The Investigation of Self-Association Reactions by Equilibrium Ultracentrifugation

HYOUNGMAN KIM,[†] RICHARD C. DEONIER,[‡] and J. W. WILLIAMS*

Departments of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, and University of Southern California, Los Angeles, California 90007

Received January 10, 1977

Contents

[†] University of Wisconsin—Madison.
[‡] University of Southern California.
* Author to whom correspondence should be addressed at University of Wisconsin—Madison.

I. Introduction

A. History

The quantitative study of interactions between molecules, alike or unlike, is difficult at best; this becomes increasingly true as the reactants are larger in size and complexity. The available techniques are usually transport methods or equilibrium methods. Of the former, those most commonly used are sedimentation transport, electrophoresis, and chromatography. Employed largely with rapidly equilibrating systems, their application has been considered at some length in recent monographs by Cann[1] and by Nichol and Winzor.[2] Quantitative information about the stoichiometry and equilibrium constant of self-association reactions is here obtained from the concentration dependence of the weight- and z-average velocities or elution volumes.

Actually, the recognition that many proteins are made up of subunits, and that some may self-associate, has depended largely upon accurate molecular weight determinations. The subject may be said to have begun with the ultracentrifugal analysis of hemoglobin and some of the hemocyanins in the 1920's,[3] the osmotic pressure investigations of hemoglobin,[4] and the conclusions of Moody[5] and Gutfreund[6] about the subunit structure of insulin, first from experiments with the ultracentrifuge and then with the osmometer.

However, these early researches were not designed to provide information about the various reaction mechanisms and corresponding energy relationships; this phase of the subject began with Tiselius.[7] His contribution was a theoretical thermodynamic treatment of rapidly reversible self-association reactions at sedimentation equilibrium; the analysis was restricted to monomer–dimer self-associations which take place in thermodynamically ideal systems.

As far as we are aware there was no attempt to broaden his teachings and to expand the subject until Squire and Li[8] performed sedimentation equilibrium experiments to elucidate the state of aggregation of ovine pituitary ACTH in acid and base solutions. Shortly thereafter, the general subject began a period

of rapid elaboration. "The development of methods to study a wide variety of systems was pioneered by Adams, Fujita, and Williams and the theory has been extensively expanded by Adams and co-workers." [9] In more recent years both theory and practice have been augmented in a number of more sophisticated ways. In this account it is some salient features of a subject as it now stands (1976) that we have attempted to portray.

Associating systems are conveniently divided into self-association systems (e.g., monomer–dimer equilibrium of a simple species; association of a monomer into a number of higher molecular weight polymeric aggregates) and mixed associations. As a class the latter group contains some of the most interesting cases, such as the formation of RNA polymerase, the reaction of antigen and antibody, and the combination of enzyme with substrate. However, the former group is not without its own significant examples, for instance, *lac* repressor, λ repressor, T4-gene 32 protein, hemocyanin structures, and many others. Such systems are far easier to treat in a theoretical way, and this report is limited to considerations of their analysis.

With certain simplifying assumptions, it has become possible to make use of observations of the combined sedimentation and chemical equilibrium data for a self-associating system to determine a stoichiometry and compute the reaction association constant for the reaction which is involved, provided that there is no volume change in the reaction during the ultracentrifugation process (or provided the molar volume change is known and suitable correction has been made).

In general a self-association may be suspected when the apparent molecular weights for the solute increase as a continuous function of increasing concentration for several different initial concentrations; it is an increase which can be readily differentiated from the behavior in the ultracentrifuge cell of a polydisperse system such as an organic high polymer. In some cases the form of apparent average molecular weight vs. concentration curve gives a clue to the actual reaction mechanism, while in others it becomes necessary to select a model which requires the specification of the number of parameters and to compute them for comparison with the experimental observations. It is important not to select more parameters for representing the data than are justified by the experimental errors, after the solution nonideality has been adequately taken into account. A priori knowledge of the mechanism in lieu of model is, of course, highly advantageous for the evaluation of the association constants.

