Biophysical Chemistry 7 (1977) 121–139

ANALYSIS OF VARIOUS INDEFINITE SELF-ASSOCIATIONS *

Lih-Heng TANG **, David R. POWELL, Barnee M. ESCOTT and E.T. ADAMS Jr. ***
Chemistry Department, Illinois Institute of Technology, Chicago, Illinois 60616, USA and Texas A&M University, College Station, Texas 77843, USA

Received 14 October 1976
Revised manuscript received 14 March 1977

Two methods have been developed for the analysis of four types of indefinite self-associations. Unlike previous treatments by others, the procedures can be applied to nonideal cases. The two methods were first tested with simulated data, and it was found that one could indeed distinguish between the four types of indefinite self-associations. For a more realistic test, sedimentation equilibrium experiments were performed on solutions of β-lactoglobulin A at 16°C in 0.15 ionic strength acetate buffer, pH 4.65. The self-association of the β-lactoglobulin A was best described by either method as a sequential indefinite self-association having two equilibrium constants and one second virial coefficient.

1. Introduction

Self-associations that appear to continue without limit, such as

$$nP_1 \rightleftharpoons qP_2 + mP_3 + hP_4 + \ldots, \quad (1)$$

$$nP_1 \rightleftharpoons qP_2 + hP_4 + jP_6 + \ldots, \quad (2)$$

and related equilibria involving an associating solute, P, are known as indefinite self-associations [1]. There are many varieties of indefinite self-associations, and we can illustrate some of the problems in studying and analyzing them by considering four different examples, which may be more frequently encountered. While it is potentially possible to develop enough equations to analyze any kind of self-association, the experimental limitations of the instruments (osmometers, ultracentrifuges, light scattering photometers) used to study self-associations force one to make some simplifying assumptions. Sometimes the experimental precision presently available may make it difficult to distinguish models for the observed self-association [2,3].

Indefinite self-associations have been encountered with a variety of different solutions. In aqueous solutions proteins, such as β-lactoglobulin A [4], insulin [5] and bovine lactate dehydrogenase [6], purine [7], N-6-dimethyl-adenine and N-6,9-dimethyladenine [8], and methylene blue [9] have been found to undergo indefinite self-associations. With nonaqueous solutions indefinite self-associations have been reported for polyethylene glycol [10], dodecylammonium propionate [11], phenol and various alcohols [12, 13], and some amides [14].

Kreuzer [15] developed some very elegant methods for analyzing a monomer-n-mer association and a sequential, indefinite self-association having all molar association constants equal (also known as a Type I indefinite or an isodesmic self-association). His methods were restricted to ideal, dilute solutions or ideal gas phase self-associations; he showed how one could evaluate both the number fraction and the weight fraction of monomer from the experimental data. Hoffmann [12] and also Coggeshall and Saier [13] found that their association data could not be analyzed as a Type I indefinite self-association. Instead they were forced to assume that K_{12}, the association constant for the dimerization, was different from the other association constants; the other associa-

* This paper is XI in the series: Sedimentation Equilibrium in Reacting Systems.

** Present address: Montefiore Hospital and Medical Center, 111 East 210th St., Bronx, New York 10467, USA.
*** Please address all correspondence to Dr. E.T. Adams Jr., Chemistry Department, Texas A&M University, College Station, Texas 77843, USA.

tion constants (K_{23}, K_{34}, etc.) were assumed to be equal, i.e., $K_{12} < K_{23} = K_{34} = \ldots = K$. An indefinite self-association of this type will be referred to as an Type III indefinite association. An indefinite self-association involving monomer, dimer, hexamer and aggregates that are integral multiples of hexamer seemed to describe the self-association of insulin in aqueous solution near neutral pH [5].

It appears at present that methods for analyzing nonideal, indefinite self-associations have been restricted to Type I indefinite self-associations. Using four different types of indefinite self-associations, we will show how these four nonideal self-associations can be analyzed. For each type we will show two different methods of analysis. Our procedures will be tested with real and simulated examples. Using data obtained from sedimentation equilibrium experiments at 16°C on a β-lactoglobulin A sample dissolved in an acetate buffer of ionic strength 0.15, we will show that the same type of association is present when either method is used. Also both methods give good agreement for the values of the association constants and second virial coefficients.

2. Quantities needed for the analysis

It will be assumed that the solution containing the associating solute has been dialyzed against the solvent solution containing the supporting electrolyte and/or buffers, so that the associating species can be defined by the Vrij–Overbeek [16,17] or Casassa–Eisenberg [18] conventions. Thus the equilibrium constant or constants, K_{ij}, the second virial coefficient, BM_1, and other physical properties refer to associating species defined according to these conventions.

We make the usual assumptions [1,4,19] regarding the self-associating species: 1) The partial specific volumes, $\bar{v}$, or the density increments, $1000(\partial\rho/\partial c)_\mu$, are equal. 2) The refractive index increments, $(\partial n/\partial c)_{T,P}$, in liters per gram are equal. 3) The natural logarithm of the activity coefficient, y_i, on the gram per liter concentration scale, c, can be represented by

$$\ln y_i = iBM_1c, \quad i = 1, 2, \ldots . \tag{3}$$

Here BM_1 is known as the second virial coefficient. For an ideal, dilute solution $BM_1 = 0$. The validity of eq. (3) has recently been examined by Ogston and Winzor [20]. They found that eq. (3) gives "a reliable estimate of the thermodynamic equilibrium constant over the ranges of stoichiometry and solute concentration that are likely to be encountered in practice". This was true even with uncharged spheres, the most unfavorable case theoretically from the point of view of the approximation.

In sedimentation equilibrium experiments, the apparent weight average molecular weight, M_{wa}, is given by

$$\frac{1}{A}\frac{\mathrm{d}\ln c}{\mathrm{d}(r^2)} = \frac{M_{wc}}{1 + BM_{wc}c} = M_{wa} \tag{4}$$

where

$$A = (1 - \bar{v}\rho)\omega^2/2RT. \tag{5}$$

The quantities making up A have their usual meaning: $\bar{v}$ is the partial specific volume of the associating solute, ρ is the solution density, $\omega = 2\pi(\mathrm{rpm})/60$ is the angular velocity of the rotor, R is the gas constant, and T is the absolute temperature. In eq. (4) c is the concentration in g/l of the associating solute at any radial position, r, in the solution column of the ultracentrifuge cell ($r_m \leqslant r \leqslant r_b$). M_{wa} is related to the true weight average molecular weight, M_{wc}, by the relation

$$M_1/M_{wa} = M_1/M_{wc} + BM_1c. \tag{6}$$

Here M_1 is the monomer molecular weight. Note that M_{wc} is the same as M_{wr}, the weight average molecular weight at any radial position r in the solution column of the ultracentrifuge cell. For self-associations M_{wr} and also M_{wa} are functions of c [1,3,4,19,21,22], so that the symbol M_{wc} is used to indicate this. In order to analyze self-associations it is necessary to do experiments at several different initial concentrations, c_0, calculate M_{wa} at various values of c ($c = c_r$), and piece them together so that one can make a plot of M_{wa} versus c or a plot of M_1/M_{wa} versus c as shown in fig. 1. Experiments are carried out to as low a value of c_0 as possible, and the plot of M_{wa} versus c is extrapolated to M_1 at $c = 0$. From the smooth curve drawn or fitted through the M_{wa} versus c data, a plot of M_1/M_{wa} versus c (see fig. 1) is constructed and used for the evaluation of M_1/M_{na}, since [1,19,22]

$$M_1/M_{na} = \frac{1}{c}\int_0^c \frac{M_1}{M_{wa}}\,\mathrm{d}c = \frac{M_1}{M_{nc}} + \frac{BM_1c}{2}. \tag{7}$$

Here M_{nc} is the number average molecular weight, and M_{na} is its apparent value. When $BM_1 = 0$, $M_{na} = M_{nc}$ and $M_{wa} = M_{wc}$.

It is also possible to evaluate $\ln f_a$, the natural logarithm of the apparent weight fraction of monomer [1,19,22] since

$$\ln f_a = \int_0^c \left(\frac{M_1}{M_{wa}} - 1\right)\frac{dc}{c} = \ln f_1 + BM_1 c. \tag{8}$$

Here $f_1 = c_1/c$ is the weight fraction of monomer. In order to use eq. (8) it is necessary to make a plot of $(M_1/M_{wa} - 1)/c$ versus c and extrapolate this plot to its limiting value at $c = 0$. One notes that

$$\lim_{c\to 0}\left(\frac{M_1}{M_{wa}} - 1\right)/c = \begin{cases} -K_2 + BM_1, & \text{if dimer is present} \\ BM_1, & \text{if dimer is absent} \end{cases}$$

With some self-associations there are problems arising in the vicinity of $c = 0$ with the plot required for eq. (8). First of all, neither $-K_2 + BM_1$ nor BM_1 are known a priori. With a large K_2 the plot becomes quite steep in the vicinity of $c = 0$, and the extrapolation becomes very difficult to do. Secondly, with a monomer-dimer-trimer association having $K_3 \gg K_2$, the plot $(M_1/M_{wa} - 1)/c$ will go through a minimum before reaching the intercept [23]. Thirdly, if no dimer is present and $BM_1 > 0$, then this plot will go through a maximum near $c = 0$. In the fourth place, the greatest contribution to the area under the curve usually comes from the region of lowest concentration, where the experimental error is greatest and where there are very few reliable data points to guide the extrapolation of $(M_1/M_{wa} - 1)/c$ versus c. These problems can be overcome by using the quantity $\ln(f_a/f_{a*})$, which is defined by [11,24]

$$\ln(f_a/f_{a*}) = \int_{c_*}^{c}\left(\frac{M_1}{M_{wa}} - 1\right)\frac{dc}{c}$$

$$= \ln(f_1/f_{1*}) + BM_1(c - c_*) \ . \tag{9}$$

Here c_* is the lowest concentration used; its choice is somewhat arbitrary.

