Biophysical Chemistry 3 (1975) 46–55

DETERMINATION OF THE EQUILIBRIUM CONSTANTS OF SELF-ASSOCIATING PROTEIN SYSTEMS

XI. The application of $C_{(r)}$ in graphical analysis and the enumeration of interacting species in the ultracentrifuge*

Paul W. CHUN** and Sung J. KIM‡

Department of Biochemistry, College of Arts and Sciences, University of Florida, Gainesville, Florida 32610, USA

Received 14 March 1974, revised manuscript received 31 July 1974

A greatly simplified procedure is proposed which employs $C = f(r)$ as determined from sedimentation equilibrium measurements in graphical analysis of self-associating protein systems and in the enumeration of interacting species in the ultracentrifuge. Basic equations given here are applicable to any self-associating system. A procedure is outlined for enumeration of interacting components independent of non-ideal behavior, using principal component analysis.

1. Introduction

In recent years the study of associating protein systems has developed in two basic directions. The most widely applied procedure considers the weight average molecular weight distribution as a function of concentration [1–18]. More recently, a second approach has been applied based on concentration distribution as a function of radial distance in sedimentation equilibrium measurements [15, 18–21]. Although significant progress has been made in recent years, evaluation of interacting components based on molecular weight distribution is still a difficult problem.

Since the equilibrium concentration profile is highly sensitive to the species distribution, the evaluation of the concentration C as a function of radial distance r can be used to determine the mode of association. In addition, evaluation of $C = f(r)$ provides information which permits determination of the number of species participating in self-association equilibria, independent of non-ideal behavior.

The procedure most widely used in molecular biology and biochemistry in the past to determine the number of linearly independent components in an interacting system employs matrix rank analysis [22–25]. In molecular sieve chromatography, a series of weight average partition coefficients, each at a different equilibrium state and gel porosity, have been used to enumerate interacting components [26]. Matrix rank analysis has been applied to a series of diffusion measurements to enumerate interacting species in the ultracentrifuge [27]. Most recently, the use of factor or principal component analysis, two long-recognized statistical techniques, has been proposed as a more effective alternative to matrix rank analysis for component enumeration [28, 29].

In this communication, we describe the application

* This work was supported by National Science Foundation Grants GB17848, GB28223-A-1,2 and in part by General Research Support, College of Medicine, University of Florida.

** All correspondence to P.W. Chun, Room 407, Nuclear Science Center, Department of Biochemistry, College of Arts and Science, University of Florida.

‡ S.J. Kim, Postdoctoral Fellow (1969–1971), Military Attache, Embassy of Korea, Washington, D.C.

of $C_{(r)}$ in graphical analysis of self-associating protein systems and in the enumeration of the interacting species. The effectiveness of principal component analysis for enumeration of these components in the ultracentrifuge is discussed.

2. Evaluation of $C_{(r)}$

The assumptions that are used in analysis of self-associating protein systems based on sedimentation equilibrium measurements have been previously described. The differential equation for sedimentation equilibrium [30–33] is given by

$$\frac{\mathrm{d}\ln C_i}{\mathrm{d}r^2} = \frac{\omega^2(\partial\rho/\partial C_i)M_i}{2RT} = LM_i, \tag{1}$$

where C_i is the concentration of the component A_i and $M_i = iM_1$. The partial differential, $\partial\rho/\partial C_i$, at constant concentration of all other species, equals $(1-\bar{v}\rho)$, where $\bar{v}$ is the partial specific volume and ρ is the density of the solution. Eq. (1) is integrated between any two limits, a and b, provided that $r_m \leqslant a$ and $r_b \leqslant b$, where r_m and r_b refer to meniscus and bottom of cell respectively. Thus

$$M_{\mathrm{wapp}} = L^{-1}\,\mathrm{d}\ln C/\mathrm{d}r^2, \tag{2}$$

where $L = (1-\bar{v}\rho)\omega^2/2RT$. Here, ω is the angular velocity. The apparent weight average molecular weight [3,31] for the self-associating system is represented by

$$\frac{1}{M_{\mathrm{wapp}}} = \frac{1}{M_{\mathrm{w(c)}}} + BC,$$

where $M_{\mathrm{w(c)}}$ is the concentration dependent weight average molecular weight and B is the second virial coefficient. By definition, the weight average molecular weight $M_{\mathrm{w}} = \Sigma_i C_i M_i/C$ and $C = \Sigma_i K_i C_1^i$ where C_1 is the concentration of monomer, and C is the total initial concentration. The expression for molecular weight, including the non-ideality term BM_1, becomes

$$M_{\mathrm{wapp}} = \frac{M_{\mathrm{w(c)}}}{1+M_{\mathrm{w(c)}}BC} = \frac{M_1}{C}\,\frac{\Sigma_i iK_iC_1^i}{1+BM_1\Sigma_i iK_iC_1^i}. \tag{3}$$

The concentration gradient at a given place in the cell column is expressed as

$$\frac{\mathrm{d}C}{\mathrm{d}r^2} = \frac{\mathrm{d}C_1}{\mathrm{d}r^2}\frac{\mathrm{d}C}{\mathrm{d}C_1} = \frac{\mathrm{d}C_1}{\mathrm{d}r^2}\sum_i iK_iC_1^{i-1} = \frac{\mathrm{d}C_1}{C_1\,\mathrm{d}r^2}\sum_i iK_iC_1^i. \tag{4}$$

Combining eqs. (2) and (4),

$$M_{\mathrm{wapp}} = \frac{1}{C}\sum_i iK_iC_1^iL^{-1}\,\frac{\mathrm{d}C_1}{C_1\,\mathrm{d}r^2}. \tag{5}$$

