Analysis of Associating Systems Using the Multinomial Theory*

Moisés Derechin

ABSTRACT: A procedure is described for the calculation of equilibrium constants of self-associating systems using equations from the multinomial theory (Derechin, M. (1969a,b), *Biochemistry 8*, 921, 927). This method is illustrated using computer-simulated model systems and experimental results. It is suggested that lysozyme (4°, pH 7.2) and cytidine generate an ideal or close to ideal monomer–dimer–trimer associating systems.

General equations for the calculation of equilibrium constants and virial coefficient of ideal and nonideal systems have been derived before (Derechin, 1968, 1969a,b). These equations require the successive derivatives of the functions $M_1/M_{k,app}(c)$ with $k = n, w, z$ representing, respectively, the number-, weight- and z-average molecular weights, at $c = 0$, with c = concentration in g/100 ml and M_1 = monomer molecular weight. In this paper, methods to obtain these derivatives and their application to the analysis of experimental results are presented and discussed, using data and equations for $M_{w,app}$.

Methods. The multinomial theory (Derechin, 1969a) defines

$$K_2 = X - a \tag{1}$$

$$K_3 = \frac{3}{2}X^2 - 3aX + b \tag{2}$$

$$K_4 = \frac{8}{3}X^3 - 8aX^2 + (8a^2 - b)X + c \tag{3}$$

$$K_5 = \frac{125}{24}X^4 - \frac{125}{6}aX^3 + \left(\frac{125}{4}a^2 - \frac{25}{8}b\right)X^2 + \left(\frac{25}{4}ab - \frac{125}{6}a^4 - \frac{5}{72}d\right)X + f \tag{4}$$

where $X = BM_1$, $a = (dR2/dc)_{c=0}$, $b = (d^{(2)}R2/dc^2)_{c=0}$, $d = (d^{(3)}R2/dc^3)_{c=0}$, $c = -(8/3)a^3 + ab - (1/18)d$, $R2 = M_1/M_{w,app} \times (c)$, $f = (125/24)a^4 - (25/8)a^2b + (5/18)ad + (5/32)b^2 - (1/96)e$, and $e = d^{(4)}R2/dc^4)_{c=0}$. Work to test the methods described here involved the (a) use of computer simulated model cases, (b) analysis of results published by other workers, and (c) analysis of our own experimental results. (a) Model systems were calculated from the definitions of $M_w(c)$ using assumed values of the equilibrium constants and virial coefficient, to give a set of error-free data. In order to simulate real experimental cases, error (random, systematic, or both) was incorporated into the data and computations were made using data within discrete concentration ranges away from $c = 0$. An error in concentration labeling was used as a model for systematic error. (b) Experimental results on cytidine published by Van Holde *et al.* (1969) were used here to establish the associating system generated by this molecule. To obtain the data measurements were made on the continuous line drawn by these workers through the experimental points in Figure 1 of their paper. Then, these measurements were converted into a table of 60 data pairs at intervals of 0.27 g/100 ml covering the range from 0 to 16 g per 100 ml. This set was studied in four different ways, namely, as a full set, and as partial sets in the ranges of 0–10, 2–7, and 8–16 g per 100 ml. In all cases, $M_{w,app} = M_1 = 243.22$ at $c = 0$ was included. (c) Our experimental work consisted in low-speed sedimentation equilibrium experiments using the reference frame method described before (Derechin, 1969c). In filling the channels of the double-sector cells equal volumes of solvent and solution, respectively, were added into each side. No immiscible fluid to produce an artificial base was added. Protein concentrations were measured at 20° in a Hilger Model 154 interferometer for liquids. The instrument was calibrated with sucrose using a specific refractive index increment (SRII) of 0.1438 ml/g calculated from the data of Gosting and Morris (1949). For the calibration of the ultracentrifuge interference system, 30 concentration measurements of lysozyme solutions in phosphate–NaCl buffer (pH 7.2), $I = 0.2$, were made in a double-sector, synthetic boundary cell capillary type. Measurements below 0.2 g/100 ml showed signs of large errors and were not used to obtain the final average (38.15 ± 0.4 for a 1.0 g/100-ml solution). For use with the Hilger interferometer, the SRII used for lysozyme was 0.1888 ml g^{-1} (Halwer *et al.*, 1951). The lysozyme employed was the six-times-crystallized commercial preparation from egg white obtained from Miles Laboratories. The work with lysozyme consisted of six experiments at 4° in the initial concentration range of 0.24–1.46 g/100 ml. A value of $\bar{v} = 0.703$ used by Adams and Filmer (1966) and $M_1 = 14315$ calculated from amino acid composition (Canfield, 1963) were used here for lysozyme.