Eqs. (6) and (7) can be combined so that BM_1 is eliminated; thus [1,11,22,24–27].

$$\xi = \frac{2M_1}{M_{na}} - \frac{M_1}{M_{wa}} = \frac{2M_1}{M_{nc}} - \frac{M_1}{M_{wc}}. \tag{10}$$

We will show how eqs. (9) or (10) can be used for the analysis of four types of indefinite self-associations.

3. Analysis of indefinite self-associations

Indefinite self-associations are associations described by eqs. (1) and (2) or by related self-associations. These associations can be considered to be made up of a series of simultaneous equilibria [1,7, 10,22]:

$$\begin{aligned} &P_1 + P_1 \rightleftharpoons P_2 \\ &P_2 + P_1 \rightleftharpoons P_3 \\ &\vdots \quad \ddots \quad \ddots \\ &P_{(n-1)} + P_1 \rightleftharpoons P_n \end{aligned} \tag{11}$$

Whenever eq. (3) is valid, the following relations obtain between the molar concentrations or the associating species:

$$K_{12} = [P_2]/[P_1]^2 \qquad \text{or} \qquad [P_2] = K_{12}[P_1]^3$$

$$K_{23} = [P_3]/[P_1][P_2]$$

$$\text{or} \quad [P_3] = K_{23}[P_1][P_2] = K_{12}K_{23}[P_1]^3$$

$$K_{(n-1)n} = [P_n]/[P_{(n-1)}][P_1]$$

$$\text{or} \quad [P_n] = K_{(n-1)n}[P_1][P_{(n-1)}]$$

$$= K_{12}K_{13}\cdots K_{(n-1)n}[P_1]^n. \tag{12}$$

Here we shall show how four types of indefinite self-association can be analyzed.

3.1. Type I indefinite self-association

This type of indefinite self-association is also known as a random, open or isodesmic self-associations. For this association it is assumed that all molar association constants are equal, i.e., $K_{12} = K_{23} = \ldots = K$. It has been shown that when eq. (3) applies, C (in g/ml) becomes [1,4,7,11,19,22].

$$C = C_1 + 2kC_1^2 + 3k^2C_1^3 + 4k^3C_1^4 + \ldots$$

$$= C_1/(1 - kC_1)^2, \quad \text{if } kC_1 < 1. \tag{13}$$

Here k, the intrinsic equilibrium constant, is defined by

$$k = 1000K/M_1. \tag{14}$$

Furthermore, when $kC_1 < 1$, one obtains

$$\frac{M_1}{M_{wa}} = \frac{1 - kC_1}{1 + kC_1} + \hat{B}M_1 C = \frac{\sqrt{f_1}}{2 - \sqrt{f_1}} + \hat{B}M_1 C \tag{15}$$

and

$$\frac{M_1}{M_{na}} = 1 - kC_1 + \frac{\hat{B}M_1 C}{2} = \sqrt{f_1} + \frac{\hat{B}M_1 C}{2}. \tag{16}$$

Note that $\hat{B}M_1 C = BM_1 c$. Since $c = 1000C$, $\hat{B}M_1 = 1000\,BM_1$. The quantity ξ (see eq. 10) becomes [1,11, 22,24–27]

$$\xi = \frac{2M_1}{M_{na}} - \frac{M_1}{M_{wa}} = 2\sqrt{f_1} - \frac{\sqrt{f_1}}{2 - \sqrt{f_1}} \tag{17}$$

and

$$\sqrt{f_1} = (1/4)\{(\xi + 3) - \sqrt{(\xi + 3)^2 - 16\xi}\}. \tag{18}$$

From eq. (13) one notes that

$$f_1 = C_1/C = (1 - kC_1)^2, \quad \sqrt{f_1} = 1 - kC_1 = 1 - kCf_1$$

and

$$1 - \sqrt{f_1} = kCf_1, \quad \text{or} \quad (1 - \sqrt{f_1})/f_1 = kC. \tag{19a,b}$$

With the aid of eq. (18) one obtains $\sqrt{f_1}$; then eqs. (19a or 19b) are used to evaluate k. A plot of $(1 - \sqrt{f_1})/f_1$ versus C will have a slope of k. If the model is correct, then this plot should be a straight line going through or close to the origin. Experimental error may cause the plot not to go through the origin. Once $\sqrt{f_1}$ is known, then $\hat{B}M_1$ can be obtained from eq. (15). It is also possible to use eq. (9) to obtain $\hat{B}M_1$, since

$$\ln(f_a/f_{a^*}) = \ln\left\{\left[\frac{2(M_1/M_{wa} - \hat{B}M_1 C)}{1 + (M_1/M_{wa} - \hat{B}M_1 C)}\right]^2 \Big/ \left[\frac{2(M_1/M_{wa^*} - \hat{B}M_1 C_*)}{1 + (M_1/M_{wa^*} - \hat{B}M_1 C_*)}\right]^2\right\} + BM_1(C - C_*) \tag{20a}$$

and

$$\sqrt{f_1} = (1 - kC_1) = \frac{2(M_1/M_{wc})}{1 + M_1/M_{wc}} = \frac{M_1}{M_{nc}} = \frac{2(M_1/M_{wa} - \hat{B}M_1 C)}{1 + M_1/M_{wa} - \hat{B}M_1 C}. \tag{20b}$$

Here one sets up an array of $\ln(f_a/f_{a^*})$ values at various values of c and find the best value of $\hat{B}M_1$ using a previously described method (11). The first step is to choose values of $\hat{B}M_1$ that one feels will include anticipated values; then we look for changes in the sign of the function

$$\epsilon_i = [\ln(f_a/f_{a^*})_{OBSVD} - \ln(f_a/f_{a^*})_{CALCD}]_i.$$

The values of $\hat{B}M_1$ giving sign changes in ϵ_i are examined in more detail. For each choise of $\hat{B}M_1$, the values of f_1, are calculated according to eq. (20b). Each set of f_1's is used to generate a k value from eqs. (19a or 19b). The values of k and $\hat{B}M_1$ from each set of f_1's are used to generate values of $(M_1/M_{wa})_{CALCD}$ as a function of ϵ. Sometimes false solutions are encountered (negative k values or negative M_1/M_{wa} values). These are discarded. Then we check the remaining set of k and $\hat{B}M_1$ values to see which set gives the best fit. Our criterion for a best fit is that $\Sigma(\delta_i)^2$ is a minimum; here

$$\delta_i = [(M_1/M_{wa})_{CALCD} - (M_1/M_{wa})_{OBSVD}]_i. \tag{21}$$

Similar procedures are used with the appropriate equation for $\ln(f_a/f_{a^*})$ in the analysis of other types of indefinite self-associations. One $\hat{B}M_1$ is known, f_1 is also known, and the other available relations can be used to evaluate the equilibrium constant or constants.

3.2. Type II indefinite self-association

This association is described by eq. (2); note that all odd species beyond monomer are absent. The series of simultaneous equilibria describing this association are

$$P_1 + P_1 \rightleftharpoons P_2 \qquad [P_2] = K_{12}[P_1]^2$$

$$P_2 + P_2 \rightleftharpoons P_4 \qquad [P_4] = K_{24}[P_2]^2 = K_{12}^2 K_{24}[P_1]^4$$

$$P_2 + P_4 \rightleftharpoons P_6$$

$$[P_6] = K_{46}[P_2][P_4] = K_{12}^3 K_{24} K_{26}[P_1]^6.$$

When all molar association constants are equal ($K_{12} = K_{24} = K_{46} = \ldots = K$), then the total solute concentration, C (in g/ml), becomes [1,11]

$$C = C_1 + 2kC_1^2 + 4k^3C_1^4 + 6k^5C_1^6 + \ldots = C_1\left\{1 + \frac{2kC_1}{(1 - k^2C_1^2)^2}\right\} = C_1\left\{1 + \frac{2x}{(1 - x^2)^2}\right\}, \tag{22}$$

if $x \equiv kC_1 < 1$, $k = 1000K/M_1$.

This equation for ξ becomes

$$\xi = \frac{2M_1}{M_{na}} - \frac{M_1}{M_{wa}} = \frac{2[1+x(1-x^2)]}{1+2x/(1-x^2)^2} - \frac{1+2x/(1-x^2)^2}{1+4x(1+x^2)/(1-x^2)^3}. \quad (23)$$

Note that ξ is a function of one unknown, $x = kC_1$. Remembering that $0 < kC_1 < 1$, we can obtain values of $kC_1 = x$ for each ξ by successive approximations. These values of kC_1 can be inserted into eq. (22) to obtain C_1 or f_1, and one can use C_1 (or f_1) to obtain k, since

$$x = kC_1 = kCf_1. \quad (24)$$

Thus a plot of x versus C_1 or x/C versus f_1 should give a straight line going through or close to the origin if a type II association is present; the slope of either plot is k. Values of $\hat{B}M_1$ can be obtained from a modification of eq. (6), since

$$\frac{M_1}{M_{wa}} - \frac{1+2x/(1-x^2)^2}{1+4x(1+x^2)/(1-x^2)^3} = \hat{B}M_1 C. \quad (25)$$

When eq. (9) for $\ln(f_a/f_{a*})$ is used, one notes that f_1 is given by

$$2f_1 = \frac{3M_1}{M_{nc}} + \frac{M_{wc}}{M_1} - 2 - \left[\left(2 - \frac{M_{wc}}{M_1} - \frac{3M_1}{M_{nc}}\right)^2 - 4\left(\frac{M_{wc}}{M_{nc}} + \frac{2M_1}{M_{nc}} - 2\right)\right]^{1/2}. \quad (26)$$

Here

$$M_1/M_{nc} = M_1/M_{na} - \hat{B}M_1 C/2 \quad (27)$$

$$M_1/M_{wc} = M_1/M_{wa} - \hat{B}M_1 C \quad (28)$$

and

$$M_{wc}/M_1 = (M_1/M_{wc})^{-1}. \quad (29)$$

The equations for f_1 and f_{1*} are given by eq. (26), evaluated at C and C_*, respectively. These equations are used in eq. (9), and one solves for $\hat{B}M_1$ by the method described previously with the Type I indefinite self-association. Once f_1 and $\hat{B}M_1$ are known, then it is a simple matter to obtain k, M_{wc} and M_{nc}. The derivation of eq. (26) is shown in Appendix A.

