

$$Y_2 = \frac{1}{N - L + 1}\left(\frac{y_2}{y_0} + \frac{y_3}{y_1} + \frac{y_4}{y_2} + \cdots + \frac{y_{N-L+2}}{y_{N-L}}\right)$$

and so on to Y_L. We then use these $L + 1$ calculated amplitudes, Y_l, to replace the original curve, y *vs.* t, with the new curve Y *vs.* t. Since $t_{i+l} = t_i + t_l$, it follows that the new curve has the same number of exponential components and the same time constants as the original curve. This can be seen by substitution of eq 1a into 6, so that

$$Y_l = \sum_{j=1}^{m} \beta_j \exp(-t_l/\tau_j) \tag{7}$$

$$\beta_j = \frac{\alpha_j}{N - L + 1} \sum_{i=0}^{N-L} \exp(-t_i/\tau_j)/y_i \tag{8}$$

and

$$Y_0 = \sum_{j=1}^{m} \beta_j \equiv 1 \tag{9}$$

As we increase the size of the data sections used to obtain the ratio, by letting Q or L get smaller, we increase the amount of smoothing, but decrease the range of the decay in the transformed equation. Since both the original and transformed equation have the same time constants, restricting the range must make analysis more difficult. However, the use of the mean displaced ratio dramatically improves the signal-to-noise ratio, more than compensating for the more restricted range. An example of the effect of the mean displaced ratio can be seen in Figure 1. The scatter shown there is the result of pseudorandom, normally distributed deviations, with a standard deviation equal to 5% of the maximum ordinate. This is more scatter than one is likely to have in most experimental situations of the type being considered here.

In addition, the mean displaced ratio eliminates one unknown quantity, through the equality given by eq 5 or 9. This alone will often make the difference between an adequate and an *inadequate* analysis. While

$$y(0) = \sum_{j=1}^{m} \alpha_j$$

could also be used in this way, without the mean displaced ratio, $y(0)$ is an experimentally observed value, and is therefore subject to error. On the other hand, $Y(0)$ is known exactly, because of the way in which the transformation is defined.

Once the m values of β_j and τ_j are known, the original amplitudes, α_j, are obtained directly from eq 4 or 8.

Analysis of Sums of Exponentials by a Method of Moments. In our earlier paper (Isenberg and Dyson, 1969) we developed a method of moments for analyzing fluorescence decay in the presence of an overlapping excitation curve. With the use of computer simulation, we were able to define those characteristics of the data most important to a satisfactory analysis, and with that as a guide, a monophoton fluorometer was constructed that has now been used to analyze real data for a number of experimental systems (Schuyler and Isenberg, 1971). In this paper we extend the concepts presented earlier to the problem at hand, the resolution of linear sums of several exponentials. Again, computer modeling must be used to define the limits of the analytical technique, and to establish guides for the collection of data. However, at present we wish only to present the theory and to document its usefulness with existing instrumentation and techniques, by showing how it may be used to analyze two-component data obtained from sedimentation equilibrium and from fluorescence decay.

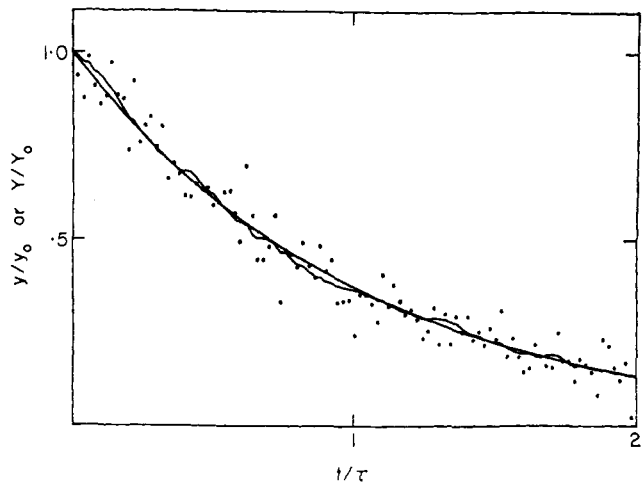

FIGURE 1: Application of the mean displaced ratio to noisy data. The smooth curve represents a single exponential decay. One-hundred pseudorandom, normally distributed deviates, with a standard deviation of 0.05 of the maximum ordinate, are superimposed on this curve to produce the points, which represent the simulated experimental data (y *vs.* t). The noisy curve is the result of using these points to generate the mean displaced ratios (Y *vs.* t according to eq 6, with $L = 0.9N$. Note that the amount of noise is large compared to most actual data.

For the moment, then, we will assume that our data can be represented as a linear sum of exactly m exponentials, according to eq 1a. Then, whether or not we precede the analysis by data smoothing, with the mean displaced ratio or some other function, we replace the experimental data with the $2m$ moments of y or Y defined, in the first case, as

$$\mu_k = \int_0^T y(t)t^k dt \qquad k = 0, 1, \ldots, 2m - 1 \tag{10}$$

The object of this transformation is to replace the original $N + 1$ equations with a smaller number of simpler and, as will be seen, more precise representations of the data.