From eqs. (3) and (5),

$$\left(1+BM_1\sum_i iK_iC_1^i\right)\frac{\mathrm{d}C_1}{C_1\,\mathrm{d}r^2} = LM_1, \tag{6}$$

$$\frac{\mathrm{d}\ln C_1}{\mathrm{d}r^2} + BM_1\frac{\mathrm{d}C}{\mathrm{d}r^2} = LM_1, \tag{7}$$

in which the $\mathrm{d}C_1/\mathrm{d}r^2$ term is obtained from eq. (4). Integration of eq. (7) yields

$$\ln C_1/C_{1r_0} + BM_1(C-C_{r_0}) = LM_1(r^2-r_0^2), \tag{8}$$

$$C_1 = C_{1r_0}\exp[LM_1(r^2-r_0^2) - BM_1(C-C_{r_0})], \tag{9}$$

where C_{1r_0} is the concentration of monomer at the radial distance r_0. Since the concentration for the ith species (see appendix I) is expressed as

$$\begin{aligned} C_i &= K_iC_{1r_0}^i\exp i[LM_1(r^2-r_0^2) - BM_1(C-C_{r_0})], \\ C_i &= C_{ir_0}\exp i[LM_1(r^2-r_0^2) - BM_1(C-C_{r_0})], \end{aligned} \tag{10}$$

where C_{ir_0} is the concentration of ith species at the radius r_0. Letting

$$\phi_r = \exp[LM_1(r^2-r_0^2)], \quad \psi_r = \exp[-BM_1(C-C_{r_0})], \tag{11}$$

then

$$C_{ir} = C_{ir_0}\phi_r^i\psi_r^i, \tag{12}$$

$$C_r = \sum_i C_{ir} = \sum_i C_{ir_0}\phi_r^i\psi_r^i, \tag{13}$$

where C_r is the concentration distribution at a given radial distance r. In order to evaluate the concentration as a function of radial distance obtained from either schlieren or Rayleigh interference fringe data [19,32,34], $(n+1)$ equidistant points are chosen,

$C_{r_0}, C_{r_1}, C_{r_2}, \cdots, C_{r_n}$, where $r_0 \geqslant a$ and $r_n \leqslant b$.

$$C_{r_1} = C_{r_0} + \Delta c,$$
$$C_{r_2} = C_{r_0} + 2\Delta c,$$
$$C_{r_3} = C_{r_0} + 3\Delta c,$$
$$\vdots$$
$$C_{r_n} = C_{r_0} + n\Delta c. \tag{14}$$

Substituting values from eq. (14) into eq. (11),

$$\psi_{r_1} = \exp[-BM_1(C_{r_1} - C_{r_0})] = \exp[-BM_1\Delta c] = \beta,$$
$$\psi_{r_n} = \exp[-BM_1(C_{r_n} - C_{r_0})] = \exp[-BM_1 n\Delta c] = \beta^n. \tag{15}$$

Therefore, eq. (13) can be written as

$$C_{r_n} = \sum_i C_{ir_0}\phi_{r_n}^i\beta^{in}. \tag{13a}$$

Eq. (13a) is fundamental in the evaluation of equilibrium constants and the mode of association for self-associating protein systems.

3. The application of $C_{(r)}$ to self-associating protein systems

To illustrate its use, eq. (13a)'s application to several cases of discrete and indefinite association will be considered in greater detail.

Case (i): $nP_i \rightleftarrows P_n, \quad (n > 1).$

A set of three equidistant points, C_{r_0}, C_{r_1}, and C_{r_2}, is selected for evaluation of C_{1r_0}, C_{nr_0}, BM_1 and K_n. A second set of points reconfirms these interaction parameters, and so on until the values of BM_1 and K_n are consistent, at which point the correct mode of association has been determined. This requires the evaluation of scores of data points from several independent sedimentation equilibrium experiments, varying the initial concentration. A similar procedure is followed for cases of three or four-species association.

$$C_{r_0} = C_{1r_0} + C_{nr_0},$$
$$C_{r_1} = C_{1r_0}\phi_{r_1}\beta + C_{nr_0}\phi_{r_1}^n\beta^n,$$
$$C_{r_2} = C_{1r_0}\phi_{r_2}\beta^2 + C_{nr_0}\phi_{r_2}^n\beta^{2n}. \tag{16}$$

Evaluation of C_{1r_0}, C_{nr_0} and β is accomplished by simultaneous elimination, using three equations and three unknowns as follows.

$$C_{r_1} - C_{r_0}\phi_{r_1}\beta = C_{nr_0}(\phi_{r_1}^n\beta^n - \phi_{r_1}\beta),$$
$$C_{r_2} - \phi_{r_2}C_{r_0}\beta^2 = C_{nr_0}(\phi_{r_2}^n\beta^{2n} - \phi_{r_2}\beta^2). \tag{17}$$

Thus,

$$(C_{r_1} - \phi_{r_1}\beta C_{r_0})(\phi_{r_2}^n\beta^{2n} - \phi_{r_2}\beta^2) = (C_{r_2} - \phi_{r_2}\beta^2 C_{r_0})(\phi_{r_1}^n\beta^n - \phi_{r_1}\beta) \tag{18}$$

may be rewritten as

$$\beta^{2n}(-C_{r_0}\phi_{r_1}\phi_{r_2}^n) + \beta^{2n-1}(C_{r_1}\phi_{r_2}^n) + \beta^{n+1}(C_{r_0}\phi_{r_1}^n\phi_{r_2}) + \beta^{n-1}(-C_{r_2}\phi_{r_1}^n) + \beta(-C_{r_1}\phi_{r_2}) + C_{r_2}\phi_{r_1} = 0. \tag{18a}$$

β, an expression of non-ideality, can be evaluated from eq. (18a) since C_{r_0}, C_{r_1}, ϕ_{r_1} and ϕ_{r_2} are known quantities. The unknown quantities C_{nr_0}, C_{1r_0}, and BM_1 may be obtained as follows

$$C_{nr_0} = \frac{C_{r_1} - \phi_{r_1}\beta C_{r_0}}{(\phi_{r_1}\beta)^n - \phi_{r_1}\beta}, \quad C_{1r_0} = C_{r_0} - C_{nr_0},$$
$$BM_1 = (-\ln\beta)/\Delta C, \quad K_n = C_{nr_0}/(C_{1r_0})^n. \tag{19}$$

A similar procedure is followed for three species and four species association. As is characteristic of the non-ideality effect, the linearity tends to increase with increasing BM_1 value.