3.3. Type III indefinite self-association

This is a variant of the type I indefinite self-association. Here it is assumed [see eqs. (12)] that $K_{12} \neq K_{23}$, K_{24}, etc., but it is assumed that $K_{23} = K_{34} = \ldots = K$. The total concentration (in g/ml) of the associating solute becomes [1,26].

$$\begin{aligned} C &= C_1 + 2k_{12}C_1^2 + 3k_{12}kC_1^3 + 4k_{12}k^2C_1^4 + \ldots \\ &= C_1[1 + k_{12}C_1(2 + 3kC_1 + 4k^2C_1^2 + \ldots)] \\ &= C_1[1 + k_{12}C_1\{(2 - kC_1)/(1 - kC_1)^2\}] \\ &= C_1[1 + y\{(2-x)/(1-x)^2\}], \quad \text{if } kC_1 < 1. \end{aligned} \quad (30)$$

Here

$$k_{12} = 1000K_{12}/M_1 \quad \text{and} \quad y = k_{12}C_1 \quad (31)$$

$$k = 1000K/M_1 \quad \text{and} \quad x = kC_1 \quad (32)$$

The pertinent equation for ξ becomes [26]

$$\xi = \frac{2 + 2[y/(1-x)]}{1 + y(2-x)/(1-x)^2} - \frac{1 + y[(2-x)/(1-x)^2]}{1 + y[(4-3x+x^2)/(1-x)^3]} \quad (33)$$

Note that this is an equation in two unknowns, $y = K_{12}C_1$ and $x = kC_1$. Nonetheless, it can be used in an iterative method to obtain k_{12} and k. This is basically how the iterative method works [26]. First estimate a value of k_{12}. Multiply eq. (30) by k_{12} to obtain

$$\begin{aligned} k_{12}C &= k_{12}C_1[1 + y(2-x)/(1-x)^2] \\ &= y + y^2(2-x)/(1-x)^2. \end{aligned} \quad (34)$$

This can be rearranged to give a quadratic equation in y, namely,

$$y^2 + y\frac{(1-x)^2}{(2-x)} - k_{12}C\frac{(1-x)^2}{(2-x)} = 0. \quad (35)$$

Solve eq. (35) for y, noting that the proper root to use is the one containing the positive square root of the discriminant. This root, $g(x)$, is inserted into eq. (33) for ξ. Thus

$$\xi = \frac{2[(1-x)^2 + (1-x)g(x)]}{(1-x)^2 + (2-x)g(x)} - \frac{(1-x)^3 + (2-x)(1-x)g(x)}{(1-x)^3 + (4-3x+x^2)g(x)} \quad (36)$$

Now solve for x using a bisection technique; remember that $0 < x < 1$. If at this point the values of x required to solve eq. (36) exceed 1, chooose another value of k_{12} (say one half of the original estimate) and repeat the process. When apparent solutions of eq. (36) are found for which $0 < x < 1$, then calculate $\Delta\xi$, where

$$\Delta\xi = \xi_{OBSVD} - \xi_{CALCD}. \quad (37)$$

Pick the value of x which gives

$$|\Delta\xi| < 1 \times 10^{-6}.$$

This is near the limit of single precision accuracy on our computer. Repeat this process at various other choices of C. Once x is known at each C, then y is also known, since it is the root, $g(x)$, of eq. (36). Since k_{12} is known now, then k can be readily obtained from the following relations:

$$f_1 = C_1/C = \frac{1}{1 + y(2-x)/(1-x)^2} \quad (38)$$

and

$$k = x/C_1 = x/Cf_1. \quad (39)$$

These procedures give pointwise values of k_{12} and k. As a final test one can check these values for consistency over the range of C that was used. In addition one can regenerate values of M_1/M_{wa} and compare it with the observed values of M_1/M_{wa}. The best choice for k_{12} and k will be the one that gives

$$\sum_i (\delta_i)^2 = \text{Minimum},$$

where δ_i is defined by eq. (21).

The other method, involving one unknown, is to use eq. (9). For the Type III indefinite self-association

$$f_1 = \left[\frac{M_1}{M_{nc}}\left(\frac{2M_1}{M_{wc}} - \frac{M_{wc}}{M_1} - 3\right) + 2\right]\left[1 - \frac{M_{wc}}{M_1}\right]^{-1}. \quad (40)$$

An outline of the derivation of eq. (40) is given in Appendix B. Note that the quantities (M_1/M_{nc}), (M_1/M_{wc}) and (M_{wc}/M_1) are defined by eqs. (27)–(29), respectively. The values of f_1 and f_{1*} are calculated from eq. (40) and inserted into eq. (9), and the evaluation of $\hat{B}M_1$ is done by the methods described for the Type I indefinite self-association.

3.4. Type IV indefinite self-association

This association is a variant of the Type II indefinite self-association. Here $K_{12} \neq K_{24}$, K_{26}, etc., but $K_{24} = K_{26} = \ldots K$. The appropriate equations for C and ξ are [1,26]

$$C = C_1 + 2k_{12}C_1^2 + 4k_{12}\,2kC_1^4 + 6k_{12}{}^3k^2C_1^6 + \ldots = C_1\left[1 + \frac{2k_{12}C_1}{(1-k_{12}kC_1^2)^2}\right] = C_1\left[1 + \frac{2k_{12}C_1}{(1-k_*^2C_1^2)^2}\right]. \quad (41)$$

Here $k_{12} = 1000K_{12}/M_1$, $k = 1000K/M_1$ and $k_*^2 = k_{12}k$.

$$\xi = \frac{2M_1}{M_{nc}} - \frac{M_1}{M_{wc}} = \frac{2[1 + k_{12}C_1/(1-k_{12}kC_1^2)]}{1 + 2k_{12}C_1/(1-k_{12}kC_1^2)^2} - \frac{1 + 2k_{12}C_1/(1-k_{12}kC_1^2)]}{1 + 4k_{12}C_1(1 + k_{12}kC_1^2)/(1-k_{12}kC_1^2)^3} \quad (42)$$

Eqs. (41) and (42) are valid only if $k_{12}C_1 < 1$ and $kC_1 < 1$. One can also write eq. (42) with k_*^2 replacing the product kk_{12}. Note that eq. (42) has two unknowns, $x = k_{12}C_1$ and $y = kC_1$. We can solve for k_{12} and k using an iterative method similar to that used with a Type III indefinite self-association. Alternatively, one can use eq. (9) for $\ln(f_a/f_{a*})$, noting that eqs. (27)–(29) apply and that $2f_1$ is also defined by eq. (26), and one would solve for f_1 and $\hat{B}M_1$ as described previously. Some details about why eq. (26) is also valid here and how one would distinguish between the Types II and IV indefinite self-associations are given in Appendix C.

4. Tests with a simulated example

Having shown how one might analyze the four types of indefinite self-associations, the next step was to test our methods with a simulated example. Here we

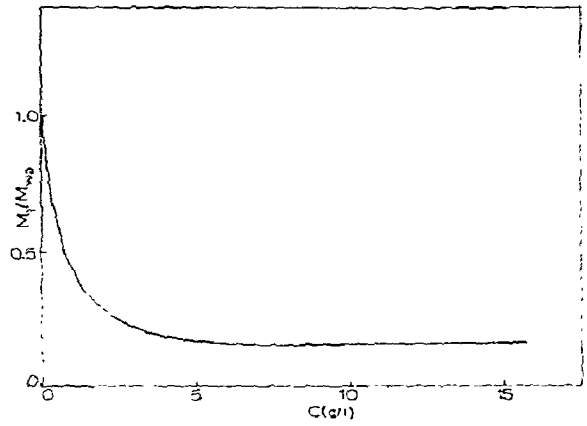

Fig. 1. Simulated example. Plot of M_1/M_{wa} versus C for a Type IV indefinite self-association having k_{12} = 709 ml/g, k = 1.159 × 10^3 ml/g and $\hat{B}M_1$ = 6.5 ml/g.

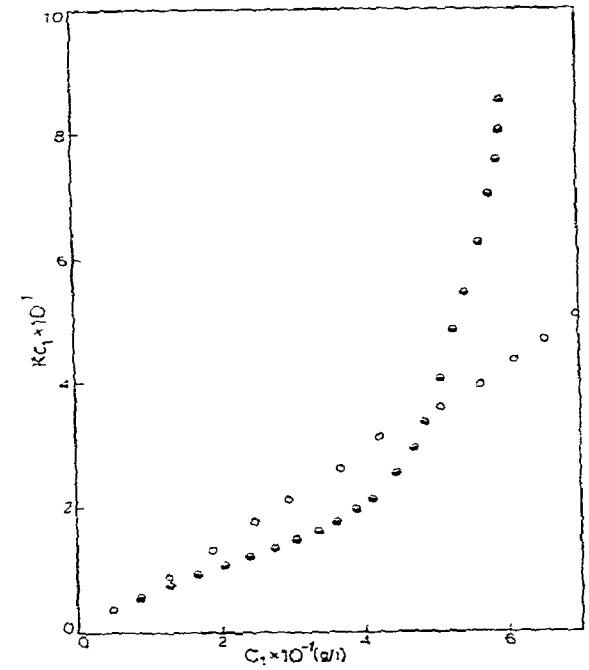

Fig. 3. Simulated example. Test for a Type II indefinite self-association [see eqs. (23) and (24)]. Plot based on eq. (24) using Method I (⊖); The failure of this plot to give a straight line going through or close to the origin argues against a Type II association being present. Plot based on eq. (24) using Method II (○). Even though this plot gave the desired straight line, the regenerated plot of M_1/M_{wa} versus C would not describe the original data. This is reflected in the value of $\Sigma_i(\delta_i)^2$ in table 1.