Where the values of the individual amplitudes and time constants, α_j and τ_j, are available, the moments are calculated as

$$\mu_k = \sum_{j=1}^{m} \alpha_j \left[k!\tau_j^{k+1} - \int_T^\infty \exp(-t/\tau_j)t^k dt \right] \tag{11}$$

Equation 11 is derived from

$$\mu_k = \int_0^T yt^k dt = \int_0^\infty yt^k dt - \int_T^\infty yt^k dt$$

where

$$\int_0^\infty yt^k dt = \sum_{j=1}^{m} \alpha_j \int_0^\infty \exp(-t/\tau_j)t^k dt = k! \sum_{j=1}^{m} \alpha_j \tau_j^{k+1}$$

The integral in eq 11 is evaluated recursively, as successively higher moments are calculated. Thus, if we let

$$I_{j,k} \equiv \int_T^\infty \exp(-t/\tau_j)t^k dt \tag{12}$$

then

$$I_{j,0} = \tau_j \exp(-T/\tau_j)$$
$$I_{j,1} = T\tau_j \exp(-T/\tau_j) + \tau_j I_{j,0}$$
$$\cdots \qquad \cdots$$
$$I_{j,k} = T^k\tau_j \exp(-T/\tau_j) + k\tau_j I_{j,k-1} \tag{13}$$

To obtain the values of the amplitudes and time constants from the transformed equation, we start with some initial guess, and seek to improve it by minimizing the function

$$\epsilon = \sum_{k=0}^{2m-1} \left[\mu_k - \sum_{j=1}^{m} \alpha_j (k! T_j^{k+1} - I_{j,k}) \right]^2 \tag{14}$$

The iterative technique employed is a simple Newton–Raphson procedure, as outlined in the Appendix.

The transformation to moments reduces the original $N + 1$ exponential equations to $2m$ equations which are much more precise, because of the integration involved, and which are no longer dominated by their exponential terms. In fact, if the decay could be followed for a long enough time, the exponential term would disappear entirely, leaving a set of equations that could be solved analytically instead of by iteration

$$\lim_{T\to\infty} \mu_k = k! \sum_{j=1}^{m} \alpha_j \tau_j^{k+1} \tag{15}$$

In most cases one will find that, although $I_{j,k}$ cannot be neglected, it represents only a small perturbation, thereby making the solution of eq 11 much easier than solving the original eq 1a.

In practice, we routinely employ the mean displaced ratio to smooth the data, thus forming the moments from (Y_i,t_i) instead of (y_i,t_i). This also eliminates one unknown, through application of eq 9, and improves the numerical integration necessary to form the moments. With the m values of β_j and τ_j, the amplitudes of the original curve are obtained from eq 8.

Materials and Methods

Fluorescence curves were obtained using the time-decay fluorometer described by Schuyler and Isenberg (1971), with 15,000 lamp flashes/sec and 4.5×10^{-9} sec/channel in the pulse height analyzer. A Corning 7-37 filter was placed between the exciting lamp and sample, and fluorescence was observed through an Optics Technology 552 high-pass interference filter. The collection rate was about 2000 photons/sec.

The samples consisted of 3×10^{-5} M pyrene (Eastman, twice recrystallized from ethanol) in Spectral Grade cyclohexane (Matheson Coleman & Bell). Two degrees of oxygenation were used to provide two different fluorescence lifetimes. One sample was sealed in a glass cuvet at 0.05 atm. The second was frozen, evacuated, thawed, frozen, and evacuated again before the final sealing, resulting in a much lower oxygen level.

Ultracentrifugation was performed in a Spinco Model E (Beckman Instruments) analytical ultracentrifuge, equipped with the interference optical system described by Gropper (1964), an electronic speed control, and temperature control. Alignment and focusing of the optical system generally followed the procedure of Richards *et al.* (1971a,b). Alignment was checked before and after the equilibrium experiments with the test described by Dyson (1970). The interference patterns were recorded on Kodak II-G spectroscopic plates, with either a Wratten 77A or Baird-Atomic 5461A interference filter. The plates were read on a Nikon 6C comparator equipped with a digitizer similar to the one described by Ansevin (1967) (L & W Electronics, Houston, Texas). The micrometer positions of the comparator were recorded either on punched paper tape, or read directly from the digitizer to the time-shared Control Data Corp. 3300 computer maintained by the Oregon State University Computer Center. (All subsequent analyses of the data were performed on this computer, in the time-shared mode, and with programs written locally.) The plates were read at selected coordinates, in order to provide equal increments of $r_0^2 - r^2$ in all experiments.

Samples for the sedimentation equilibrium experiments consisted of crystalline bovine serum albumin (Sigma Chemical Co.) and hen ovalbumin (Nutritional Biochemicals Corp.). The proteins were freed of dimers and higher aggregates by chromatography on Sephadex G-75 (Pharmacia Fine Chemicals, Inc.). Approximately 10 mg of the protein, in 0.5 ml, was placed on a 15×250 mm column, preequilibrated with 0.1 M sodium acetate buffer (pH 5.05 at 22°) containing 0.005 M β-mercaptoethanol. The material was chromatographed at 9 ml/hr at 4°. The same buffer was used in the sedimentation equilibrium experiments.