Case (ii): $qP_1 \rightleftarrows rP_m + SP_n, \quad (n > m > 1)$

In order to analyze a case in which three species are involved in chemical equilibrium, using concentration as a function of radial distance, one must select several sets of data points r_0, r_1, r_2 and r_3 so that $C_{r_3} - C_{r_2} = C_{r_2} - C_{r_1} = C_{r_1} - C_{r_0} = \Delta C$.

$$C_{r_0} = C_{1r_0} + C_{mr_0} + C_{nr_0},$$
$$C_{r_1} = C_{1r_0}\phi_{r_1}\beta + C_{mr_0}(\phi_{r_1}\beta)^m + C_{nr_0}(\phi_{r_1}\beta)^n,$$
$$C_{r_2} = C_{1r_0}\phi_{r_2}\beta^2 + C_{mr_0}(\phi_{r_2}\beta^2)^m + C_{nr_0}(\phi_{r_2}\beta^2)^n,$$
$$C_{r_3} = C_{1r_0}\phi_{r_3}\beta^3 + C_{mr_0}(\phi_{r_3}\beta^3)^m + C_{nr_0}(\phi_{r_3}\beta^3)^n. \tag{20}$$

For simplicity, let $\zeta_1 = \phi_{r_1}\beta, \zeta_2 = \phi_{r_2}\beta^2, \zeta_3 = \phi_{r_3}\beta^3$. Substituting these values into eq. (20) gives

$$\begin{aligned} C_{r_0} &= C_{1r_0} + C_{mr_0} + C_{nr_0}, \\ C_{r_1} &= C_{1r_0}\zeta_1 + C_{mr_0}\zeta_1^m + C_{nr_0}\zeta_1^n, \\ C_{r_2} &= C_{1r_0}\zeta_2 + C_{mr_0}\zeta_2^m + C_{nr_0}\zeta_2^n, \\ C_{r_3} &= C_{1r_0}\zeta_3 + C_{mr_0}\zeta_3^m + C_{nr_0}\zeta_3^n. \end{aligned} \tag{20a}$$

These four equations must be solved simultaneously, using four unknowns. Thus, elimination of C_{1r_0} gives

$$\begin{aligned} C_{r_1} - \zeta_1 C_{r_0} &= (\zeta_1^m - \zeta_1)C_{mr_0} + (\zeta_1^n - \zeta_1)C_{nr_0}, \\ C_{r_2} - \zeta_2 C_{r_0} &= (\zeta_2^m - \zeta_2)C_{mr_0} + (\zeta_2^n - \zeta_2)C_{nr_0}, \\ C_{r_3} - \zeta_3 C_{r_0} &= (\zeta_3^m - \zeta_3)C_{mr_0} + (\zeta_3^n - \zeta_3)C_{nr_0}. \end{aligned} \tag{21}$$

Next, eliminating C_{mr_0},

$$\begin{aligned} &(\zeta_1^m - \zeta_1)(C_{r_2} - \zeta_2 C_{r_0}) - (\zeta_2^m - \zeta_2)(C_{r_1} - \zeta_1 C_{r_0}) \\ &\quad = [(\zeta_1^m - \zeta_1)(\zeta_2^n - \zeta_2) - (\zeta_2^m - \zeta_2)(\zeta_1^n - \zeta_1)]C_{nr_0}, \\ &(\zeta_1^m - \zeta_1)(C_{r_3} - \zeta_3 C_{r_0}) - (\zeta_3^m - \zeta_3)(C_{r_1} - \zeta_1 C_{r_0}) \\ &\quad = [(\zeta_1^m - \zeta_1)(\zeta_3^n - \zeta_3) - (\zeta_3^m - \zeta_3)(\zeta_1^n - \zeta_1)]C_{nr_0}. \end{aligned} \tag{22a}$$

Finally, eliminating C_{nr_0},

$$\begin{aligned} &[(\zeta_1^m - \zeta_1)(C_{r_2} - \zeta_2 C_{r_0}) - (\zeta_2^m - \zeta_2)(C_{r_1} - \zeta_2 C_{r_0})] \\ &\quad \times [(\zeta_1^m - \zeta_1)(\zeta_3^m - \zeta_3) - (\zeta_3^m - \zeta_3)(\zeta_1^n - \zeta_1)] \\ &\quad = [(\zeta_1^m - \zeta_1)(C_{r_3} - \zeta_3 C_{r_0}) - (\zeta_3^m - \zeta_3)(C_{r_1} - \zeta_1 C_{r_0})] \\ &\quad \times [(\zeta_1^m - \zeta_1)(\zeta_2^n - \zeta_2) - (\zeta_2^m - \zeta_2)(\zeta_1^n - \zeta_1)], \end{aligned} \tag{22b}$$

where $\zeta_i = \phi_{r_i}\beta^i$.

β may be evaluated from eq. (22b), since C_{r_0}, C_{r_1}, C_{r_2}, C_{r_3}, ϕ_{r_1}, ϕ_{r_2} and ϕ_{r_3} are all known quantities. Then it is possible to evaluate the five quantities which follow:

$$\begin{aligned} C_{nr_0} &= \frac{(\zeta_1^m - \zeta_1)(C_{r_3} - \zeta_3 C_{r_0}) - (\zeta_3^m - \zeta_3)(C_{r_1} - \zeta_1 C_{r_0})}{(\zeta_1^m - \zeta_1)(\zeta_3^n - \zeta_3) - (\zeta_3^m - \zeta_3)(\zeta_1^n - \zeta_1)}, \\ C_{mr_0} &= \frac{(C_{r_3} - \zeta_3 C_{r_0}) - (\zeta_3^n - \zeta_3)C_{nr_0}}{(\zeta_3^m - \zeta_3)}, \\ C_{1r_0} &= C_{r_0} - C_{mr_0} - C_{nr_0}, \quad BM_1 = (-\ln\beta)/\Delta C, \\ K_m &= C_{mr_0}/(C_{1r_0})^m \text{ and } K_n = C_{nr_0}/(C_{1r_0})^n. \end{aligned} \tag{23}$$

Again negative values of BM_1 may be determined only from a single root of eq. (20), a procedure made easier by the use of a computer.