Computations. The computations were done on a CDC 6400 computer using various computer programs described below, and plots were produced on a Model 563 Calcomp digital incremental plotter. Experimental information is analyzed with the program WEIGHT, for use with data obtained using either absorption or interference optics at low or high speed. This program calculates point concentrations in the required units and ln $c(r)$. Also it produces the least-squares coefficients, a_j, for various polynomials

$$P_i(x) = a_0 + a_1x + a_2x^2 + \ldots + a_ix^i \qquad i = 1, 2, \ldots, 10, \quad x = r^2 \tag{5}$$

where $P_i(x)$ stands for ln $c(r)$.

* From the Departments of Biochemistry and Biochemical Pharmacology and the Center for Theoretical Biology, State University of New York at Buffalo, Amherst, New York 14226. *Received June 28, 1971.* This work was supported by the NIH General Research Support Grant No. 50-8112-D, NIH Grant GM-16726, and the Institutional Research Support Grant No. 50-8865F of this university, and the NASA Grant No. NGR-33-015-002 to the Center for Theoretical Biology. This is paper IV of the series Self-Associative Systems by this author; the preceding three were published in this journal (1968, 1969a,b).

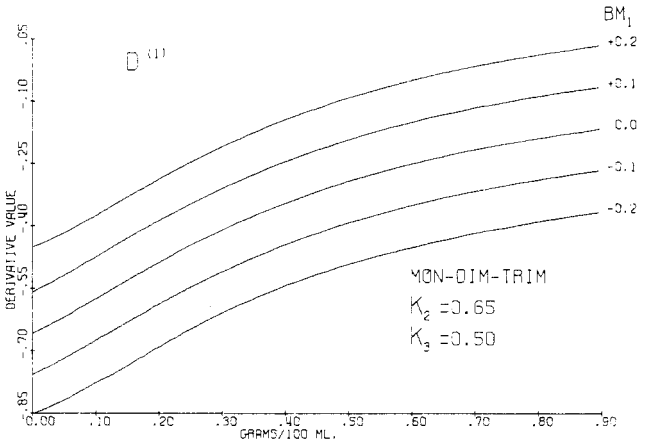

FIGURE 1: First derivative for Adams system. Degree 8 was used to represent $M_{w,app}(c)$.

The molecular weight at each point is then calculated using the equation

$$M_{w,app}(c) = M_{w,app}(r^2) = AP'_i(r) \quad (6)$$

where P'_i is the derivative of the polynomial, and $A = 2RT/(1 - \bar{v}\rho)\omega^2$. Point computations according to eq 6 are made using all polynomials obtained. The program WEIGHT produces plots for all polynomial degrees of ln c *vs.* r^2 and $M_{w,app}(c)$ *vs.* c. The program MWEXR2 works with a data set of the function $M_{w,app}(c)$ *vs.* c, including the vital point $(M_{w,app})_{c=0} = M_1$. This data set is analyzed by the least-squares method to produce polynomials $M_{w,app}(c) = P_i(c)$ (of the form of eq 15 with $x = c$) usually up to degree ten. Beginning with a preselected polynomial, say of degree 2, this computer program calculates for this and each successive higher degree polynomial tables of the functions $M_{w,app}(c)$, $R2(c)$, and the first, second, third, and fourth derivatives, as described below, respectively of $R2(c)$ that start at $c = 0$. The values of these derivatives at $c = 0$ are then incorporated into eq 1–4 as partial coefficients of the various X^i and used to calculate the roots of these equations. For this, each equilibrium constant, one at a time, is assumed zero. Also, the assumption $X = 0$ is made and the equilibrium constants of a hypothetical ideal case are calculated. This procedure leads to the calculation of eleven sets of equilibrium constants and virial coefficients for each polynomial degree used to represent $M_{w,app}(c)$, from which one set is to be selected as representative of the experimental case. For the production of model systems the program LVOUT is used. This program makes use of a set of constants supplied to generate their corresponding $M_{w,app}$ and then proceeds essentially in the same manner as the program MWEXR2. Also, hypothetical systems suggested from the analysis of experimental information are used to produce tables and plots of $M_{w,app}(c)$ and $d^{(r)}R2(c)$. These $M_{w,app}(c)$ are analyzed for correspondence (M_w test) with the experimental $M_{w,app}(c)$. A similar test was conducted for the derivatives (D test) of the various hypothetical systems. Finally, once a very good approximation to some equilibrium constants, say, K_2 and K_3 was available, this program was used to define in greater detail the properties of the successive derivatives of variants of the system (S test) within a restricted range, say holding K_2 constant and varying K_3, or using K_2 and K_3 constants and varying K_4.