Ultracentrifugation was carried out using the high-speed technique of Yphantis (1964), including the six-channel centerpiece described by him. Sapphire windows were employed. Base lines were obtained after each experiment by shaking the cells and returning them to speed, or by refilling the cells with water, or both. Equilibrium was assumed when no detectable changes occurred in the pattern (as measured with the microcomparator) over a 2- to 4-hr period. The reported molecular weights of bovine serum albumin and hen ovalbumin are based on partial specific volumes ($\bar{v}$) of 0.734 and 0.748, respectively (*cf.* Sober, 1970).

All graphs presented here were plotted with a Hewlett-Packard 7200A computer-driven X–Y recorder, with programs written for, and executed on, the time-shared system mentioned above.

Results

Fluorescence Decay. The fluorescence data, taken directly from the pulse height analyzer but corrected for background, are shown in Figure 2. The two lower curves, representing two states of oxygenation of the pyrene solutions, were added point by point to obtain the upper curve, thus simulating a two-component system. To test the method of analysis, this simulation is preferable to using a sample in which the two components are actually present in the same solution, since it avoids the possibility of species interaction, and makes it possible to independently determine the individual amplitudes and time constants.

The resulting two-component curve was analyzed by the procedures of this paper. The characteristics of the individual components, both predicted and found, are given in Table I. The effect of smoothing, through the mean displaced ratio, is clearly seen in the results. Note that when $L + 1$ is reduced to below 135, the greater degree of smoothing does not compensate for the accompanying reduction in the range of observation. (The apparent trend in the parameters can be attributed to the particular pattern of random errors in the data and the way in which they are weighted by the mean displaced ratio.) While no one analysis gave the closest value for all four parameters simultaneously, the differences be-

TABLE I: Analysis of Two-Component Fluorescence Data.[a]

	α_1 (Photons/Channel)	τ_1 (nsec)	α_2 (Photons/Channel)	τ_2 (nsec)	RMS[b]
Separately determined[c]	33,830	124.7	22,106	331.6	
Combined curve					
Without smoothing[d]	25,340	103.3	30,660	286.0	198.1
Smoothed[e] $L + 1 = 145$[f]	32,330	125.8	22,940	324.4	143.1
140	33,070	127.6	22,170	329.5	143.3
135	33,370	128.4	21,860	331.4	143.5
130	33,880	130.0	21,280	335.2	144.9
125	34,090	130.6	21,040	336.8	145.9
120	34,190	130.9	20,930	337.5	145.9
110	35,490	134.7	19,470	348.6	154.4
100	34,760	132.0	20,400	341.5	146.0

[a] See Figure 2. [b] Root-mean-square residual, from eq 16. [c] From a linear least-squares fit to the individual logarithmic curves shown in Figure 2. [d] See Figure 3 for a graphical representation. [e] Smoothed by MDR, eq 9, from the $(N + 1) = 150$ observed points. [f] See Figure 4 for a graphical representation.

tween the first three analyses shown in Table I are very small, and may be considered to be experimental error. For simplicity, one may choose the analysis yielding the lowest root-mean-square (RMS) difference between the original and reconstructed curves. This parameter is calculated from the residuals, R_i, according to

$$RMS = \sqrt{\sum_{i=0}^{N} R_i^2/(N+1)}$$

where

$$R_i = y_i - \sum_{j=1}^{m} \alpha_j e^{-t_i/\tau} \qquad (16)$$

Two of the analyses are shown in Figures 3 and 4. In both cases, the calculated curve is plotted on top of the original data, and the residual plot is presented above the decay curve. As can be seen, the residuals are small in both plots, though the better analysis is characterized by their more random distribution.

Sedimentation Equilibrium. The sedimentation equilibrium patterns of bovine serum albumin alone, ovalbumin alone, and a mixture of the two were analyzed. Again, a composite two-component curve was constructed from the point-by-point addition of two individual curves, as shown in Figure 5. This avoids the possibility of protein interaction and permits an independent determination of the apparent amplitudes—in this case the $C_j(0)$ from eq 1b. However, an actual mixture of the two proteins, at equilibrium in the same solution, was also analyzed.

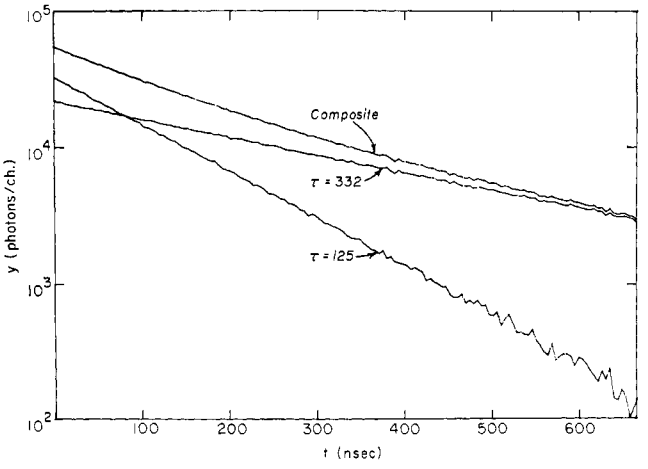

FIGURE 2: The fluorescence decay of pyrene solutions at two states of oxygenation. The upper curve is the point-by-point addition of the two lower curves. The latter represent the actual fluorescence decay, corrected for background. The analysis of these curves gives the results presented in Table I.