Case (iii): $qP_1 \rightleftarrows tP_l + rP_m + SP_n, \quad (n > m > l > 1)$

When four species are involved in chemical equilibrium, choose five points on the $C = f(r)$ curve, r_0, r_1, r_2, r_3 and r_4 such that $C_{r_4} - C_{r_3} = C_{r_3} - C_{r_2} = C_{r_2} - C_{r_1} = C_{r_1} - C_{r_0} = \Delta C$.

$$\begin{aligned} C_{r_0} &= C_{1r_0} + C_{lr_0} + C_{mr_0} + C_{nr_0}, \\ C_{r_1} &= C_{1r_0}\zeta_1 + C_{lr_0}\zeta_1^l + C_{mr_0}\zeta_1^m + C_{nr_0}\zeta_1^n, \\ C_{r_2} &= C_{1r_0}\zeta_2 + C_{lr_0}\zeta_2^l + C_{mr_0}\zeta_2^m + C_{nr_0}\zeta_2^n, \\ C_{r_3} &= C_{1r_0}\zeta_3 + C_{lr_0}\zeta_3^l + C_{mr_0}\zeta_3^m + C_{nr_0}\zeta_3^n, \\ C_{r_4} &= C_{1r_0}\zeta_4 + C_{lr_0}\zeta_4^l + C_{mr_0}\zeta_4^m + C_{nr_0}\zeta_4^n, \end{aligned} \tag{24}$$

where $\zeta_i = \phi_{r_i}\beta^i$, $\phi_{r_i} = \exp[LM_1(r_i^2 - r_0^2)]$, $\beta = \exp(-BM_1\Delta C)$.

Equation set (24) is solved as five simultaneous equations in which the quantities C_{1r_0}, C_{lr_0}, C_{mr_0}, and BM_1 are unknown. Elimination of C_{1r_0} through C_{nr_0} yields one equation in which BM_1 can be evaluated. Once β is known, the quantity ζ_i can be evaluated. A set of equations to evaluate C_{1r_0}, C_{lr_0}, C_{mr_0} and C_{nr_0} are determined using column matrix transformation. Thus, $K_n = C_{nr_0}/(C_{1r_0})^n$, $K_m = C_{mr_0}/(C_{1r_0})^m$ and $K_l = C_{lr_0}/(C_{1r_0})^l$ yield the equilibrium constants when four species are involved.

The practical limits of this method become obvious in the evaluation of data points for three or four species in chemical equilibrium. It should be noted that error in the measurements will reflect on the accurate determination of BM_1.

Case (iv): Indefinite association

The theoretical considerations to be applied to indefinite association are similar to the procedure outlined by Elias and Lys [9] and Van Holde and Rossetti [35]. The quantity $kC > 1$, where k is the intrinsic equilibrium constant denoting quantities consistent with Elias and Lys' notation.

The total concentration, C, can be expressed as

$$C = \sum_i ik^{i-1}C_1^i = \frac{C_1}{(1 - kC_1)^2}. \tag{25}$$

Three points, r_0, r_1 and r_2 are selected so that the concentration increment gives $C_{r_1} - C_{r_0} = C_{r_2} - C_{r_1} = \Delta C$. In some cases, several such sets must be evaluated.

$$C_{r_0} = \frac{C_{1r_0}}{(1-kC_{1r_0})^2},$$

$$C_{r_1} = \frac{C_{1r_1}}{(1-kC_{1r_1})^2} = \frac{C_{1r_0}\zeta_1}{(1-kC_{1r_0}\zeta_1)^2}, \qquad (26)$$

$$C_{r_2} = \frac{C_{1r_0}\zeta_2}{(1-kC_{1r_0}\zeta_2)^2}.$$

Letting $kC_{1r_0} = x$,

$$C_{r_0} = \frac{C_{1r_0}}{(1-x)^2}, \; C_{r_1} = \frac{C_{1r_0}\zeta_1}{(1-x\zeta_1)^2}, \; C_{r_2} = \frac{C_{1r_0}\zeta_2}{(1-x\zeta_2)^2}. \qquad (27)$$

Elimination of C_{1r_0} gives

$$C_{r_0}(1-x)^2 = C_{r_1}(1-x\zeta_1)^2/\zeta_1$$

or $$\zeta_1 C_{r_0}(1-x)^2 = C_{r_1}(1-x\zeta_1)^2,$$

$$C_{r_0}(1-x)^2 = C_{r_2}(1-x\zeta_2)^2/\zeta_2$$

or $$\zeta_2 C_{r_0}(1-x)^2 = C_{r_2}(1-x\zeta_2)^2.$$

Eliminating x,

$$(1-\zeta_2)(\sqrt{\zeta_1 C_{r_0}/C_{r_1}} - 1) = (1-\zeta_1)(\sqrt{\zeta_2 C_{r_0}/C_{r_2}} - 1), \qquad (28)$$

$$\zeta_1 = \phi_{r_1}\beta = \exp[LM(r_1^2 - r_0^2)] \exp(-BM_1 \Delta C),$$

$$\zeta_2 = \phi_{r_2}\beta^2 = \exp[LM_1(r_2^2 - r_0^2)] \exp(-BM_1 2\Delta C). \qquad (29)$$

Eqs. (28) and (29) are used for evaluation of BM_1, and eq. (29) gives ζ_1. Knowing ζ_1, it is possible to evaluate

$$\frac{x}{1-x} = \frac{1}{1-\zeta_1}(\sqrt{\zeta_1 C_{r_0}/C_{r_1}} - 1). \qquad (30)$$

From eqs. (25) and (30),

$$C_{1r_0} = C_{r_0}(1-kC_{1r_0})^2 = C_{r_0}(1-x)^2.$$

The equilibrium constant k is computed from the following expression

$$k = x/C_{1r_0}. \qquad (31)$$

This procedure, which incorporates the non-ideality term BM_1 as well as the ideal case, can be applied to numerous cases of ideal and non-ideal association in a biological system. Its greatest advantage, beyond its wide applicability, is the relative simplicity of the computation involved compared to previously known procedures (see appendix II).