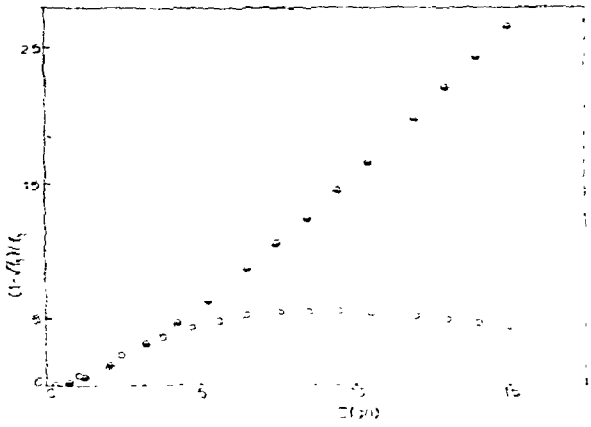

Fig. 2. Simulated example. Test for a Type I indefinite self-association [see eqs. (18) and (19b)]. The failure of this plot to give a straight line going through or close to the origin rules out a Type I self-association.

used a Type IV indefinite self-association having the following parameters: k_{12} = 709 ml/g, k = 1.159 × 10^3 ml/g and $\hat{B}M_1$ = 6.5 ml/g. Fig. 1 shows a plot of M_1/M_{wa} versus C for this example. For the nonideal association there is a broad minimum in the plot of M_1/M_{wa} versus C over the concentration range displayed, since $\hat{B}M_1$ is small. The values of M_1/M_{wa} at various values of C were used to calculate the corresponding values of M_1/M_{na} and $\ln(f_a/f_{a^*})$ [see eqs. (7) and (9)]. The question to be answered is: Can one really distinguish between the various types of indefinite self-associations? Our results with this simulated example and the real example that follows indicate that it may be possible to distinguish between the four kinds of indefinite self-associations. The results or our tests with the simulated example are shown in table 1. The first part of this table is labelled Method I. Here are displayed the results of the analysis based on the quantity ξ [see eqs. (17), (23), (33) and (36)] that are appropriate for each of the four types of indefinite self-associations. It is evident from this table that the Type IV indefinite self-association is the best choice. Types I and II can be eliminated by inspection of the plots shown in figs. 2 and 3. If either model were correct, then one of the plots shown in figs. 2 or 3 would give a straight line going through or close to the origin. The curvature in these plots impeach these models. The Type III in-

Table 1
Analysis of simulated data using a Type IV model system

Type	No. of pts. [a]	BM_1 [b] (l/g)	R [c]	$\mp\sigma$ [d]	k_{12} [e] (l/g)	k [f] (l/g)	R	$\pm\sigma$	$\Sigma\lvert\delta\rvert$ [g]	$\Sigma(\delta)^2$	Comments
A. Method I [k]											
I	30	0.015 [h]	0.97084	0.035	N/A	1.85	0.99657	0.39	4.621	1.352	Rejected by inspection of fig. 2
II	30	−0.013 [i]	0.9546	0.022	N/A	1.18	0.98713	0.17	2.429	3.906×10^{-1}	Rejected on the basis of fig. 3
III	30	−0.035 [i]	0.87622	0.051	0.709	0.96	0.92567	0.30	4.725	1.592	Rejected on the basis of $\Sigma(\delta)^2$
IV	30	6.5×10^{-3} [i]	0.99997	4.1×10^{-5}	0.709	1.160	0.99993	1.7×10^{-3}	3.25×10^{-3}	3.30×10^{-7}	Gives the best fit.

Type	No. of pts.	BM_1 [j] (l/g)	k_{12} (l/g)	R	k (l/g)	R	$\Sigma\lvert\delta\rvert$	$\Sigma(\delta)^2$	Comments
B. Method II [l]									
I	50	−0.0218	N/A	N/A	0.755 [i]	N/A	3.315	5.597×10^{-1}	Rejected by inspection of fig. 5
II	50	6.5×10^{-3}	N/A	N/A	0.710	1.000	4.096	4.836×10^{-1}	Rejected on the basis of $\Sigma(\delta)^2$ and by the inspection of fig. 4
III	50	-2.8×10^{-3}	0.949	0.752	0.590	0.9995	1.307	5.956×10^{-2}	Rejected on the basis of $\Sigma(\delta)^2$
IV	50	6.50×10^{-3}	0.713	1.000	1.16	1.000	6.395×10^{-3}	2.138×10^{-6}	Gives the best fit.

a) No. of points used in the evaluation of BM_1, k_{12} or k.
b) Calculated from the appropriate form of eq. (6) for Method I and appropriate form of eq. (9) for Method II.
c) The coefficient of correlation of the linear regression line.
d) Standard deviation.
e) The assumed value of k_{12} that gave the best fit using Method I.
f) See text for the appropriate equation.
g) δ is defined by eq. (21).
h) This value was obtained from linear regression analysis.
i) Average pointwise value.
j) These values of BM_1 were obtained through use of a bisection technique. The actual bounds on each BM_1 vary.
k) Analysis based on the quantity ξ [see eq. (10)].
l) Analysis based on the quantity $\ln f_a/f_a^*$ [see eq. (9)].

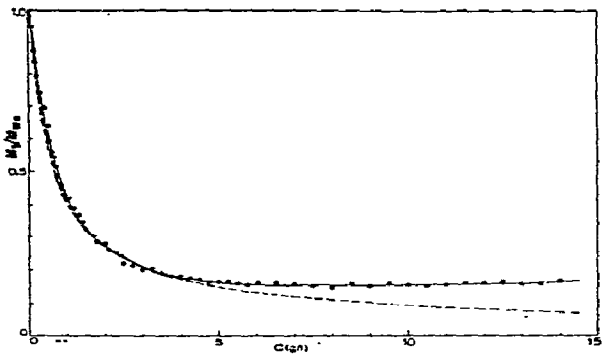

Fig. 4. Simulated example with random error. Here it was assumed that the maximum error in M_1/M_{wa} for the association shown in fig. 1 was ±3%. This error was input randomly, and the noisy data was smoothed using B-splines. A Type IV indefinite self-association having k_{12} = 0.709 l/g, k = 1.196 l/g and BM_1 = 0.0067 l/g seemed to give a very good description of the data; the regenerated M_1/M_{wa} versus c curve using these values is shown by the solid curve. No other type of indefinite self-association seemed to describe the data. The dashed line plot in fig. 4 shows the regenerated curve of M_1/M_{wa} versus c for the spurious Type II self-association having k = 0.710 l/g and BM_1 = 0.0065 l/g (see table 1, Method II). It is evident from this plot that a Type II association does not describe the data. However, if the analysis was based on data collected at very low concentrations (4 g/l and lower), it would be virtually impossible to distinguish between the Type II and the Type IV indefinite self-association.

definite self-association gives a much poorer fit to the data, as judged by the values of $\Sigma_i(\delta_i)^2$, than does a Type IV association. The quantity δ_i is defined by eq. (21).

The second half of table 1, labelled Method II, lists the results obtained from the use of $\ln(f_a/f_{a^*})$ at various values of C; the appropriate equations for f_1 for the four types of indefinite self-associations are given by eqs. (20b), (26), (38) and (26). It is evident from the second part of table 1 that the Type IV indefinite self-association is the best choice, since it has the lowest value of $\Sigma_i(\delta_i)^2$. It should be noted that the choice of C_* or J_* can influence the goodness of fit shown in the table; best results were obtained with a value of C_* as low as possible for the simulated example.

With regard to the plots shown in fig. 3, it should be noted that we rejected a Type II indefinite self-association for several reasons. First, the analysis based on Method I failed to give a straight line as required by the theory; this is evident from the half-filled circle plot. Second, although the analysis based on Method II indicated the possibility of a Type II indefinite self-association being present (see the open circle plot in fig. 3), a regenerated curve of M_1/M_{wa} versus C (using the K and $\bar{B}M_1$ from part B of table 1 for a Type II association) gave very poor fit with the true curve as judged by the values of $\Sigma(\delta_i)^2$ recorded in the next to the last column in part B of Table 1. The best value of $\Sigma(\delta_i)^2$ for Method I or Method II was obtained with a Type IV indefinite self-association. Third, if a Type II model were correct, then the plots in fig. 3 based on Methods I and II should be consistent; they aren't. Finally, the most critical test is how does the regenerated M_1/M_{wa} versus C curve agree with the original M_1/M_{wa} versus C curve. To illustrate this, we have included in fig. 4 the regenerated curve of M_1/M_{wa} versus C (the dashed curve) for this spurious (Type II) solution, and it is evident that it is a poor choice. The regenerated plot of M_1/M_{wa} versus C for the Type IV indefinite self-association, using the k and $\bar{B}M_1$ in part A of table 1, gave a plot indistinguishable from the one shown in fig. 1.