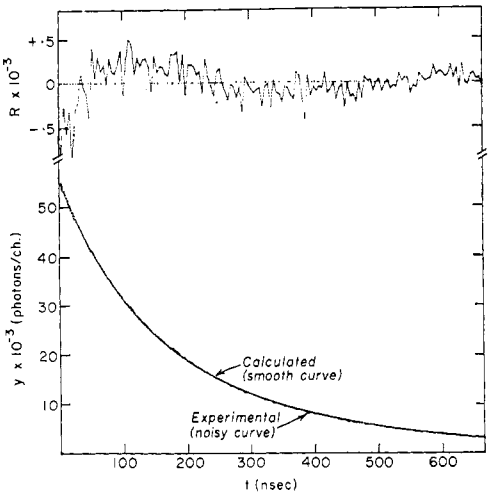

FIGURE 3: Analysis of two-component fluorescence data without prior smoothing. The smooth line is calculated from the constants given in Table I. The noisy line directly on top of it represents the 150 points of original data, with the points connected. The residuals, greatly magnified, appear above the corresponding part of the decay. Both y and R are given in photons per channel of 4.5 nsec. Note that although the original and calculated curves appear to coincide at most points, the time constants and amplitudes of the calculated curve are not satisfactory estimates of the known values.

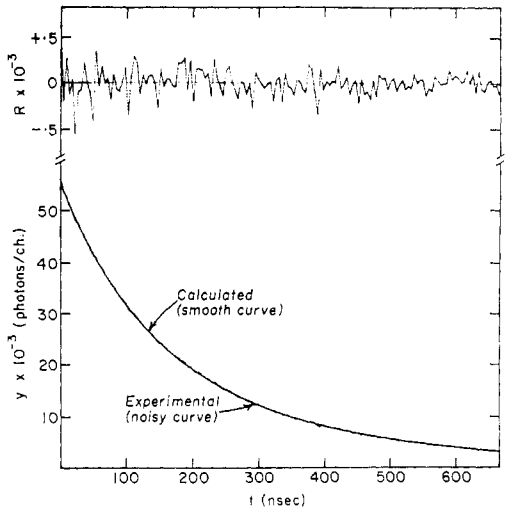

FIGURE 4: Analysis of two-component fluorescence data with smooth ing. Same as Figure 3, but mean displaced ratios are used to smooth the data prior to analysis. The smooth curve is calculated from the constants given in Table I with $L + 1 = 145$.

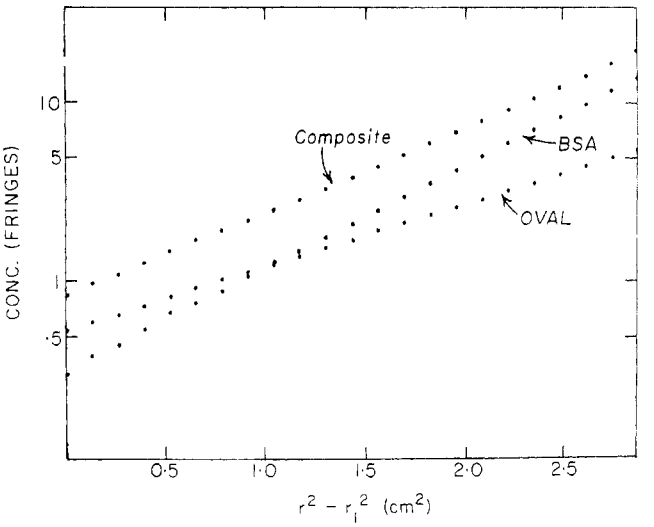

FIGURE 5: The sedimentation equilibrium of bovine serum albumin (BSA) and hen ovalbumin, at 18,054 rpm and 4.9°. Initial concentrations were 1.3 and 1.0 mg per ml, respectively, with each protein run separately in the acetate buffer described in the Methods section. The equilibrium patterns were measured at selected coordinates to provide 23 equal increments of Δr^2 in each case. The two individual curves were then added point by point to obtain the composite curve shown. Concentrations are given as fringe displacements, where 1 mg/ml is approximately four fringes.

An analysis of the constructed two-component data is presented in Figure 6 and Table II. The equivalent analysis of the actual two-component mixture is shown in Figure 7. As can be seen, the predicted and calculated values for the molecular weights and $C_j(0)$ agree very closely, though the value of M for the bovine serum albumin is, in every case, slightly higher than the generally accepted value of 66,500. This could be due to a failure to completely free the protein from aggregates before carrying out the experiments.