Polynomial and Data Selection. A tentative polynomial selection to represent the experimental data is first made after examining plots of ln c *vs.* r, and choosing that polynomial degree that best fitted the data. Usually these data produced polynomials up to degree four, and preference being for degree four or three. Then, plots of $M_{w,app}$ *vs.* c obtained with the various polynomials were compared and sections of these data showing obvious discrepancies of representation were removed. In addition the polynomials and data for $M_{w,app}(c)$ from various experiments were examined so that the final selection should result in the most continuous and smooth curve for experiments in successive concentration ranges. This condition is based on the assumption that the associating species being analyzed belong to only one thermodynamic system.

Derivatives. After a polynomial degree had been selected to represent the function $M_{w,app}(c)$, the equation was used to calculate values of the function at small intervals of the independent variable (*e.g.*, at 0.001 g/100-ml increments). Using these values, $R2(c)$ is computed and therefrom, its first derivative by making

$$\left(\frac{dR2}{dc}\right)_i = \frac{\Delta R2}{\Delta c} = \frac{R2_{i+1} - R2_{i-1}}{c_{i+1} - c_{i-1}} \quad (7)$$

where i stands for the ith increment of the independent variable. The data produced by eq 7 were analyzed by the least-squares method to produce polynomials of the form of eq 5 whose first coefficient is the desired value, $(dR2/dc)_{c=0}$.

A polynomial so produced was selected and used to calculate the final tables of the function $(dR2/dc)(c)$. The process of differentiation, polynomial regression, recalculation of the calculated function using a polynomial, and new differentiation was repeated as required. In all cases, the first coefficient of the selected polynomial, that is, the intercept of the function with the y axis was taken as $(d^{(r)}R2/dc)_{c=0}$.

As shown below (D and S tests), although $(d^{(r)}R2/dc)_{c=0}$ are the only numbers used in the computation of equilibrium constants, the various $(d^{(r)}R2/dc)(c)$ are also calculated since they are valuable tools in the analysis and can reveal "hidden" properties of the various self-associating systems that are important for diagnostic purposes. Unless otherwise stated, *data* refer to sets of $M_{w,app}(c)$ *vs.* c from either experiment or computer-simulated systems; *derivatives* refer to $d^{(r)}R2/dc$ and may also be found written as $D^{(r)}$, with $r = 1$–4. The abbreviated expressions monomer–dimer, monomer–dimer–trimer, and monomer–dimer–trimer–tetramer, respectively, are used in place of monomer–dimer system, monomer–dimer–trimer system, and monomer–dimer–trimer–tetramer system. The monomer–dimer–trimer system with $K_2 = 0.65$ and $K_3 = 0.50$ that was extensively used by Adams *et al.* (1966) and Adams (1967) to test their method is referred to here as Adams system.

Results

Computer-Simulated Systems. The plots of $M_{w,app}$ *vs.* c showed the macroscopic appearance expected from the constants used to generate each system. Also, some additional properties of the data generated with the various systems became apparent after polynomial regression and successive differentiation. (a) The correct calculation of equilibrium constants and virial coefficient for error-free systems was possible using values of $M_{w,app}(c)$ calculated with only certain of the polynomials that appeared to fit the data. Al-

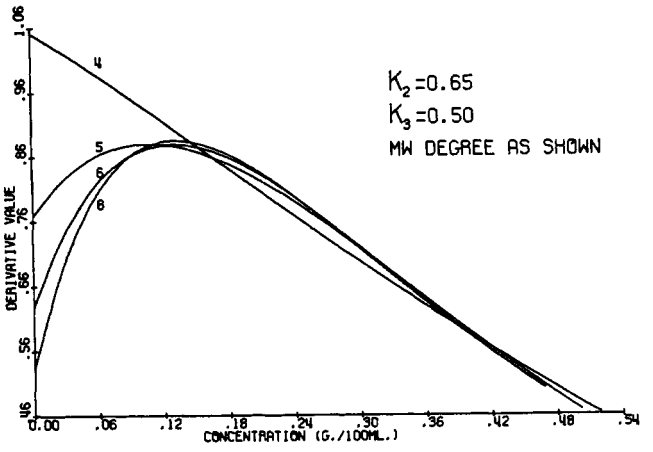

FIGURE 2: Second derivative for all monomer–dimer–trimer with $K_2 = 0.65$ and $K_3 = 0.50$ ideal or nonideal. Results from various degrees to represent $M_{w,app}(c)$ are compared.