The simulated example was also analyzed by introducing random error; this would give an insight into problems that might arise with the analysis of real data. The maximum possible error in any M_1/M_{wa} values was chosen to be ±3%; however, the error was input randomly. These noisy data were subsequently smoothed using a B-spline smoothing procedure developed by Snodgrass and Smith (personal communication) of the Texas A&M University Department of Mathematics. An analysis of the smoothed data was done for the four types of indefinite self-associations using Method I only. Again, only the Type IV indefinite self-association gave the best description of the data as judged by the values of $\Sigma(\delta_i)^2$ and from regeneration of the M_1/M_{wa} versus C curve. Fig. 4 shows the plot of the noisy M_1/M_{wa} versus C data. The solid curve through these data were obtained from a Type IV association having k_{12} = 709 ml/g, $k = 1.173 \times 10^3$ ml/g and $\bar{B}M_1 = 7.18 \times 10^{-3}$ ml/g. The true values were k_{12} = 709 ml/g, $k = 1.159 \times 10^3$ ml/g and $\bar{B}M_1 = 6.50 \times 10^{-3}$ ml/g. The values of $\Sigma(\delta_i)^2$ obtained for all four indefinite self-associations for noisy, simulated data are listed in table 2. It is evident from this table, as well as from fig. 4, that a Type IV indefinite self-association describes the plot of M_1/M_{wa} versus C best.

Table 2
Analysis of simulated data containing ±3% random error assuming a type IV model system. *Method I* (†)

Type (*)	No. a) of Pts.	k_{12} e) (l/g)	k f) (l/g)	R c)	BM_1 b) (l/g)	R	$\Sigma\lvert\delta\rvert$ g)	$\Sigma(\delta)^2$	Comments
I	50	–	1.975	0.9975	0.1399	0.9735	–	–	See below
II	50	–	4.689	0.8133	0.0105	0.9692	2.565	0.410	Rejected on the basis of graphs like fig. 3 and $\Sigma(\delta)^2$
III	50	0.709	9.309	0.8376	0.0145	0.8954	4.555	1.191	Rejected on the basis of $\Sigma(\delta)^2$
IV	50	0.709	1.196	0.9915	0.0067	0.9817	0.119	4.85×10^{-4}	Gives the best fit
Type (+)	**No. of Pts.**	**k_{12} (l/g)**	**k (l/g)**	**$\pm\sigma$ d)**	**BM_1 (l/g)**	**$\pm\sigma$**	**$\Sigma\lvert\delta\rvert$**	**$\Sigma(\delta)^2$**	**Comments**
I	50	–	1.39	0.35	−0.009	0.024	4.103	0.5272	Rejected on the basis of $\Sigma(\delta)^2$
II	50	–	1.33	0.15	−0.001	0.014	2.104	0.1066	See above
III	50	0.709	1.17	0.22	−0.008	0.034	4.157	0.5270	Rejected on the basis of $\Sigma(\delta)^2$
IV	50	0.709	1.18	0.01	0.0067	0.0024	0.1068	5.89×10^{-4}	Gives the best fit.

† See table 1 for footnotes a, b, c, d, e, f, g.
* These values were obtained using regression analysis.
+ These values were obtained using pointwise averages.

5. Results with β-lactoglobulin A in acetate buffer

The critical test of our method for analyzing indefinite self-associations is how well it works with a real case; here we show some results with the self-association of β-lactoglobulin A in an acetate buffer. The β-lactoglobulin A used in this study was graciously furnished to us by Dr. Edwin J. Kalan and Dr. J.J. Basch of the Eastern Utilization Research and Development Division of the U.S. Department of Agriculture. For this study the buffer solution consisted of 0.1 M sodium acetate, 0.1 M acetic acid and 0.05 M KCl; at 23°C it had a pH of 4.65 and an ionic strength of 0.15. The protein solution was prepared and dialyzed in the same manner as described previously [4,27]. Protein concentrations were determined at 20°C on a Brice–Phoenix differential refractometer using the green line of mercury (λ = 546 nm). The refractive index increment was assumed to be 1.82×10^{-3} dl/g [4,27,28]. At 20°C the initial concentration, J_0, in 12 mm fringes is related to c_0 in g/dl by

$$J_0 = (h/\lambda)(\partial n/\partial c)_{T,P} c_0 = 40 c_0. \qquad (43)$$

The J_0 at 20°C was corrected to that at 16°C by previously used methods [4,27]. At 20°C the partial specific volume, $\bar{v}$, of the β-lactoglobulin A was assumed to be $\bar{v} = 0.751$ ml/g; the value of $\bar{v}$ at 16°C was estimated by the methods of Bull and Breese [29]. At 16°C the density of the buffer was $\rho_0 = 1.0060$ g/ml. Sedimentation equilibrium experiments were carried out in a Beckman/Spinco Model E analytical Ultracentrifuge at 16°C; the speeds used were 9000 and 11 000 RPM. Sedimentation equilibrium experiments were performed using a 12 mm multichannel, equilibrium centerpiece; we used the same experimental procedures and precautions described previously [4,27].

Six different solutions, ranging from $J_0 = 4.04$ to 37.69 fringes (12 mm centerpieces), were used in these experiments. Values of M_{wa} were calculated according to eq. (4). We used a value of $M_1 = 18\,422$ daltons, which was based on the amino acid sequence analysis of Frank and Braunitzer [30]. Two methods were used to put smooth curves through plots of the M_1/M_{wa} versus J data: fig. 5a shows a smoothed curve obtained using ship curves, whereas fig. 5b shows the curve obtained with B-splines [11]. The dashed line in each plot indicates the extrapolation to

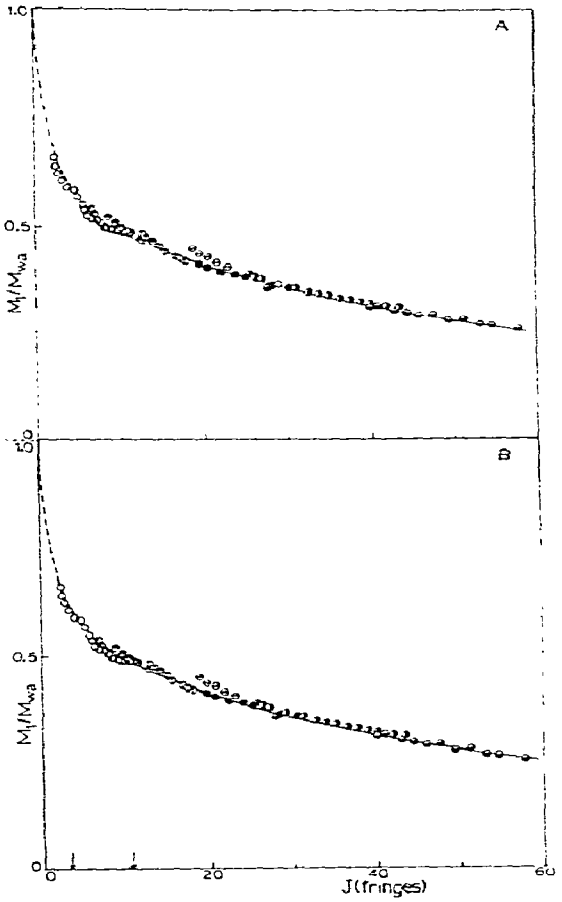

Fig. 5. β-Lactoglobulin A. Plot of M_1/M_{wa} versus J for β-lactoglobulin A at 16°C in 0.15 ionic strength acetate buffer, pH 4.65 (A). The raw data and a ship's curve smoothed plot through the data are displayed here. (B). The raw data with a B-spline smoothed plot through the raw data are displayed here; the arrows indicate the location of the splines or knots. Note that both methods give essentially the same result. The different symbols represent the values of M_{wa} versus J encountered with solutions of different initial concentrations.

$M_1/M_{wa} = 1$ at $J = 0$. Note that both methods give essentially the same result.

Values of M_1/M_{na} and $\ln(f_a/f_{a^*})$ were calculated by means of eqs. (6) and (9), respectively. The experimental values of M_{wa} ranged from $27.9_6 \times 10^3$ daltons at $J = 2.49$ fringes to $71.9_5 \times 10^3$ daltons at $J = 57.98$ fringes. Examination of fig. 5 indicates that

Table 3
Analysis of the self-association of β-lactoglobulin A in 0.15 ionic strength acetate buffer pH 4.65 at 16°C

Type	No. [a] of Pts.	BM_1 [b] (l/g)	R [c]	σ [d]	k_{12} [e] (l/g)	k [f] (l/g)	R	σ	$\Sigma(\delta)^2$	Comments
I	77	-4.26×10^{-4}	0.0545	N/A	N/A	0.238	0.9164	N/A	9.358×10^{-2}	*Rejected by inspection of fig. 5*
II	77	-6.16×10^{-3}	0.4918	N/A	N/A	0.375	0.9369	N/A	7.943×10^{-2}	Rejected by inspection of fig. 6
III	53	-3.59×10^{-3}	N/A	$\pm 9.4 \times 10^{-4}$	0.8125	0.280	N/A	$\pm 6.3 \times 10^{-3}$	9.404×10^{-4}	Gave the best fit based on $\Sigma(\delta)^2$
IV	53	9.12×10^{-3}	N/A	$\pm 7.5 \times 10^{-3}$	0.9918	0.465	N/A	$\pm 6.7 \times 10^{-2}$	4.172×10^{-2}	Rejected on basis of $\Sigma(\delta)^2$

Type	No. of Pts.	BM_1 [j] (l/g)	k_{12} (l/g)	σ	k (l/g)	R	σ	$\Sigma(\delta)^2$	Comments
I	71	1.43×10^{-2}	N/A	N/A	0.536	0.9866	N/A	1.128×10^{-1}	Rejected on basis of $\Sigma(\delta)^2$; values of k and BM_1 would not regenerate the data well
II	71	3.71×10^{-2}	N/A	N/A	0.571	0.7747	N/A	3.600×10^{-1}	Rejected by inspection of fig. 6
III	71	-5.26×10^{-4}	0.634	±0.12	0.315	N/A	±0.035	3.013×10^{-3}	Gives the best fit based on $\Sigma(\delta)^2$
IV	71	3.71×10^{-2}	0.992	±0.28	0.464	N/A	±0.11	1.763×10^{-2}	Rejected on basis of $\Sigma(\delta)^2$

a) No. of points used in the evaluation of BM_1, k_{12} or k.
b) Calculated from the appropriate form of eq. (6) for Method I and appropriate form of eq. (9) for Method II.
c) The coefficient of correlation of the linear regression line.
d) Standard deviation.
e) The assumed value of k_{12} that gave the best fit using Method I.
f) See text for the appropriate equation.
j) These values of BM_1 were obtained through use of a bisection technique. The actual bounds on each BM_1 vary.