Discussion

The advantages of the mean displaced ratio (MDR) as a means of smoothing data are twofold. (1) It reduces random deviations from the true decay, without the risk of introducing systematic errors by impressing an inappropriate analytical model on the data. (2) It eliminates one unknown by exactly defining the sum of the β_j's (*i.e.*, the amplitudes—*cf.* eq 5 or 9). Most other numerical filters fail in one or the other of these categories. The nearest relative of the MDR would be the autocorrelation function, defined as

$$\bar{y}(s) = \lim_{T\to\infty} \int_{-T}^{T} y(t)y(t+s)\mathrm{d}t \qquad (17)$$

With the limits of integration adjusted to equal the actual limits of the data, the autocorrelation function would have many of the properties of the mean-displaced ratio. However, because it uses the product, rather than the ratio of the data points, autocorrelation does not uniquely define the sum of the amplitudes, and therefore does not reduce the number of unknown quantities that must be established by the subsequent analysis.

In using the MDR on real experimental data, one is faced with the problem of assigning the limit of integration—that is, the value of Q or L in eq 2 or 6, respectively. In the example given in Table I, the RMS deviation (or residual) between the data and smoothed curve adequately describes the best choice of integration limit. However, we have found some instances in which this is not true, so it appears that one cannot choose an RMS test as a goodness of fit cirterion, valid for all cases. As an alternative, one could calculate the correlation coefficient or regression coefficient of the residuals, but we find that these parameters are even less reliable than the RMS.

At the present time, therefore, we continue to use the RMS of the residuals to choose the optimum amount of smoothing, but we emphasize that caution is required in this procedure. Our experience to date seems to indicate that Q or L may be set to about 0.9 of T or N, respectively, and that further smoothing of the data, by reducing the value of Q or L, is not likely to compensate for the accompanying restriction of range.

TABLE II: Analysis of Sedimentation Equilibrium Data.[a]

	Ovalbumin		Serum Albumin		
	$C_1(0)$	M_1	$C_2(0)$	M_2	RMS[b]
Separately determined[c]	5.52	44,380	14.1	67,640	
Analysis of combined data[d]	5.66	43,230	14.0	68,590	0.022
Analysis of actual mixture[e]	(4.23)	41,420	(9.90)	68,261	0.011

[a] Concentrations are given as fringe displacements, where 1 mg/ml is approximately four fringes. [b] Root-mean-square deviation, in fringes (see eq 16). [c] From the individual data sets shown in Figure 5, by a linear least-squares procedure. [d] Represented graphically in Figure 6. Total number of points $(N + 1) = 23$. Mean displaced ratios $(L + 1) = 21$ (*cf.* eq 9). [e] Represented graphically in Figure 7. Note that these amplitudes $[C_j(0)]$ cannot be compared to those above. Number of points $(N + 1) = 47$. MDRs $(L + 1) = 42$.

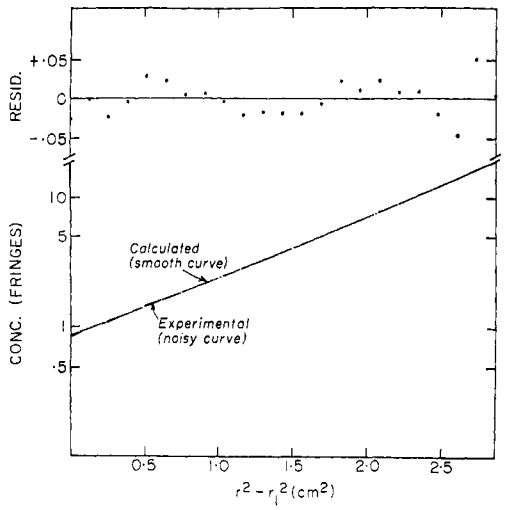

FIGURE 6: Analysis of sedimentation equilibrium data from a constructed two-component curve. The composite data shown in Figure 5 are analyzed, with the results given in Table II. Note that the original and calculated curves follow each other within the width of the pen line throughout most of the pattern. R and C are in fringes.

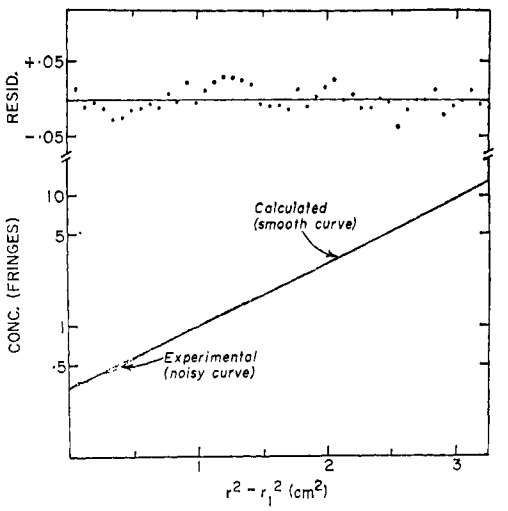

FIGURE 7: Analysis of sedimentation equilibrium data obtained from a mixture of two proteins. A mixture containing approximately 0.5 mg/ml of serum albumin and 0.7 mg/ml of ovalbumin was allowed to reach sedimentation equilibrium at 18,054 rpm and 4.9°. The resulting interference pattern was examined at 47 radial positions, chosen to provide equal increments of $(r^2 - r_1^2)$. The results of the analysis are given in Table II. Note that the original and calculated curves appear to coincide at most points.