4. Basic equations for the application of $C_{(r)}$ in graphical analysis

4.1. Graphical approach to non-ideal cases

The quantities C versus r and dC/dr versus r are experimentally determined from sedimentation equilibrium measurements. From the previous section, it is apparent that when the equilibrium conditions are met the following expressions are applicable.

$$\frac{d}{dr^2}(\ln C_i + iBM_1 C) = iLM_1,$$

$$\frac{d}{dr^2}(\ln C_1 + BM_1 C) = LM_1. \qquad (32)$$

Letting $\alpha_i = C_i \exp iBM_1 C$ and $\alpha_1 = C_1 \exp BM_1 C$, eq. (32) becomes

$$\frac{d\alpha_i}{d\sigma^2} = i\alpha_i, \quad \frac{d\alpha_1}{d\sigma^2} = \alpha_1 \quad \text{and} \quad \alpha_i = \alpha_{i_0} \exp i(\sigma^2 - \sigma_0^2), \qquad (33)$$

where $\sigma^2 = LM_1 r^2$. Thus, eq. (33) may be expressed as*

$$\sum_i \frac{dC_i}{d\sigma^2} = \sum_i iC_i\left(1 - BM_1\frac{dC}{d\sigma^2}\right) = \left(1 - BM_1\frac{dC}{d\sigma^2}\right)\sum_i iC_i. \qquad (34)$$

Eq. (34), expressed in terms of C and r, may be rearranged to yield an equation formally equivalent to the Goldberg equation (3) for sedimentation equilibria data. Hence $dC/d\rho^2$ is equivalent to $M_{w\text{app}}C/M_1$.

$$\frac{d\sigma^2}{dC} = \frac{1}{\Sigma_i iC_i} + BM_1$$

* $$\frac{d\alpha_i}{\alpha_i d\sigma^2} = \frac{d\ln\alpha_i}{d\sigma^2} = \frac{d\ln C_i}{d\sigma^2} + iBM_1\frac{dC}{d\sigma^2} = i.$$

$$\text{or} \quad \frac{d\sigma^2}{d\ln C} = \frac{1}{\Sigma_i i f_i} + BM_1 C. \tag{35}$$

This expression is equivalent to the weight average molecular weight for the self-associating solute. Thus, we may define a new weight average quantity I, obtained by rectangular approximation.

$$I = \lim_{C_m \to 0} \int_{C_m}^{C_{(\sigma)}} C d\sigma^2 = \lim_{C_m \to 0} \int_{C_m}^{C_{(\sigma)}} \left(\sum_i dC_i/i + BM_1 C dC \right)$$

$$= \sum_i C_i/i + \tfrac{1}{2} BM_1 C^2. \tag{36}$$

Eq. (36) is applicable only when the meniscus C_m is depleted in the sedimentation run. If this is not the case, the quantity I must be plotted as a function of concentration and further extrapolation is required. In order to evaluate the quantity I in this equation, experiments must be performed at several concentrations.

$$I/C = \sum_i f_i/i + \tfrac{1}{2} BM_1 C. \tag{37}$$

Eq. (37) is valid where $C \geqslant C_m$. The apparent weight fraction is given by

$$f_\alpha = \alpha_1 / C$$

and $$d\ln f_\alpha = (d\ln \alpha_1 / d\ln C - 1)\, d\ln C.$$

Thus,

$$d\ln f_\alpha = \left(\frac{d\sigma^2}{d\ln C} - 1 \right) \frac{dC}{C}. \tag{38}$$

Integration of eq. (38) gives

$$\lim_{C_m \to 0} \int_{C_m}^{C} d\ln f_\alpha \equiv \ln f_\alpha, \tag{39}$$

$$f_\alpha = \alpha_1 / C = f_1 \exp BM_1 C. \tag{40}$$

Equations given to this point will suffice to evaluate the unknown quantities C_i and BM_1. Once these quantities are known, the mode of association and the equilibrium constant can be evaluated by graphical analysis as follows.

Case (i): $nP_i \rightleftarrows P_n$, $(n > 1)$

For monomer-n-mer association, the following sets of equations must be evaluated.

$$f_n = 1 - f_1, \quad \frac{d\sigma^2}{d\ln C} = \frac{1}{f_1 + n f_n} + BM_1 C,$$

$$\frac{I}{C} = f_1 + \frac{f_n}{n} + \tfrac{1}{2} BM_1 C, \quad \ln f_\alpha = \ln f_1 + BM_1 C. \tag{41}$$

Elimination of the non-ideality term from these equations yields two fundamental expressions which can be applied to any self-associating system.

$$\xi = (2I/C - d\sigma^2/d\ln C)$$

$$= 2(f_1 + f_n/n) - 1(f_1 + n f_n), \tag{42}$$

$$\eta = (d\sigma^2/d\ln C - \ln f_\alpha) = 1/(f_1 + n f_n) - \ln f_1. \tag{43}$$

After substitution of $f_n = (1 - f_1)$, eq. (42) becomes

$$\xi = 2[f_1 + (1/n)(1 - f_1)] - 1/[f_1 + n(1 - f_1)]. \tag{44}$$

Rearrangement of eq. (44) yields the quadratic equation

$$f^2 - \frac{n(2+\xi) - 2}{2(n-1)} f_1 + \frac{n(n\xi - 1)}{2(n-1)^2} = 0. \tag{45}$$

Solving for f_1 from eq. (44)

$$f_1 = \tfrac{1}{2}[1 + n\xi/2(n-1)]$$

$$\pm \tfrac{1}{2} \sqrt{[1 + n\xi/2(n-1)]^2 - 2n(n\xi - 1)(n-1)^2}.$$

At $C_m = 0$, $\xi = 1$, f_1 approaches

$$\tfrac{1}{2}[1 + n/2(n-1) \pm (n-2)/2(n-1)] = 1.$$

At $C \to \infty$, $\xi \to 1/n$, f_1 approaches

$$\tfrac{1}{2}[1 + 1/2(n-1) \pm (1 + 1/2(n-1))] = 0.$$

From eq. (45), letting $a = \tfrac{1}{2}[1 + n\xi/2(n-1)]$ and

$$b = \tfrac{1}{2} \sqrt{a^2 - 2n(n\xi - 1)/(n-1)^2}, \tag{46}$$

as C approaches zero, $f_1 = a + b$. At higher concentrations, $f_1 = a - b$. It is apparent that the value of ξ changes its sign at a point $d\xi/df_1 = 0$. Therefore,

$$\frac{2}{n} = \frac{1}{[n - (n-1) f_1]^2}$$

and

$$f_1 = \frac{n - \sqrt{n/2}}{(n-1)}.$$

where f_1 is less than 1.