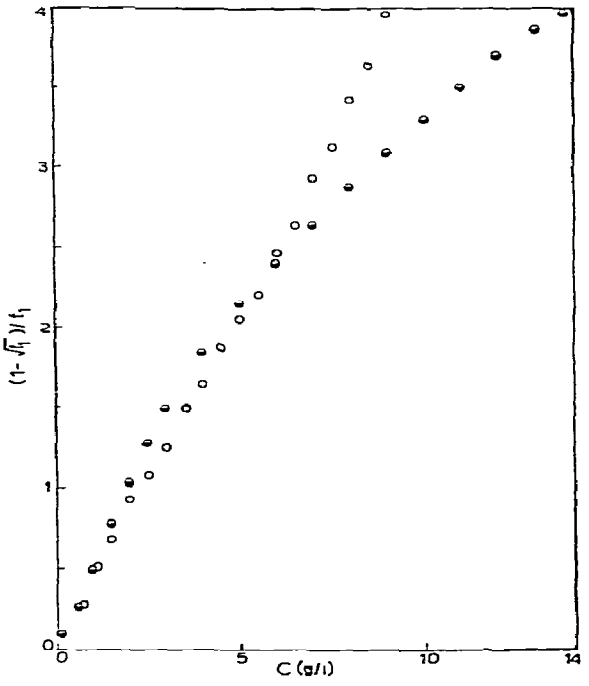

Fig. 6. β-Lactoglobulin A. Test for a Type I indefinite self-associations (see eqs. 18 and 19b); the curvature in this plot based on eq. (19b) rules out a Type I association when Method I (•) was used. When Method II (○) was used, the plot based on eq. (19b) gave the desired straight line; however, the regenerated values of M_1/M_{wa} versus J would not describe the original data. This is reflected in the value of $\Sigma_i(\delta_i)^2$ in table 3.

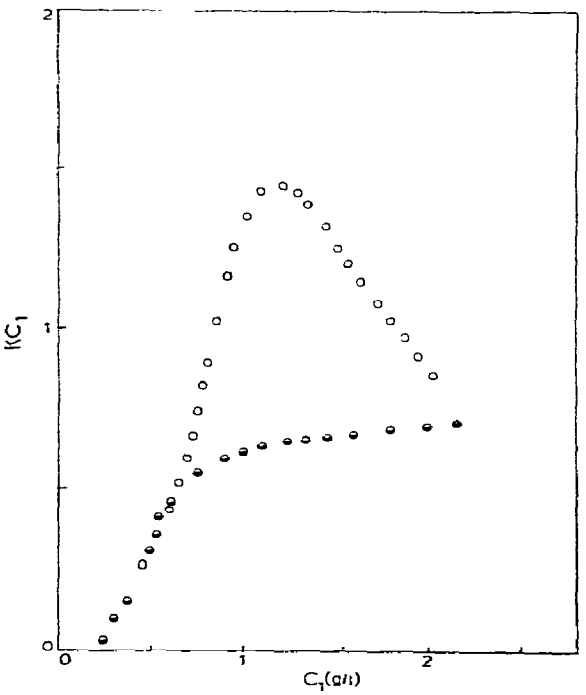

Fig. 7. β-Lactoglobulin A. Test for a Type III indefinite self-association [see eqs. (23) and (24)]; the curvature in these plots (• Method I; ○ Method II) based on eq. (24) rules out a Type II association.

the lowest value of $M_1/M_{wa} = 0.2594$ at $J \approx 58$ fringes and that the trend of the plot of M_1/M_{wa} versus J suggested that still lower values of M_1/M_{wa} might be encountered at higher concentrations. Thus monomer-n-mer associations with $n < 5$ were not considered; tests with other monomer-n-mer association from $n = 5$ to $n = 8$ using methods based on $\xi = (2M_1/M_{na}) - (M_1/M_{wa})$ indicated that none of these models gave a description of the observed self-association. We also considered various 1, m, n type associations (1,2,4; 1,2,6 and 1,2,8), and none of these seemed to give a good description of the observed self-association. Here our analysis was based on the quantity $\ln(f_a/f_{a^*})$. We were unable to find a solution for the 1,2,4 association. A 1,2,6 self-association could be rejected for the following reasons: 1) If $\hat{B}M_1 < 0$, then some of the values of K_6 became negative, which is physically unreal. 2) If $\hat{B}M_1 > 0$, then a plot based on $(C - C_1)/C_1^2 = K_2 + K_6C_1^4$ gave very poor correlation (0.4 to 0.65 correlation coefficient) with a straight line, or else negative values of f_1 were encountered, which is impossible since $0 \leqslant f_1 \leqslant 1$. Similar results were obtained with a 1,2,8 association.

Since previous studies by Adams and Lewis [4] had indicated the presence of a Type I indefinite self-association, this model was considered next. Values of ξ were calculated and used for evaluation of f_1 as described earlier (see section 3.1). A plot of $(1 - \sqrt{f_1})/f_1$ versus C shown in fig. 6 gave pronounced curvature instead of a straight line going through or close to the origin, as required by the theory. Thus this model was ruled out by the plot shown in fig. 6. Other tests using the appropriate form of $\ln(f_a/f_{a^*})$ supported this observation. The results of these tests and tests for other indefinite self-associations are displayed in table 3. If

one assumed a Type II indefinite self-association were present, then values of $x = kC_1$ and C_1 could be calculated as described earlier (see section 3.2.). Fig. 7 shows a plot of x versus C_1; the curvature in this plot rules out the presence of a Type II association. Tests with a Type III and a Type IV indefinite self-association indicated that the Type III association was the better choice, since the value of $\Sigma_i(\delta_i)^2$ listed in the first part of table 3 is lower for the Type III association. The results of tests with the appropriate forms of $\ln(f_a/f_{a^*})$ are listed in the second part of table 3, and it is evident from the values of $\Sigma_i(\delta_i)^2$ that a Type III association was the best model to describe the observed self-association.

6. Discussion

We have shown with real and simulated examples how one might analyze four types of indefinite self-associations. The results we obtained are very encouraging; however, there still may be situations in which it may be impossible to distinguish between two self-associations. This may be due to the fact that only a very limited range of concentration was studied, imprecision of the data or a combination of both factors. In some cases scarcity of material may be a limiting factor in what concentration range may be studied. Attempts to analyze any self-association from one or two solutions of different initial concentrations, or from one or two loadings in a high speed equilibrium experiments should be avoided. Nonideal, non-associating heterogeneous solutes could give plots of M_{wa} versus J that resemble a self-association [31], and at low concentrations two or more models may show almost identical variation of M_{wa} with c or J (here one can encounter ill-conditioned equations). The treatment presented here can deal with non-ideal self-associations, whereas early treatments of Type I and III indefinite self-associations have been restricted to ideal dilute solutions [13–15,32]. For the ideal case ($\hat{B}M_1 \approx 0$) our analysis is consistent with earlier methods. For example, our eq. (19a) can be shown to be formally identical to Kreuzer's [15] eq. (23a), and our eq. (19b) is identical to Coggeshall and Saier's [13] eq. (1).

Heyn and Bretz [33] have referred to Type I indefinite self-associations as noncooperative self-associations and to Type III indefinite self-associations as cooperative. These designations could also be applied to Types II (noncooperative) and Types IV (cooperative) Thus any indefinite self-association involving two equilibrium constants could be referred to as a cooperative association. The need for two equilibrium constants may imply that the bonding between the monomers in the formation of dimers is different from the subsequent bonding for the higher aggregates. It should also be realized that these indefinite self-associations are simple mathematical models that may describe a more complicated real situation, especially for linear associations. For example, Tobolsky and Thach [32] reanalyzed the data of White and Kilpatrick [34] for the self-association of 2-n-butylbenzimidazole and benzotriazole in benzene. They showed that two equilibrium constants, i.e., a Type III indefinite self-association, gave ±1.2% average deviation from the experimental data, whereas White and Kilpatrick [34] had to invoke eight equilibrium constants to achieve the same precision.

Muller [35] has invoked a Type II indefinite self-association to describe the micellization of ionic surfactants in nonaqueous solvents. From calculations of electrostatic potentials he concludes that odd species beyond monomer are less stable than the even species; thus two molecules of an odd species (trimer, pentamer, etc.) would disproportionate to form one molecule of a higher aggregate and one molecule of a lower aggregate. For example, the trimer would disproportionate as follows: $2P_3 \rightarrow P_2 + P_4$. The electrostatic potential energy of the dimer plus tetramer is lower than the electrostatic potential of two trimers, hence the disproportionation. It should be noted, however, that Lo et al. [11] found that the self-association of the surfactant, dodecylammonium propionate, was best described as a Type I indefinite self-association. Nonetheless, Muller's [35] treatment offers one explanation of how a Type II self-association could arise.

A Type IV self-association could be caused by having a very strong association constant for dimer, so that at real concentrations more dimer than monomer would be present, and the subsequent self-association would involve dimers. Alternatively, one can still have a strong dimerization constant, but there may be energetic or steric reasons making the trimer and other higher odd species unfavorable.