While the mean displaced ratio results in a smoother curve, this only facilitates the subsequent analysis; it does not provide the values of the individual decay constants or amplitudes. For that we employ a method of moments, as described. By replacing the original curve with its moments, we reduce the amount of data to be manipulated during the analysis, working instead with a smaller number of more precise points.

The effect of this transformation on the level of noise is marked. If

$$y(t)^{\text{obsd}} = y(t)^{\text{true}} + R(t)$$

where $R(t)$ is a residual, then

$$\mu_k^{\text{obsd}} = \mu_k^{\text{true}} + \int_0^T R(t)t^k \mathrm{d}t$$

In other words, the moments calculated from the observed data will deviate from the actual moments of the decay process only to the extent that the moments of the residuals do not vanish. With a large number of points and a random distribution of residuals, the observed and true moments will be very nearly equal.

The use of moments has several properties that recommend it over other possible transformations. First, it may be seen from eq 15 that the moments are related in a particularly simple way to the amplitudes and time constants. This relationship is not dominated by exponentials, avoiding the most objectionable feature of the original representation (eq 1a or 1b). Second, although each moment is formed from the entire set of data, each is sensitive to a different portion of the curve. This can be seen in Figure 8, which plots the integrand, yt^k, against time. Although the actual shape, maximum ordinate, and spacing will depend on the number of components present and their individual characteristics, Figure 8 shows that the higher moments emphasize that portion of the data collected at longer times.

One feature of the analysis that has been neglected here, is the problem of determining m, the number of components. In our earlier paper (Isenberg and Dyson, 1969) we presented a graphical and an analytical procedure to aid in the process. A simpler approach is merely to analyze for successively higher values of m. When m is less than the actual number of components, the residual plot will often indicate this by showing a definite trend. When m is larger than the actual number of components, the analysis is apt to produce an amplitude or time constant that is physically meaningless—for example, a very small or negative amplitude, perhaps combined with a very large time constant, the effect of which could be interpreted as a base-line drift. In any case, the ability to distinguish between two time constants separated by an increment $\Delta\tau$, or the ability to find a decay with an amplitude

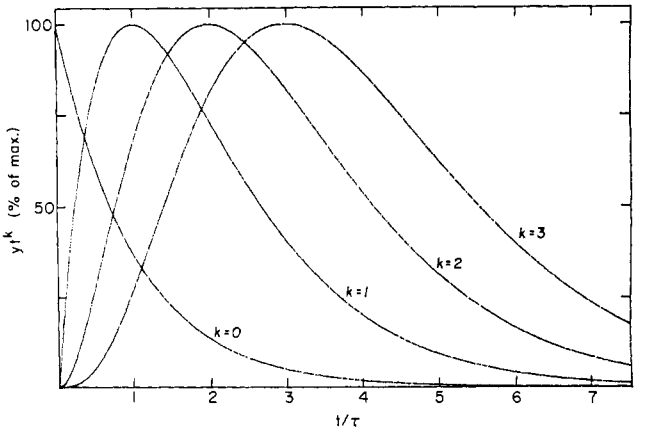

FIGURE 8: The sensitivity of the moments to various portions of a decay curve. The equal spacing of the maximum ordinates is characteristic of a single exponential. The actual shape, spacing, and magnitude of real data will depend on its source, but the higher moments will always be more sensitive to data taken at longer times.

that is small with respect to others, is obviously related to the level of noise.

While we have demonstrated that these procedures are adequate for the analysis of two-component data, as it is obtained from time-decay fluorometry and sedimentation equilibrium using the instrumentation described, this does not represent the limits of their usefulness. For example, we have successfully analyzed simulated three-component decays. The conditions under which one can rely upon such an analysis, and other limiting features of the procedures, are being tested by computer modeling. In addition, applications to more specific cases, such as the sedimentation equilibrium of nonideal associating solutes, are also being investigated.

Appendix

A simple Newton-Raphson iteration is used to solve eq 11 for the m values of α_j and τ_j. The program consists of setting up a linear system of equations, using estimates of α_j and τ_j, followed by iterative updates of the estimates. The linear system is the $2m$ equations

$$f_k = \sum_{j=1}^{m}\left[\left(\frac{\delta\mu_k}{\delta\alpha_j}\right)\Delta\alpha_j + \left(\frac{\delta\mu_k}{\delta\tau_j}\right)\Delta\tau_j\right] \qquad k = 0, 1, \ldots, 2m-1 \quad (18)$$