$$\xi = 2(1 - \sqrt{1/2n} + 1/n)$$

$$- \frac{1}{n - (n - \sqrt{n/2})} = 2(1 - \sqrt{2/n} + 1/n). \quad (47)$$

Given values of ξ at various concentrations, f_1 is readily obtained. f_1 may also be evaluated from eq. (43) as follows

$$\frac{1}{\eta + \ln f_1} - f_1 = n f_n, \quad (48)$$

where $f_1 = f(\xi)$. From eq. (33),

$$f_n = (\alpha_n / C) e^{n\beta C} \quad \text{and} \quad f_1 = f_\alpha e^{\beta C}.$$

Thus,

$$f_1 / f_\alpha = e^{\beta C} \quad \text{and} \quad f_n = (\alpha n / C)(f_1 / f_\alpha)^n.$$

Eqs. (47) and (48) yield

$$\alpha_n = \frac{C}{n(f_1/f_\alpha)^n}\left(\frac{1}{\eta + \ln f_1} - f_1\right) \equiv \bar{Y}_n,$$

$$\alpha_{n_0} \exp n(\sigma^2 - \sigma_0^2) = \frac{C}{n(f_1/f_\alpha)^n}\left(\frac{1}{\eta + \ln f_1} - f_1\right).$$

These expressions may be combined to yield

$$\ln \bar{Y}_n = \ln \alpha_{n_0} + n(\sigma^2 - \sigma_0^2). \quad (49)$$

In brief summary, the graphical procedure for plotting $\ln \bar{Y}_n$ versus $(\sigma^2 - \sigma_0^2)$ from eq. (49) is as follows

(1) For each n, eq. (45) gives

$$f_{\ln} = a + b, \quad 1 < \xi < 2\left(1 + \frac{1}{n} - \sqrt{\frac{2}{n}}\right), \quad C < C_{\xi_{\max}},$$

$$= a - b, \quad \frac{1}{n} < \xi < 2\left(1 + \frac{1}{n} - \sqrt{\frac{2}{n}}\right), \quad C < C_{\xi_{\max}}.$$

(2) Plotting $\ln \bar{Y}_n$ versus $(\sigma^2 - \sigma_0^2)$ should yield a straight line with a slope n, and intercept $\ln \alpha_{n_0}$; then $BM_1 = \ln(f_1/f_\alpha)/C$ for each given concentration. If no straight line is obtained, $1-m$ association is not operating in the system.

(3) Equilibrium constants are evaluated from $\alpha_{1_0} = f_{\alpha_0}/C_0$ and $K_n = C_{1_0} + C_{n_0}$.

Case (ii): Indefinite association

In cases of indefinite association, in which a single equilibrium constant applies to monomer addition for all species, the basic assumptions and notation are consistent with those of Elias and Lys [9], Van Holde and Rossetti [35], and Adams and Lewis [40].

$$C = C_1/(C_1 - kC_1)^2,$$

$$\frac{d\sigma^2}{d \ln C} = \frac{1 - kC_1}{1 + kC_1} + BM_1 C, \quad (50)$$

$$I/C = (1 - kC_1) + \tfrac{1}{2} BM_1 C.$$

Elimination of the non-ideality term from these equations yields

$$\xi = 2I/C - d\sigma^2/d \ln C = 2(1 - kC_1) - (1 - kC_1)/(1 + kC_1). \quad (51)$$

Letting $kC_1 = \lambda$, then eq. (51) becomes

$$\xi = \frac{(1-\lambda)(1+2\lambda)}{(1+\lambda)}$$

or

$$2\lambda^2 + (\xi - 1)\lambda + (\xi - 1) = 0. \quad (52)$$

This quadratic equation yields

$$\lambda = \tfrac{1}{4}[-(\xi - 1) \pm \sqrt{(\xi-1)^2 - 8(\xi-1)}]. \quad (53)$$

At $C_m \to 0$, $\xi = 1$ and $\lambda = kC_1 \to 0$. In this case, eq. (53) carries a negative sign at low concentration. As the concentration increases $kC_1 \to 1$ and $\xi \to 0$, and only a positive sign of the quadratic eq. (53) will satisfy these conditions. To evaluate the apparent weight fractions, eq. (33) and λ determined from eq. (53) are combined, giving the expression

$$\ln \alpha = 2 \ln(1 - \lambda) + \ln C - 2[I/C - (1 - \lambda)], \quad (54)$$

which is equivalent to

$$\ln \alpha = \ln[\alpha_0 \exp(\sigma^2 - \sigma_0^2)] = \ln \alpha_0 + (\sigma^2 - \sigma_0^2). \quad (55)$$

Eq. (55) may be rewritten as

$$\bar{Y}_{\text{inde}} = \ln \alpha_0 + (\sigma^2 - \sigma_0^2). \quad (56)$$

$\bar{Y}_{\text{inde}}$ can be evaluated from this expression, providing that C, ξ and I/C are known. In graphical analysis, a plot of $\bar{Y}_{\text{inde}}$ versus $(\sigma^2 - \sigma_0^2)$ is constructed. If indefinite association is operating in the system, the plot should yield a straight line.

4.2. Graphical approach to ideal cases (see appendix II)

The methods outlined may be readily applied to ideal cases of self-association in which $BM_1 = 0$.

5. Enumeration of interacting species in self-associating protein systems

Eq. (13a) in section (2) is a fundamental expression in the enumeration of components by principal component and factor analysis. These methods determine in principal only linearly independent components or species. All linearly dependent species appear as one component.