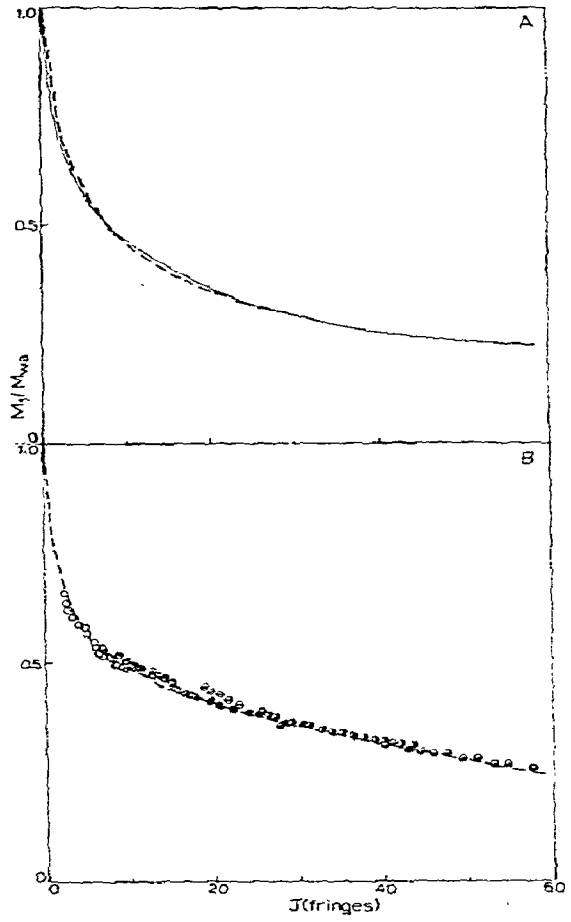

Fig. 8. β-Lactoglobulin A. Comparison plots of M_1/M_{wa} versus J for β-lactoglobulin A at 16°C and pH 4.65 in two different acetate buffers. (A). 0.1 M acetic acid, 0.1 M sodium acetate, $I = 0.10$; this plot is based on the data of Adams and Lewis [4]. (B). 0.1 M acetic acid, 0.1 M sodium acetate, 0.05 M KCl, $I = 0.15$; the plot is taken from this work. Plot A shows a comparison between the smoothed data (solid line) and the data regenerated (dashed line) using $k = 434$ ml/g and $\hat{B}M_1 \approx 1.6$ ml/g. The dashed line is based on a re-analysis of their data using the quantity ξ. Plot B shows how well the Type III indefinite self-association (dashed line), with $k_{12} = 8.13 \times 10^2$ ml/g, $k = 0.280 \times 10^3$ ml/g and $\hat{B}M_1 = -3.59$ ml/g fits the experimental data.

The results reported here with the β-lactoglobulin A (βA) in the 0.15 ionic strength acetate buffer were best interpreted as a Type III sequential indefinite self-association. Earlier studies at the same pH and temperature by Adams and Lewis [4] with βA in an acetate buffer of 0.1 ionic strength had indicated that the self-association was best characterized as a Type I sequential indefinite self-association. At that time Adams and Lewis used a different method for the analysis [4,19] than the one suggested here, but subsequent reanalysis of the Adams and Lewis data, using the quantity ξ and the procedures suggested here, have shown the Type I indefinite self-association to be an excellent model [1,25]. Fig. 8 also shows a comparative plot of M_1/M_{wa} versus J data for the two studies. The difference in solution behavior, while not too pronounced, is probably due to the change in chemical environment caused by the differences in ionic strength.

It should be noted that Timasheff and Townend [36,37] found that the ΔG° for the self-association of βA at low temperature, 4–5°C, and in the pH range of 4–5, was independent of changes in ionic strength of the supporting electrolyte for ionic strengths between 0.01–0.30. In 0.1 ionic strength acetate buffer at pH 4.65, βA shows a temperature-dependent self-association, the association being strongest at 10°C and weakest at 30°C. The β-lactoglobulin C (βC) variant on the other hand shows no temperature dependence of the self-association over the same temperature range. In fact changing the ionic strength from 0.1 to 0.2 had no effect on the self-association. All of the data could be put on the same plot of M_1/M_{wa} versus J and seemed to be best described as a very strong monomer–dimer association [42]. The work of Timasheff and Townend [38–41], the chemical modification studies of Armstrong and McKenzie [43], and the sedimentation equilibrium studies of Sarquis and Adams [42], suggest quite strongly that the aspartic acid residue at sequence position 64 is necessary for the anomalous association (association beyond dimer) of βA between pH 3.5 and 5.2. Both βC and βD, which have a glycine residue at sequence position 54, do not show anomalous self-association in this region.

Earlier light scattering studies by Timasheff and Townend [39] suggested that βA undergoes a dimer–octamer association in the pH range 3.7–5.2. However, later light scattering studies by Kumosinski and Timasheff [40] suggested that these data were better interpreted as a progressive octamerization (they re-

ferred to it as a progressive tetramerization, since their monomer was the dimer, i.e., the 36 844 dalton molecular weight unit) of the form dimer–tetramer–hexamer–octamer. They assumed that the association constants were equal, and they invoked a steric factor of 1/4 for the last association constant. The sedimentation equilibrium studies of Adams and Lewis [4], the studies reported here, as well as the studies reported by Tang [26] clearly indicate that at pH 4.7 the monomeric unit is the 18 422 dalton unit. This has been confirmed by sedimentation equilibrium experiments at low concentrations using a photoelectric scanner. It should be noted that Timasheff and Townend [37] report that no association has been detected in the β-lactoglobulin C and D variants in the pH region between 3.8 and 5.1, whereas Sarquis and Adams [42] observed a very strong monomer–dimer self-association for βC at pH 4.65. The self-association to dimer was so strong that it was only detected with the aid of an ultracentrifuge equipped with a photoelectric scanner. Discrepancies between light scattering and sedimentation equilibrium experiment results have been observed with the self-association of insulin [21,44–46]. Here one encounters another strong self-association in the low concentration region. On the other hand sedimentation equilibrium [27,47,48] and light scattering [49] studies with βA and βB between pH 2–3, where not as strong a monomer–dimer association exists, gave reasonable agreement with each other. Thus it appears, when one is dealing with a strong self-association, the sedimentation equilibrium experiment may be a more sensitive, experimental method for these studies than is the light scattering experiment.

Although the Adams and Lewis [4] data strongly suggest the presence of a Type I indefinite self-association, and the data presented here strongly support a Type III indefinite self-association, the real situation may only involve a limited number of associating species. The precision and accuracy of the experiment is limited by the experimental equipment currently available, and this forces one to use simpler models. We would need a much higher degree of precision (which would have its attendant problems) to really resolve the situation.

Although we have used B-splines to smooth the data, and then used methods involving relations between M_{wa}, M_{na} and $\ln f_a$ for the subsequent analysis of self-associations, one could also use nonlinear least squares [50] on the primary data, i.e., M_{wa} and c, for the analysis of the self-association. Professor P.W. Smith (personal communication), of the Mathematics Department at Texas A&M University, has told us that nonlinear least squares may be the best method for analyzing self-associations, but it also may be computationally very expensive, particularly if one has to check several models for the observed self-association. One must use the appropriate equation for M_{wa} for each model tested and do the nonlinear least squares smoothing each time for each model. Our procedures may be computationally less expensive, since the experimental data, M_{wa} versus C or M_1/M_{wa} versus C, is smoothed only one time using B-splines, which is a least squares method. Another advantage of our procedure is that it can be more readily used by others who have less sophisticated computational facilities or less mathematical expertise in this area.

Acknowledgement

Part of this work is based on the dissertation submitted by Lih-Heng Tang to the Graduate School, Illinois Institute of Technology in partial fulfillment for the degree of Doctor of Philosophy (Dec. 1971). This work was supported by grants from the National Institutes of Health (GM15551, GM17611) and The Robert A. Welch Foundation (A485). We thank Drs. E.B. Kalan and J.J. Basch for furnishing the β-lactoglobulin A used in this work.

Appendix

A. Derivation of eq. (26)

For the Type II indefinite self-association one notes that

$$\frac{C}{C_1} - 1 = \frac{2kC_1}{(1-k^2C_1^2)^2} = \frac{2x}{(1-x^2)^2} \tag{A1}$$

$$\frac{CM_1}{C_1M_{nc}} - 1 = \frac{kC_1}{1-k^2C_1^2} = \frac{x}{1-x^2} \tag{A2}$$

and

$$\frac{CM_{wc}}{C_1M_1} - 1 = \frac{4kC_1(1+k^2C_1^2)}{(1-k^2C_1^2)^3} = \frac{4x(1+x^2)}{(1-x^2)^3}. \quad \text{(A3)}$$

Here $x = kC_1$. These equations can be combined to give

$$R_1 = \frac{CM_1/C_1M_{nc} - 1}{C/C_1 - 1} = \frac{1-x^2}{2} \quad \text{(A4)}$$

and

$$R_2 = \frac{CM_{wc}/C_1M_1 - 1}{C/C_1 - 1} = \frac{2(1+x^2)}{1-x^2} \quad \text{(A5)}$$

Eqs. (A4) and (A5) can be combined to yield

$$R_2R_1 + 2R_1 - 2 = 0 \quad \text{(A6)}$$

This can also be written as

$$\frac{(C_1M_{wc}/C_1M_1 - 1)(CM_1/C_1M_{nc} - 1)}{(C/C_1 - 1)^2} + \frac{2(CM_1/C_1M_{nc} - 1)}{C/C_1 - 1} - 2 = 0 \quad \text{(A7)}$$

Now do the following steps: 1) multiply eq. (A7) by $(C/C_1 - 1)^2$; 2) then multiply the resulting equation by C_1^2; and 3) finally, divide by C^2 to obtain

$$\left(\frac{M_{wc}}{M_1} - f_1\right)\left(\frac{M_1}{M_{nc}} - f_1\right) + 2\left(\frac{M_1}{M_{nc}} - f_1\right) \times (1-f_1) - 2(1-f_1)^2 = 0. \quad \text{(A8)}$$

Eq. (A8) can be rearranged to give

$$f_1^2 - f_1\left[\frac{3M_1}{M_{nc}} + \frac{M_{wc}}{M_1} - 2\right] + \left(\frac{M_{wc}}{M_{nc}} + \frac{2M_1}{M_{nc}} - 2\right) = 0.$$

The solution to this quadratic equation in f_1 is given by eq. (25).