where $f_k \equiv \mu_k^{obsd} - \mu_k^{calcd}$. The estimate of the kth moment, μ_k^{calcd}, is found from the current value of the α_j and τ_j by using eq 11–13, while μ_k^{obsd} is obtained from the original data, according to eq 10. The system is solved by linear algebra for the error terms $\Delta\alpha_j$ and $\Delta\tau_j$, which are then used to update the estimates of α_j and τ_j: $\alpha_j^{new} = \alpha_j^{old} + \Delta\alpha_j$ and $\tau_j^{new} = \tau_j^{old} + \Delta r_j$. The required partial derivatives

$$\frac{\partial\mu_k}{\partial\alpha_j} = k!\tau_j^{k+1} - \int_T^\infty e^{-t/\tau_j}t^k dt$$

$$\frac{\partial\mu_k}{\partial\tau_j} = (k+1)!\alpha_j\tau_j^{k} - (\alpha_j/\tau_j^2)\int_T^\infty e^{-t/\tau_j}t^{k+1}dt$$

are readily calculated from the current values of α_j and τ_j at each pass, using eq 12 and 13 to evaluate the integrals. Each time the unknowns are updated, the error term

$$\epsilon' = \sum_{k=0}^{2m-1} f_k^2$$

is calculated. The iteration may be stopped when this term is arbitrarily small, or when the changes in α_j and τ_j become insignificant.

The Newton–Raphson iteration requires reasonable initial guesses, which can often be supplied from experience. In other cases, a one-component equivalent is obtained by fitting a straight line to ln y_i *vs.* t_i, and then guesses for multiple components are chosen such that the *sum* of the α_j's equals the hypothetical one-component α, while the *average* of the τ_j's is equal to the corresponding one-component τ.

We routinely scale the data before analysis by dividing each y by the maximum value of y, and each t by the one-component τ mentioned in the last paragraph. This restricts the range of the moments, and prevents disproportionate weighting. With this modification, we have always achieved rapid convergence with two-component data, usually to one part in a million after less than a dozen iterations.

Formation of the moments by numerical integration of the original data (eq 10) necessarily introduces some error. On very noisy data, trapezoidal integration may be as satisfactory as any other quadrature algorithm, but on smooth data, we use

$$\mu_k = \int_0^T yt^k dt = \sum_{i=0}^{N-1} A_{i,k} \qquad \text{for } k \geq 0 \quad (19)$$

where the $A_{i,k}$ are generated recursively, starting with

$$A_{i,0} = (t_i - t_{i+1})(y_i - y_{i+1})/\ln(y_{i+1}/y_i)$$

and proceeding to

$$A_{i,k} = (t_i - t_{i+1})[(y_i t_i^k - y_{i+1}t_{i+1}^k) + kA_{i,k-1}]/\ln(y_{i+1}/y_i)$$

This procedure assumes that the data can be represented as a single exponential between any two adjacent points.

Acknowledgments

The authors thank Mrs. Susan Roxby for technical assistance, Mrs. Linda Haley for her help in preparing the graphs, and Mr. James Hinds for assistance with the computer programming. The fluorescence data was furnished by Mr. Robert Schuyler.

References

Amdur, I., and Hammes, G. G. (1966), Chemical Kinetics: Principles and Selected Topics, New York, N. Y., McGraw Hill.

Ansevin, A. T. (1967), *Anal. Biochem. 19*, 498.

Bay, Z. (1950), *Phys. Rev. 77*, 419.

Bennett, W. R., Jr., Kindlmann, P. J., and Mercer, G. N. (1965), *Appl. Opt. Suppl. 2*, 34.

Caldin, E. F. (1964), Fast Reactions in Solution, New York, N. Y., Wiley.

Dyson, R. D. (1970), *Anal. Biochem. 33*, 193.

Friess, S. L., Lewis, E. S., and Weissberger, A., Ed. (1963), Techniques of Organic Chemistry, Vol. VIII, Investigation of Rates and Mechanisms of Reactions, New York, N. Y., Interscience.

Gardner, D. G., Gardner, J. C., Laush, G., and Meinke, W. W. (1959), *J. Chem. Phys. 31*, 978.

Glass, H. I., and de Garreta, A. C. (1967), *Phys. Med. Biol. 12*, 379.

Gropper, L. (1964), *Anal. Biochem. 7*, 401.

Haschemeyer, R. H., and Bowers, W. F. (1970), *Biochemistry 9*, 435.

Hildebrand, F. B. (1956), Introduction to Numerical Analysis, New York, N. Y., McGraw Hill.

Hvidt, A., and Neilsen, S. O. (1966), *Advan. Protein Chem. 21*, 287.

Isenberg, I. I., and Dyson, R. D. (1969), *Biophys. J. 9*, 1337.

Jablonski, A. (1935), *Z. Physik. 95*, 53.

Laiken, S. L., and Printz, M. P. (1970), *Biochemistry 9*, 1547.

Lanczos, C. (1956), Applied Analysis, Englewood Cliffs, N. J., Prentice Hall.

Marzollo, A. (1969), *Int. J. Control. 9*, 17.