though $M_{w,app}(c)$ obtained with the Adams' system appeared in the plots as equally well fitted with all polynomials starting with a quadratic, essentially error-free constants could be calculated with polynomials of degree six to ten only. This result permitted us to define as "correct" the polynomial that led to the correct reproduction of the original constants. It follows that the derivatives were correct in these cases only. (b) All $D^{(1)}$ of a given system obtained with a "correct" polynomial for $M_{w,app}(c)$ were found to consist of a set of parallel curves (Figure 1), with that for any particular nonideal case being displaced along the y axis by a value equal to BM_1. In the particular case of Adams system an inflection point could also be seen in the plots of $D^{(1)}$. (c) For any given system, ideal or nonideal, all $D^{(2)}$ were identical (Figure 2). In the particular case of the Adams system, a maximum (to be expected from the inflection point in $D^{(1)}$) could be seen at least when "correct" polynomials (degree 6–10) had been used. Small variations were obtained in the results obtained with these when different degrees were compared but these differences had a negligible effect. In particular, if $K_2 = 0.65$ and K_3 is varied in a monomer–dimer–trimer the maximum in plots of $D^{(2)}$ was displaced toward $c = 0$ on decreasing K_3 to disappear for $K_3 = 0.32$ and lower values (Figure 3). (d) After data (error free) were omitted in the concentration range close to $c = 0$, up to various concentrations, the least-squares method was capable of completely reproducing the original full data set and of correctly calculating the successive derivatives and all the constants used to generate the system. Starting with data from 0 to 0.9 g per 100 ml this result was completely independent of the extent of data removed up to a critical concentration, $c = 0.39$ g/100 ml. The removal of one single additional data pair prevented the correct reproduction of the original data or the correct calculations to be derived therefrom.

When error was introduced in the data, the computations with higher degree (*e.g.*, 8–10) polynomials resulted in large error in the computed constants. Smoothing of data containing random error resulted in equilibrium constants with greater error than were obtained without smoothing.

Cytidine. As expected from good data the analysis produced very consistent results from polynomials up to degree ten for all concentration ranges studied. The monomer–dimer–trimer–tetramer and monomer–dimer–trimer systems, respectively, with $K_2 = 0.06$, $K_3 = 0.003$, and $K_4 = 4 \times 10^{-4}$ or

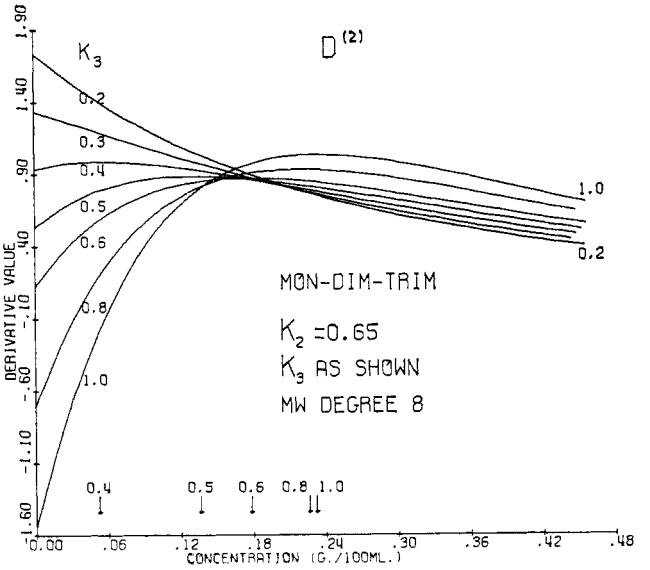

FIGURE 3: Effect of K_3 on the shape of $D^{(2)}$ in a monomer–dimer–trimer. Arrows show the concentration at which the maximum occurred in each case.

$K_4 = 0$, appeared to have an equal chance of representing the experimental case. The value calculated for K_5 was extremely small in all cases and was taken as zero. Some results suggested also the possibility of a slight nonideality for the same two cases. Within the range of variation observed for the various equilibrium constants (ideal), it was possible to (M_w test) satisfactorily fit calculated to experimental data in both cases. Close correspondence was also observed in both cases for calculated and hypothetical $D^{(1)}$. This was an indication that cytidine generates a nearly ideal associating system but did not resolve the ambiguity about the presence or absence of tetramer. This was done using $D^{(2)}$ (Figure 4). The close correspondence in geometrical shape of $D^{(2)}$ for a monomer–dimer–trimer compared with the sharp divergence for the system that assumes the tetramer is evident (Figure 4). The nonideal cases, irrespective of the assumed presence of tetramer, failed to fit (M_w test) the data. It was concluded that cytidine in the conditions studied by Van Holde *et al.*