B. Derivation of eq. (39).

For the Type III indefinite self-association the analogs of eqs. (A1)–(A3) are

$$\frac{C}{C_1} - 1 = \frac{k_{12}C_1(2-kC_1)}{(1-kC_1)^2} = \frac{y(2-x)}{(1-x)^2} \quad \text{(B1)}$$

$$\frac{CM_1}{C_1M_{nc}} - 1 = \frac{k_{12}C_1}{1-kC_1} = \frac{y}{1-x} \quad \text{(B2)}$$

and

$$\frac{CM_{wc}}{C_1M_1} - 1 = \frac{k_{12}C_1(4-3kC_1+k^2C_1^2)}{(1-kC_1)^3} = \frac{y(4-3x+x^2)}{(1-x)^3} \quad \text{(B3)}$$

Here $y = k_{12}C_1$ and $x = kC_1$. These equations can be combined to give

$$R_3 = \frac{C/C_1 - 1}{CM_1/C_1M_{nc} - 1} = \frac{1-x}{2-x} \quad \text{(B5)}$$

and

$$R_4 = \frac{C/C_1 - 1}{CM_{wc}/C_1M_1 - 1} = \frac{2+3x+x^2}{4-3x+x^2}. \quad \text{(B6)}$$

From eq. (B5) one notes that

$$x = \frac{1-2R_3}{1-R_3} \quad \text{(B7)}$$

Insertion of eq. (B7) into eq. (B6) yields after rearrangement

$$2R_4 - 3R_3R_4 + 2R_3^2R_2 - R_3 = 0 \quad \text{(B8)}$$

After substituting eqs. (B5) and (B6) into eq. (B8) and doing various algebraic manipulations one obtains eq. (39) for f_1.

C. Type IV indefinite self-association

For the Type IV indefinite self-association, the analogs of eqs. (A1)–(A3) are

$$\frac{C}{C_1} - 1 = \frac{2kC_1}{(1-k_{12}kC_1^2)^2} = \frac{2x}{(1-xy)^2} = \frac{2x}{(1-z^2)^2}, \quad \text{(C1)}$$

$$\frac{CM_1}{C_1M_{nc}} - 1 = \frac{kC_1}{(1-k_{12}kC_1^2)} = \frac{x}{1-xy} = \frac{x}{1-z^2} \quad \text{(C2)}$$

and

$$\frac{CM_{wc}}{C_1M_1} - 1 = \frac{4kC_1(1+kk_{12}C_1^2)}{(1-k_{12}kC_1^2)^3} = \frac{4x(1+xy)}{(1-xy)^3} = \frac{4x(1+z^2)}{(1-z^2)^3}. \quad \text{(C3)}$$

Thus

$$R_1 = \frac{CM_1/C_1MC_{nc} - 1}{C/C_1 - 1} = \frac{1-z^2}{2} \quad \text{(C4)}$$

which is formally identical with eq. (A4), and the corresponding equation for R_2 would be formally identical to eq. (A5). Hence eq. (25) also is the equation for f_1 for this self-association. To distinguish between Types II and IV, one would have to see whether one or two equilibrium constants are needed to describe eqs. (C1)–(C3), since these are the analogs of eqs. (A1)–(A3) for a type IV association. If the association is Type II, then one has C_1 from eq. (25), and one can use C_1 to evaluate x^2 and x from eqs. (A4) or (A5). These values of C_1 and x can be used in various values of eqs. (A1)–(C3) to see if they satisfy these equations. If these equations are satisfied within reasonable error limits, then a Type II association is present, and this should be confirmed by the use of eq. (23), since one can evaluate $x = kC_1$ directly from eq. (23) for a Type II association. If these criteria are not met and a plot of x versus C_1, using x values obtained from eq. (23), is curved instead of being a straight line going through or close to the origin, then a Type IV association may be present.

References

[1] E.T. Adams Jr., W.E. Ferguson, P.J. Wan, J.L. Sarquis and B.M. Escott, Separation Science 10 (1975) 175.
[2] M.S.N. Rao and G. Kegeles, J. Amer. Chem. Soc. 80 (1958) 5724.
[3] D.K. Hancock and J.W. Williams, Biochemistry 8 (1969) 2598.
[4] M.S. Lewis and E.T. Adams Jr., Biochemistry 7 (1968) 1044.
[5] A.H. Pekar and B.H. Frank, Biochemistry 11 (1972) 4013.
[6] P.W. Chun, S.J. Kim, C.A. Stanley and G.K. Ackers, Biochemistry 8 (1969) 1625.
[7] K.E. Van Holde and G.P. Rossetti, Biochemistry 6 (1967) 2189.
[8] R. Bretz, A. Lustig and G. Schwarz, Biophys. Chem. 1 (1974) 237.
[9] E. Braswell, J. Phys. Chem. 78 (1968) 2477.
[10] H.-G. Elias and R. Bareiss, Chimia 21 (1967) 53.
[11] Fred Y.-F. Lo, B.M. Escott, E.J. Fendler, E.T. Adams Jr. R.D. Larsen and P.W. Smith, J. Phys. Chem. 79 (1975) 2609.
[12] E.G. Hoffman, Z. Physik. Chem. B53 (1943) 179.
[13] N.D. Coggeshall and E.L. Saier, J. Amer. Chem. Soc. 73 (1951) 5414.
[14] M. Davies and D.K. Thomas, J. Phys. Chem. 60 (1956) 763, 767.
[15] J. Kreuzer, Z. Phys. Chem. B53 (1943) 213.
[16] A. Vrij, Ph.D. Dissertation, University of Utrecht, 1969.
[17] A. Vrij and J.Th.G. Overbeek, J. Colloid Sci. 17 (1962) 570.
[18] E.F. Casassa and H. Eisenberg, Advan. Protein Chem. 19 (1964) 287.
[19] E.T. Adams, Jr., Fractions No. 3, 1967.
[20] A.G. Ogston and D.J. Winzor, J. Phys. Chem. 79 (1975) 2496.
[21] E.T. Adams, Jr., and Hiroshi Fujita, in: Ultracentrifugation in theory and experiment, ed. J.W. Williams (Academic Press, New York, 1973) p. 119.
[22] Hiroshi Fujita, Foundations of ultracentrifugal analysis (Wiley–Interscience, New York, 1975) Chap. 6.
[23] B.K. Millthorpe, P.D. Jeffrey and L.W. Nichol, Biophys. Chem. 3 (1975) 169.
[24] W.E. Ferguson, C.M. Smith, E.T. Adams Jr. and G.H. Barlow, Biophys. Chem. 1 (1974) 325.
[25] P.W. Chun, S.J. Kim, J.D. Williams, W.T. Cope, L.-H. Tang and E.T. Adams Jr., Biopolymers 11 (1972) 197.
[26] Lih-Heng Tang, Ph.D. Dissertation, Illinois Institute of Technology, Chicago, 1971.
[27] L.-H. Tang and E.T. Adams Jr., Arch. Biochem. Biophys. 157 (1973) 520.
[28] M. Halwer, G.C. Nutting and B.A. Brice, J. Amer. Chem. Soc. 73 (1951) 2786.
[29] H.B. Bull and K. Breese, J. Phys. Chem. 72 (1968) 1817.
[30] V.G. Frank and G. Braunitzer, Hoppe-Seyler Z. Physiol. Chem. 348 (1967) 1691.
[31] P.D. Jeffrey, Biochemistry 7 (1968) 3345, 3352; 8 (1969) 5217.
[32] A.V. Tobolsky and R.E. Thach, J. Colloid Sci. 17 (1962) 410.
[33] M.P. Heyn and R. Bretz, Biophys. Chem. 3 (1975) 35.
[34] N.E. White and M. Kilpatrick, J. Phys. Chem. 59 (1955) 1044.
[35] N. Muller, J. Phys. Chem. 69 (1975) 287.
[36] T.T. Herskovits, R. Townend and S.N. Timasheff, J. Amer. Chem. Soc. 86 (1964) 4445.
[37] S.N. Timasheff and R. Townend, Proteides Biol. Fluids 16 (1969) 33.
[38] R. Townend and S.N. Timasheff, J. Amer. Chem. Soc. 82 (1960) 3168.
[39] S.N. Timasheff and R. Townend, J. Amer. Chem. Soc. 83 (1961) 464.
[40] T.F. Kumosinski and S.N. Timasheff, J. Amer. Chem. Soc. 88 (1966) 5635.
[41] T. Townend, R.J. Winterbottom and S.N. Timasheff, J. Amer. Chem. Soc. 82 (1960) 3161.
[42] J.L. Sarquis and E.T. Adams Jr., Biophys. Chem. 4 (1976) 181.
[43] J. McD. Armstrong and H.A. McKenzie, Biochem. Biophys. Acta 147 (1967) 93.
[44] R.F. Steiner, Arch. Biochem. Biophys. 39 (1952) 333.
[45] P. Doty and G.E. Myers, Discussions Faraday Soc. 13 (1953) 51.
[46] P.D. Jeffry and J.H. Coates, Nature 197 (1963) 1104; Biochemistry 5 (1966) 489, 3820.

[47] D.A. Albright and J.W. Williams, Biochemistry 7 (1968) 1044.
[48] J. Visser, R.C. Deonier, E.T. Adams Jr. and J.W. Williams Biochemistry 11 (1972) 2634.
[49] S.N. Timasheff and R. Townend, J. Amer. Chem. Soc. 83 (1961) 470.
[50] M.S. Lewis and G.C. Knott, Biophys. Chem. 5 (1976) 171.