5.1. Generation of the matrix from the concentration profile as a function of radial distance

After the initial sedimentation equilibrium runs have been made as described below to generate the matrix, duplicate runs are made at varying concentrations and the profiles of $C = f(r)$ recorded in tabular form, as follows:

Radial distance	r_1	r_2	r_3	...	r_l
Conc. (g/dl)	C_{11}	C_{12}	C_{13}	...	C_{1l}
Error	ϵ_{11}	ϵ_{12}	ϵ_{13}	...	ϵ_{1l}

In order to establish the error matrix, several initial runs are made at a given concentration. It is assumed that the error values determined for these initial runs remain the same for all runs at other concentrations. Indeed this is probably not the case, but determination of the error values at every concentration is not feasible. The error value for a single element is determined from the standard deviation:

$$\delta^2 = \sum_{i=1}^{n} (C_{il} - C_{nl})^2/(n-1),$$

where C_{nl} is the mean value of n measurements.

All concentration profiles are sampled over the entire length of the liquid column and at as many points as is practical (15–20 points of the radial distance are minimal).

From a series of such tables at varying concentrations, the data matrix is generated, with the number of rows being equal to the number of duplicate sample runs, and with each column representing a specific value of C_{r_n} from eq. (13a).

5.2. Setting up the matrix for principal component [36] analysis

In order to set up the matrix, eq. (13a) can be rewritten as

$$\bar{C}_{sk} = \sum_{j=1}^{l} C_{sj}\zeta_{kj}, \quad (s = 1,2,...,n; k = 1,2,...,l). \quad (57)$$

A set of $n \times l$ equations of the following matrix can be set up based on C versus r data as described above.

$$[\bar{C}_{sk}] = \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} \cdots \bar{C}_{1l} \\ \bar{C}_{21} & \bar{C}_{22} \cdots \bar{C}_{2l} \\ \vdots & \vdots & \vdots \\ \bar{C}_{nl} & \bar{C}_{n2} \cdots \bar{C}_{nl} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & R_{12} & R_{13} \cdots R_{1l} \\ R_{21} & 1 & R_{23} \cdots R_{2l} \\ \vdots & \vdots & 1 \quad \cdots R_{3l} \\ \vdots & \vdots & \vdots \quad \vdots \\ R_{l1} & R_{l2} & R_{l3} \cdots 1 \end{bmatrix}. \quad (58)$$

A simplified procedure based on the $n \times l$ matrix follows, to demonstrate the relative ease with which it can be used to enumerate components in a system using a Biomed computer program (03M, by Dixon, UCLA) [39]. $[\bar{C}_{sk}]$ is converted to the correlation matrix, $[R_{sk}]$, which determines the correlation between $\bar{C}_{.s}$ and $\bar{C}_{.k}$ where the first subscript denotes an observation. Let λ_1 be the largest eigenvalue of $[R_{sk}]$ between $\bar{C}_{.s}$ and $\bar{C}_{.k}$. This correlation matrix is replaced with the eigen matrix which follows.

$$\begin{bmatrix} S_{11} & R_{12} & R_{13} \cdots R_{1l} \\ R_{21} & S_{22} & R_{23} \cdots R_{2l} \\ \vdots & \vdots & S_{33} \cdots R_{3l} \\ \vdots & \vdots & \vdots \quad \vdots \\ R_{l1} & R_{l2} & R_{l3} \cdots S_{ll} \end{bmatrix} \quad (59)$$

Diagonal elements of this matrix, S_{ll}, are the square multiple correlations of $\bar{C}_{\cdot j}$ with $\bar{C}_{\cdot 1} \ldots \bar{C}_{\cdot l}$. Let $\theta_1, \theta_2, \ldots, \theta_l$ be the eigenvalues and $\alpha_1, \alpha_2, \ldots, \alpha_l$ be the associated eigenvectors of this matrix. Although eigenvalues may be negative or positive, only positive values are considered. The number of components will be equivalent to the number of eigenvalues greater than unity accounting for most of the total.

Enumeration of components by principal component analysis is a simple matter, and after considering statistical significance and criteria used in the analysis, it is apparent that the eigenvalue is one of the best tools available for delineating the number of components present in a system [28,29,37]. Perhaps the biggest drawback of this simple criterion is the problem posed when eigenvalues are close to unity. For example, it does not appear altogether proper to retain an eigenvalue of 1.00 and drop a subsequent one of 0.97, or to consider that an eigenvalue of 0.98, for example, does not belong to a component or factor contributing significantly to the result. In such instances, other empirical criteria or perhaps rigorous statistical tests may be useful to supplement the analysis. When principal component and factor analysis are used, these methods determine in principal only linearly independent components. All linearly dependent components appear as a single component.

Acknowledgement

The authors wish to express their thanks to Dr. Magar E. Magar for his critical review and suggestions in the preparation of this manuscript.

Appendix I: Non-linear least square analysis for evaluation of BM_1 and C_{ir_0}

In order to compute $C = f(r)$, eq. (10) becomes

$$C_j = \sum_{i=1}^{n} C_{ir_0} \exp i[LM_1(r_j^2 - r_0^2) - BM_1(C - C_{r_0})] + \delta_j, \quad (j = 1, \ldots, m), \qquad \text{(A.1)}$$

where δ_j, the experimental error in making a measurement, can be written as

$$\delta_j = C_j - \sum_{i=1}^{n} C_{ir_0} \exp i[LM_1(r_j^2 - r_0^2) - BM_1(C - C_{r_0})]. \qquad \text{(A.2)}$$

As noted from eq. (A.2), we are trying to determine C_{ir_0} and BM_1 so as to minimize the sums of the square of the experimental error function, i.e.,

$$\Phi(C_{ir_0}, C_{2r_0}, \ldots, C_{nr_0}, BM_1) = \sum_{j=1}^{m} \left\{ C_j - \sum_{i=1}^{n} C_{ir_0} \exp i[LM_1(r_j^2 - r_0^2) - BM_1(C - C_{r_0})] \right\}^2 \qquad \text{(A.3)}$$

or

$$\Phi(C_{1r_0}, C_{2r_0}, \ldots, C_{nr_0}, BM_1) = \sum_{j=1}^{m} \delta_j^2. \qquad \text{(A.4)}$$

The most common procedure for non-linear least square analysis is based on a modified Gauss–Newton method (BMD X85, UCLA, No.3, Biomedical computer program of Dixon, 1969). In order to compute eq. (10) in which C_{ir_0} and BM_1 are unknown, the concentration as a function of radial distance, initial values of C_{ir_0} and BM_1 must be assigned. Once the initial estimates have been put in, the computer program makes continuous changes in the parameters, with each iteration resulting in successively smaller mean square deviations of the observed C_i value of eq. (10) with respect to $f(r)$. The procedure is continued until the mean square deviation is minimized.