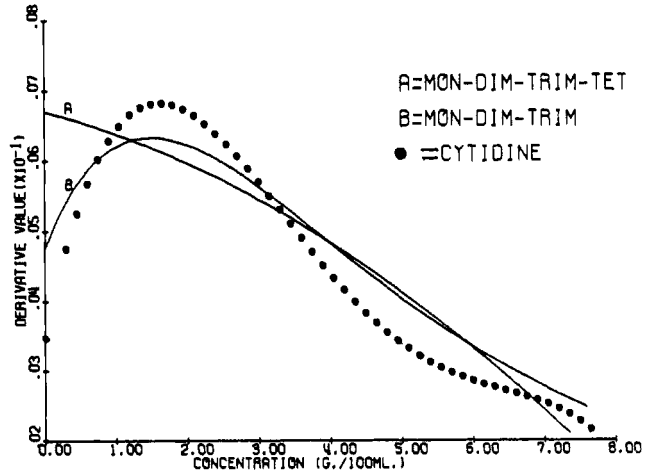

FIGURE 4: D tests for cytidine using $D^{(2)}$. Symbols indicate results from "experimental" cytidine and lines correspond to model systems, assuming tetramer present (A) or absent (B).

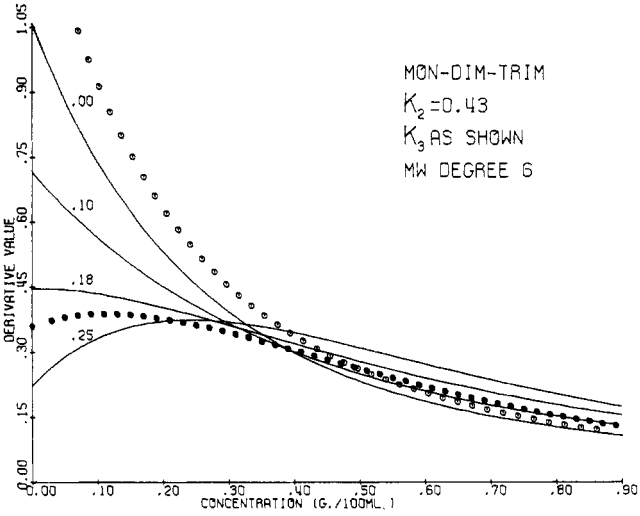

FIGURE 5: Tests using $D^{(2)}$. Filled symbols correspond to $D^{(2)}$ calculated from experimental data (lysozyme). These are compared with $D^{(2)}$ from model systems: open circles are from a monomer-dimer with $K_2 = 0.49$; the continuous lines correspond to various monomer-dimer-trimer using $K_2 = 0.43$ and K_3 as indicated on the figure.

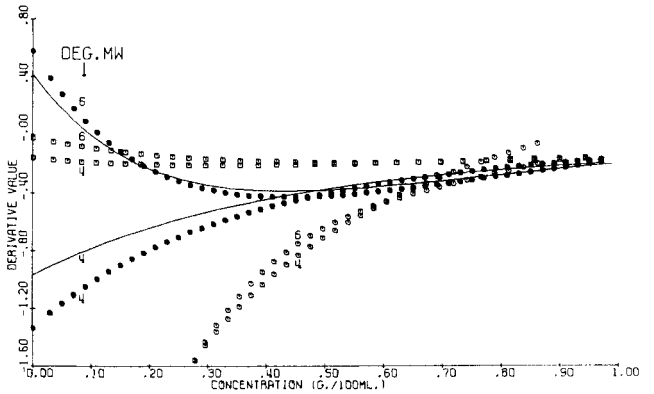

FIGURE 6: Combined D and S test using $D^{(3)}$. Filled symbols show results using experimental data for lysozyme. The model systems are for a monomer-dimer-trimer with $K_2 = 0.43$, $K_3 = 0.20$ (continuous line), monomer-dimer with $K_2 = 0.49$ (open circles), and the monomer-dimer-trimer-tetramer with $K_2 = 0.43$, $K_3 = 0.18$, and $K_4 = 0.10$ (open squares).

(1969) generates an ideal monomer-dimer-trimer with values of K_2 and K_3 as indicated above.

Lysozyme. Data from four experiments were used in the final computations. Polynomials for $M_{w,app}(c)$ up to degree eight were very consistent as to the associating system they suggested. This was taken as an indication of the good quality of the experimental data used. Some error in degree nine and larger error in degree ten was apparent.