McDonough, R. N., and Huggins, W. H. (1968), *IEEE Trans. Auto. Control AC-13*, 408.
Myhill, J. (1965), *Biophys. J. 5*, 89.
Perl, W. (1960), *Int. J. Appl. Radiat. Isotopes 8*, 211.
Richards, E. G., Teller, D., Hoagland, V. D., Jr., Haschemeyer, R., and Schachman, H. K. (1971a), *Anal. Biochem. 41*, 215.
Richards, E. G., Teller, D., and Schachman, H. K. (1971b), *Anal. Biochem. 41*, 189.
Robertson, J. S. (1957), *Physiol. Rev. 37*, 133.
Schuyler, R., and Isenberg, I. (1971), *Rev. Sci. Instrum.* (in press).
Smith, R. E., and Morales, M. F. (1944), *Bull. Math. Biophys. 6*, 133.
Sober, H. A., Ed. (1970), Handbook of Biochemistry, Cleveland, Ohio, The Chemical Rubber Co.
Svedberg, T., and Pederson, K. O. (1940), The Ultracentrifuge, New York, N. Y., Oxford University Press, Johnson Reprint Corp.
Tao, T. (1969), *in* Molecular Luminescence, Lim, E. C., Ed., New York, N. Y., W. A. Benjamin, p 851.
Van Liew, H. D. (1967), *J. Theor. Biol. 16*, 43.
Wahl, P., and Lami, H. (1967), *Biochim. Biophys. Acta 133*, 233.
Worsley, B. H. (1964), *Commun. Ass. Comput. Mach. 7*, 39.
Worsley, B. H., and Lax, L. C. (1962), *Biochim. Biophys. Acta 59*, 1.
Yphantis, D. (1964), *Biochemistry 3*, 297.

Equilibrium Centrifugation of Nonideal Systems. The Donnan Effect in Self-Associating Systems*

Dennis E. Roark† and David A. Yphantis‡

ABSTRACT: The sedimentation-equilibrium concentration distribution of charged macromolecules is coupled to the equilibrium distribution of other ions, causing the macromolecular concentration to be a more gradual function of radius than if the macromolecules possessed no charge. This Donnan effect decreases the apparent molecular weight averages from their ideal values. A rigorous thermodynamic treatment of this effect is presented for systems containing one or more macromolecular species, such as self-associating systems, and in which the supporting electrolyte may sediment. A general restriction of this theory is that the charge-to-mass ratio of all macromolecular species be the same. An extension of the thermodynamic component definitions of Scatchard leads to simple expressions for the reciprocal apparent weight-average molecular weight. Except when the supporting electrolyte of a highly nonideal solute sediments appreciably, this reciprocal molecular weight average is separable into the sum of an ideal and a nonideal contribution, the latter in the form of a nonideal virial expansion. The nonideal virial coefficients depend on the stoichiometry of the supporting electrolyte: for a uni-univalent salt, the third nonideal virial coefficient is zero; while for a uni-divalent salt, the fourth nonideal virial coefficient is zero. Expressions are given for estimating the limiting macromolecular concentration, below which a particular nonideal virial coefficient has a negligible influence on the experimentally measured molecular weight average.

The equilibrium distribution in a centrifugation field of an ideal monodisperse neutral species sedimenting in an incompressible solvent is a simple exponential function

$$c_i = A_i e^{\sigma_i \xi} \tag{1a}$$

where

$$\sigma_i = \frac{M_i(1 - \bar{v}_i\rho)\omega^2}{RT} \tag{1b}$$

and

$$\xi = r^2/2$$

Here, c_i is the concentration on a weight basis of species i; σ_i has been termed the "effective reduced molecular weight" (Yphantis and Waugh, 1956; Yphantis, 1964); M_i and $\bar{v}_i$ are the molecular weight and partial specific volume of the species, respectively; ρ is the solution density; ω the angular velocity; and r the distance from the center of rotation. This equation can be derived by various mechanistic or, more properly, thermodynamic approaches (see, for instance, Svedberg and Pedersen, 1940; Goldberg, 1953; Schachman, 1959; Williams *et al.*, 1958). The resultant distribution is a consequence of the opposing tendencies of diffusion and sedimentation: the centrifugal field induces solute transport in the radial direction, while diffusion tends to equalize any concentration inbalance (and thus leads to a transport in the centripetal direction). Equation 1 reflects the balance of these two flows and depends

* Contribution from the Departments of Biophysics and Biology, State University of New York at Buffalo, Buffalo, New York (D. E. R. and D. A. Y.), and from the Biochemistry and Biophysics Section of the Biological Sciences Group and Institute of Materials Science at the University of Connecticut, Storrs, Connecticut (D. A. Y.). *Received April 2, 1971.* This investigation has been supported, in part, by Public Health Service Training Grant 5T1-GM-718-09, and by Grants GB-6130, GB-8164, and GB-13790 from the National Science Foundation. A portion of this material was taken from a dissertation submitted by D. E. R. in partial fulfillment of requirements for a Ph.D. degree, State University of New York at Buffalo.

† Present address: Department of Biological Chemistry, Milton S. Hershey Medical Center, Hershey, Pa. 17033.
‡ To whom to address correspondence.