Appendix II: Graphical approach to ideal cases

Case (i): Monomer-n-mer association

(1) A plot of $\ln(C - C_1)$ versus $(\sigma^2 - \sigma_0^2)$ will give a straight line with a slope of n. $\ln C_{n_0}$ is evaluated at the intercept, where $\sigma = \sigma_0$.

(2) A plot of $\ln(dC/d\sigma^2 - C)$ versus $(\sigma^2 - \sigma_0^2)$ gives a slope of n. The intercept will yield $\ln(n-1)C_{n_0}$; thus $K_n = C_{n_0}/(C_{1_0})^n$.

Case (ii): Indefinite association

Plotting $\ln(1 - I/C)$ versus $(\sigma^2 - \sigma_0^2)$ should yield a straight line with a slope of kC_{1_0}. C_{1_0} may be directly evaluated from

$$C_1 = C \exp \int_0^C (d\sigma^2/d \ln C - 1) C^{-1} dC.$$

Case (iii): Three species association

(1) A plot of $\ln \bar{Y}_m = \ln(n-m)C_{n_0} + n(\sigma^2 - \sigma_0^2)$ versus $(\sigma^2 - \sigma_0^2)$ will yield a straight line, where m is assumed and n is obtained from the slope. Thus, $\ln(n-m)C_{n_0}$ may be evaluated at the intercept.

(2) $\bar{Y}_2 = (dC/d\sigma^2) - 2C - C_1$ for 1–2–n association, $\bar{Y}_3 = (dC/d\sigma^2) - 3C + 2C_1$ for 1–3–n association, etc. $K_m = C_{m_0}/C_1^m$, $K_n = C_{n_0}/C_{1_0}^n$.

References

[1] R.F. Steiner, Arch. Biochem. Biophys. 44 (1953) 120; 39 (1952) 33.
[2] R.F. Steiner, Arch. Biochem. Biophys. 47 (1953) 56.
[3] R.J. Goldberg, J. Phys. Chem. 57 (1953) 194.
[4] M.S.N. Rao and G. Kegeles, J. Am. Chem. Soc. 80 (1958) 5724.
[5] E.T. Adams, Jr. and H. Fujita, in: Ultracentrifugal analysis in theory and experiment, ed. J.W. Williams (Academic Press, New York, 1963).
[6] P.G. Squire and C.H. Li, J. Am. Chem. Soc. 83 (1961) 3521.
[7] E.T. Adams, Jr., Biochemistry 4 (1965) 1646.
[8] E.T. Adams, Jr., Biochemistry 4 (1965) 1966.
[9] H.G. Elias and H. Lys, Makromol. Chem. 96 (1966) 64.
[10] A.J. Sophianopoulos and K.E. Van Holde, J. Biol. Chem. 239 (1964) 2516.
[11] See Ann. N.Y. Acad. Sci. (1968) in: Advances in ultracentrigual analysis, ed. D.A. Yphantis, Vol. 164, Part 1, pp. 1–305.
[12] P.O.P. Ts'o and S.I. Chan, J. Am. Chem. Soc. 86 (1969) 4176.
[13] R.F. Steiner, Biochemistry 7 (1968) 2201.
[14] P.W. Chun and S.J. Kim, Biochemistry 8 (1969) 1633.
[15] P.W. Chun and S.J. Kim, J. Phys. Chem. 74 (1970) 899.
[16] P.W. Chun, S.J. Kim, G.D. Williams, W.T. Cope, L.H. Tang and E.T. Adams, Jr., Biopolymers 11 (1972) 197.
[17] D.E. Roark and D.A. Yphantis, Biochemistry 10 (1971) 3241.
[18] D.A. Yphantis and D.E. Roark, Biochemistry 11 (1972) 2925.
[19] H.K. Schachman, Ultracentrifugation in biochemistry (Academic Press, New York, 1959).
[20] R.H. Hashmeyer and W.F. Bower, Biochemistry 9 (1970) 435.
[21] D.C. Teller, T.A. Horbett, E.G. Richard and H.K. Schachman, Ann. N.Y. Acad. Sci. 164 (1969) 66.
[22] R.M. Wallace, J. Phys. Chem. 64 (1960) 899.
[23] R.M. Wallace and S.M. Katz, J. Phys. Chem. 68 (1964) 3890.
[24] G. Weber, Nature (London) 190 (1961) 27.
[25] D.W. McMullen, S.R. Jaskunas and J. Tinoco, Jr., Biopolymers 5 (1967) 589.
[26] G.K. Ackers, J. Biol. Chem. 243 (1968) 2056.
[27] W. Godschalk, Biochemistry 10 (1971) 3284.
[28] M.E. Magar, Biopolymers 11 (1972) 2187.
[29] M.E. Magar and P.W. Chun, Biophys. Chem. 1 (1973) 18.
[30] T. Svendberg and K.O. Pedersen, Ultracentrifuge (Oxford University Press, London, 1940).
[31] E.F. Casassa and H. Eisenberg, Advan. Prot. Chem. 19 (1964) 287.
[32] D.A. Yphantis, Biochemistry 3 (1964) 297.
[33] E.T. Adams, Jr. and J.W. Williams, J. Am. Chem. Soc. 86 (1964) 3454.
[34] E.G. Richard and H.K. Schachman, J. Phys. Chem. 63 (1959) 1578.
[35] K.E. Van Holde and G.P. Rossetti, Biochemistry 6 (1967) 2189.
[36] H.H. Harman, Modern factor analysis (University of Chicago Press, Chicago, 1967).
[37] P.W. Chun, W.P. Herschleb, D.J. Downing and M.L. Krista, Biophys. Chem. 1 (1973/74) 141.
[38] P.W. Chun and S.J. Kim, Biochemistry 9 (1970) 1957.
[39] M. Dixon, Biomed Program (1969) BMD 03M, UCLA.
[40] E.T. Adams, Jr. and M.S. Lewis, Biochemistry 7 (1968) 1044.