Three systems deserved analysis, namely, the monomer-dimer, the monomer-dimer-trimer, and the monomer-dimer-trimer-tetramer. The amount of tetramer suggested by the results in the latter group was very small but this ambiguity permitted examining the fine discriminating capability of the method. The M_w test showed that the monomer-dimer equilibrium, grossly departed from experiment whether using constants as calculated here ($K_2 = 0.25$, $BM_1 = -0.12$) or assuming some computational error and trying "adjusted" constants, say, $BM_1 = -0.08$. The various monomer-dimer-trimer-tetramer and monomer-dimer-trimer suggested by our computations were capable of reproducing the experimental data within the range of variation of the constants ($K_2 = 0.40$–0.44, $K_3 = 0.12$–0.20, and $K_4 = 0.0$–0.07) for assumed ideal cases but failed to reproduce the data if significant nonideality was assumed. A first approximation to the description of the system was obtained with an S test for $K_2 = 0.43$ using $D^{(2)}$ (Figure 5). This test suggested that $K_3 = 0.18$–0.25. Using $K_2 = 0.43$ and $K_3 = 0.18$ or $K_3 = 0.20$, $D^{(3)}$ was used to investigate the presence of tetramer. This test showed (Figure 6) best correspondence for $K_4 = 0$. On changing the values of K_4 by discrete increments the plots of $D^{(3)}$ departed increasingly from the corresponding plots obtained from the experimental data. It was concluded that lysozyme at pH 7.4 and 4° generates a monomer-dimer-trimer with $K_2 = 0.43$ and $K_3 = 0.20$. When $D^{(1)}$ for this hypothetical system was compared with its experimental counterpart a close coincidence was observed, thus indicating that in the experimental conditions used here the associating system of lysozyme is ideal or very close to ideal. For comparison the error-free monomer-dimer with $K_2 = 0.49$ and $BM_1 = -0.03$ was examined for M_w and D correspondence. These values were suggested by Adams *et al.* (1966) to represent the associating system of lysozyme at pH 6.7 and 15°. Although with these values of K_2 and BM_1 the plots of M_w are substantially lower than our experimental results, it was possible within the deviations suggested for these constants by Adams *et al.* (1966) to obtain an acceptable correspondence. On the contrary D correspondence was not obtained (Figure 5-6).

Discussion

Initial uncertainty concerning the applicability of the multinomial theory arose from the belief that experimental information was needed in the concentration range close to $c = 0$ where it was either not available or suspected of containing relatively large error. Results using computer-simulated systems showed that this initial information at low solute concentration is not essential at least for some systems since it can be derived by polynomial regression from accurate measurements in a concentration range well removed from $c = 0$. In this respect the observation of an all-or-none behavior of the least-squares method in its ability to reproduce the original function and its successive derivatives after data had been removed up to a critical concentration was a most striking observation. (For lysozyme, the critical point was at 0.70 g/100 ml.)

In addition the analysis of experimental work has shown that information at high solute concentration may be more useful than in low concentration range since it greatly helps to (1) define the associating system, and (2) achieve precision in determining the constants.

It is usually recognized that error may be random or systematic. The main source of random error in our computations arises from the measurement of the photographic plates. Except near the meniscus end of the interferogram where the interference fringes tend to be flat this error is usually very small and can be minimized to a negligible value by measuring the plates five or more times and averaging before doing the computations.

Systematic error may originate from a number of sources and is frequently inaccessible to detection or precise correction. Each individual experiment may contain simultaneously several forms of systematic error; some of which

have been discussed to varying extents by several workers. Adams and Filmer (1966) examined briefly a variety of sources of systematic error; Yphantis (1964) discussed problems arising from solvent gradient mismatch on the two sides of the double-sector cell and the correction for cell distortion in experiments at high speed; Teller *et al.* (1969) rigorously discussed the effect of error in concentration "labeling" on the computation of the various averages molecular weights. Perhaps, not emphasized enough, errors in concentration labeling may arise from measurements of the initial solute concentration using the synthetic boundary cell, as suggested by the scattering observed in this type of measurements. Although the use of an average from many such measurements in a cross calibration of the ultracentrifuge and Hilger interferometers has greatly reduced the error in our work, further improvement seems desirable. A conceivable possibility is the use of a continuous solid medium possessing two sections differing in refractive index (or thickness) within the accessible range of measurement of the two instruments to be cross calibrated. The two optical sections could be joined by a portion of refractive index (or thickness) gradient. This method would allow the most precise transfer of information from the external to the ultracentrifuge interferometer and completely remove the need for synthetic boundary cell measurements. Although Yphantis' (1964) high-speed sedimentation equilibrium experiments offers an alternative to synthetic boundary cell calibration, the method has its own sources of error, such as those arising from cell distortion. Improved ultracentrifuge cells for high-speed equilibrium experiments such as those described (Ansevin *et al.*, 1970) recently promise results accessible to treatment by these procedures.

The need of high-degree polynomials to represent the function $M_{w,app}(c)$ and provide a basis from which to calculate the equilibrium constants was another unexpected result, particularly since in most cases no difference was apparent upon examination of plots produced with any polynomial starting with one of a much lower degree. When significant error was present in the data, however, agreement in the results was apparent for low-degree polynomials only. Thus, the ability of the experimental data to permit the calculation of uniform results from polynomials up to the higher degrees was taken as an indication of their good quality and suitability for computations using the multinomial theory. So far as possible the higher degree polynomials were preferred for calculating the function $M_{w,app}(c)$ in view of their higher information content relative to the successive derivatives. In the analysis of lysozyme experiments it was observed that for the successful computation of polynomials of higher degree no overlap of data should be allowed when pooling information from various experiments, however close the coincidence of the data may appear and whether this overlap occurs in the full concentration range (duplicate experiments) or only at the ends of the data sets.

No rigorous mathematical treatment or formulae is known to us to lead to an infallible selection of the correct polynomial. The statistical F test (see Snedecor, 1959) is commonly used for the selection of the best polynomial. The F test was the criterion that permitted our programs to carry out the complete set of cycles and computations before options were selected from other criteria. In our experience this test is a good guide for polynomial selection when error-free data are analyzed. When the data contained random error, it was occasionally observed that the best polynomial by the F test was that which best reproduced the erroneous data.

The limit of the applicability of this method coincides with the limit of our ability to adequately represent the data by means of a function that can then be used for further computations. The systems examined in this work were accessible to correct representation by means of least-squares polynomials within degree one to ten and using data in the concentration range (for proteins) of 0–2.0 g/100 ml. However, if $K_2 = 1.0$ no least-squares polynomial of degrees 2–10 was able to correctly represent $M_{w,app}(c)$ in the range 0–2.0 g/100 ml, but correct computations were still possible in ranges below 1.0 g/100 ml. Similarly, if $K_2 = 10$ then, the least-squares polynomials degrees 2–10 was suitable only in concentration ranges below 0.2 g/100 ml. Data of this type would require either a polynomial of degree higher than ten, or some other method of representation. The positive observations of this work have been dependent also on the ability to carry large-scale, high-speed computations and the availability of extensive and rapid plotting facilities. In addition to permitting the visualization of the influence of a given parameter, the examination of diagrams permitted the succession of decision making and planning of this research in a minute fraction of the time that would have been required otherwise by the examination of jungles of tables.

Our conclusions for cytidine disagree with those of Van Holde *et al.* (1969). Their scheme assumes that all possible species are present and all equilibrium constants are equal. Possibly, these two preconditions of the analysis are the source of the disagreement. No assumption was made here as to the values of the constants or as to the possible species present.

Although we have worked with a commercial preparation of lysozyme we feel confident about the purity of this preparation. Sophianopoulos *et al.* (1964) failed to detect any difference tn the behavior of commercial lysozyme preparations when compared with their highly purified protein. The excellent coincidence of both ends of the plots with five of our experiments reinforces our confidence in the purity of the protein used.

Bruzzesi *et al.* (1965) suggested from light-scattering and sedimentation velocity experiments that lysozyme solutions at pH 4.5–7 consisted of equilibria containing species higher than the dimer. Although the results of Adams *et al.* (1966) with lysozyme appear to disagree with ours their experimental conditions are somewhat different from ours. At pH 6.7 and 15° they suggested this protein generates a monomer–dimer with $K_2 = 0.49$ and $BM_1 = -0.03$. However, since decreased association should be expected at pH 6.7 (Sophianopoulos *et al.*, 1964) and at 15° (Adams *et al.*, 1966) the difference in the associating system suggested by the work of Adams *et al.* (1966) and ours would be understood.

Sophianopoulos *et al.* (1964) studied lysozyme equilibria using short-column sedimentation equilibrium experiments and concluded that this protein undergoes a reversible dimerization in the pH range 5–9. Their conclusion is based on the analysis of plots of M_Z *vs.* $1/M_w$, from experiments at various pH values assuming ideal behavior for lysozyme. Since the plots of M_Z *vs.* $1/M_w$ involve an extrapolation to a point far removed from the data any scattering in these may greatly change the results. It is possible therefore that some scattering present in the data of Sophianopoulos *et al.* may have lead to an intercept and slope for a true monomer–dimer.

It has been shown here that in addition to supplying theoretical equations for the equilibrium constants of self-associating systems, the multinomial theory can be applied to the analysis of experimental data to yield results that are com-

patible with the experimental system being studied and free from serious ambiguity. Some experimental results may be unsuitable for work with the present methods. The confidence to be held in the final results is dependent on the quality of the experimental data: perfect measurements will lead to essentially perfect results and error-containing data will produce approximations, good or poor, depending on the amount of error in the data.

Acknowledgments

The author thanks Dr. Robert A. Spangler (Department of Biophysics, State University of New York at Buffalo) for a valuable discussion of this manuscript, Mr. Brian A. Jordan for invaluable skilled technical assistance with the ultracentrifuge, Mrs. Herriet J. Vender for her efficient work with our computer programs and in the ultracentrifuge, and Miss Arline S. Dash for her valuable assistance in the initial stages of our computer programming.

References

Adams, E. T., Jr. (1967), Fractions, Palo Alto, Calif., No. 3 Beckman Instruments.

Adams, E. T., Jr., and Filmer, D. L. (1966), *Biochemistry 5*, 2971.

Ansevin, H. T., Roark, D. E., and Yphantis, D. A. (1970), *Anal. Biochem. 34*, 237.

Bruzzesi, M. R., Chiancone, E., and Antonini, E. (1965), *4*, 1796.

Canfield, R. (1963), *J. Biol. Chem. 238*, 2698.

Derechin, M. (1968), *Biochemistry 7*, 3253.

Derechin, M. (1969a), *Biochemistry 8*, 927.

Derechin, M. (1969b), *Biochemistry 8*, 921.

Derechin, M. (1969c), *Anal. Biochem. 28*, 385.

Gosting, L. J., and Morris, M. S. (1949), *J. Amer. Chem. Soc. 71*, 1998.

Halwer, M., Nutting, G. C., and Brice, B. A. (1951), *J. Amer. Chem. Soc. 73*, 2786.

Snedecor, G. W. (1959), Statistical Methods, Iowa State College Press.

Sophianopoulos, A. J., and Van Holde, K. E. (1964), *J. Biol. Chem. 239*, 2516.

Teller, D. C., Horbett, T. A., Richards, E. G., and Schachman, H. K. (1969), *Ann. N. Y. Acad. Sci. 164*, 166.

Van Holde, K. E., Rossetti, Q. P., and Dyson, R. D. (1969), *Ann. N. Y. Acad. Sci. 164*, 279.

Yphantis, D. A. (1964), *Biochemistry 3*, 297.

Endonuclease II of *Escherichia coli*. Degradation of Partially Depurinated Deoxyribonucleic Acid*

Sheikh-Mumtaz Hadi and David A. Goldthwait†

ABSTRACT: Endonuclease II of *Escherichia coli*, which was partially purified on the basis of its ability to catalyze single- and double-strand breaks in DNA alkylated with methylmethanesulfonate, has been shown to degrade partially depurinated DNA. DNA either entrapped in an acrylamide gel or free in solution was converted to a substrate by heating at various temperatures at pH 3.5. Depurination and molecular weight studies, a melting curve, viscosity, and density studies indicated that the heating at various temperatures produced either double-stranded or single-stranded partially depurinated DNA. The exact nature of the single- and double-stranded forms is not clear. The depurinated DNA was unstable in alkali, but could be stabilized by treatment with $NaBH_4$ or with NH_2OH. After either treatment, the DNA was still a substrate for the enzyme.

Endonuclease II of *Escherichia coli* hydrolyzes phosphodiester bonds of alkylated DNA. The enzyme makes both single- and double-strand breaks in DNA alkylated with the monofunctional alkylating agent methylmethanesulfonate. It also makes a limited number of single-strand breaks in native T-4 and T-7 DNA (Friedberg and Goldthwait, 1968, 1969; Friedberg *et al.*, 1969). Work in this laboratory, which has characterized this enzyme from *E. coli*, was started with the belief that distortions in DNA structure were responsible for the endonucleolytic scissions required for recombination (Friedberg and Goldthwait, 1968). Alkylation of DNA was considered to be a chemical means of producing distortions which might be analogous to those produced biologically. Previous work by others (Reiter *et al.*, 1967; Strauss and Robbins, 1968) aided in the initiation of our work.

Endonuclease II has been purified some 1600-fold on the basis of its ability to degrade alkylated DNA. Experiments have been designated to determine whether the enzyme will hydrolyze phosphodiester bonds of double-stranded DNA the structure of which is altered by other methods. In this paper, the activity of the enzyme on depurinated and depurinated reduced DNA is described.

Materials and Methods

Enzyme Preparation. The original purification procedure (Friedberg and Goldthwait, 1969) has been modified and the

* From the Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106. *Received June 25, 1971.* This study was supported by grants from the National Institutes of Health (Grant CA-11322), National Science Foundation (GB-1309), and the American Cancer Society (E 537).

† National Institutes of Health Research Career Award Fellowship KG-gm-21444; to whom to address correspondence.