B. Plan for the Text

The study of the self-association processes between molecules has long been a popular one, so that in an article of this type there must be a restriction in coverage. Indeed, one could hardly make reference to all of the literature articles which describe some phase of the subject. A considerable portion of the recent progress, especially as it pertains to molecules of biological interest, has been achieved through the use of the sedimentation equilibrium experiment, and our remarks will be restricted to a description, both of the fundamental theory and its application in the continuing growth of knowledge of the mechanisms of these reactions. Even with this restriction it is difficult to avoid some arbitrariness in the choice of material.

There are several logical routes by which one may outline and describe the basic mathematical theory. We have selected what appears to us to be the most general route to this end, arranging matters so that working equations for the interpretation of the data in monomer–dimer, monomer–dimer–trimer . . . , monomer–*n*-mer, and isodesmic reaction processes can be taken from a single overall mathematical statement. This plan makes it unnecessary to provide an isolated description for the isodesmic case, which has been a common procedure. For the actual analysis of the data section, however, we shall write more in terms of a methodological argument, separating the description of the procedures into diagnostic plots and analytical treatments of both the ideal and nonideal systems.

Following this introductory chapter there is presented a description of the fundamental theory of the sedimentation equilibrium itself and of the necessary modifications required for the application to the self-association reaction. The third chapter provides information about the various methods by which the data for the combined sedimentation and chemical equilibrium are analyzed. A description of the complications and of the auxiliary data forms the subject matter of section IV. It is followed by a record of tests, with actual experimental data, of the several computational methods which have been described. In concluding chapters the question of the thermodynamic interpretation of the equilibrium constants is briefly treated and some general interpretations are attempted.

C. List of Symbols

c	concentration on volume-based scales, usually g/100 mL, but also g/mL
c_0	original concentration in cell
c_i	concentration of the ith component
c_n	concentration of oligomer
c_1	concentration of monomer
c_r	total concentration at radial distance r
c_a	concentration of equilibrium mixture at meniscus
c_b	concentration of equilibrium mixture at cell bottom
C_i	concentration, mole basis, of species i
f_i	weight fraction of species i in the cell
g_i	amount of constituent i in cell
h'	cell thickness, parallel to the axis of revolution
k	self-association constant, isodesmic reaction
k_i	self-association constant, dimerization, trimerization, etc. (units: dL/g).
n	refractive index
r	arbitrary radial distance along the cell column
r_a	radial position at the meniscus (or a)
r_b	radial position at the bottom of the cell (or b)
r_d, r_e, r_0	arbitrary cell positions
t	time
$\bar{v}$	partial specific volume of a solute in the solution
y_i	activity coefficient of the ith species
z	valence of an ion or macroion
A	factor $(1 - \bar{v}\rho)\omega^2/RT$
A'	factor $(1 - \bar{v}\rho)\omega^2/2RT$
A_i	species containing i monomers: ($i = 1, \ldots, n$)
B	second virial coefficient (or B_1, etc.)
G	Gibbs free energy per mole; ΔG_i° = standard Gibbs free energy change, species i
I	ionic strength
K	self-association constant, isodesmic reaction (units: liter/mole)
$K_{n,m}$	self-association constant for discrete reactions, molar basis
M	molecular weight
M_1	monomer molecular weight
M_2, M_3, . . . , M_n, etc.	molecular weight of dimer, trimer, . . . , n-mer

M_n number-average molecular weight over the cell
M_w weight-average molecular weight over the cell
M_z z-average molecular weight over the cell
$M_{n(c)}$, $M_{w(c)}$, $M_{z(c)}$, etc. point average molecular weights corresponding to a radial distance where the concentration is c, ideal solutions
$M^a_{n(c)}$, $M^a_{w(c)}$, $M^a_{z(c)}$, etc. apparent values of same in nonideal solutions
P pressure
R molar gas constant; refractive index increment
R_a $M^a_{w(c)}/M_1$
V cell volume
$\bar{V}$ partial molal volume
δ small increment
η defined function for extension of two-species plot to nonideal systems
λ $(1 - \bar{v}_2\rho)\omega^2(r_b^2 - r_a^2)/2RT$
$\hat{\mu}_i$ chemical potential, per mole of the ith species
μ_i chemical potential, weight per volume scale
$\bar{\mu}_i$ total potential, weight per volume scale
ρ density of solution, g/cm^3
ρ_1 density of solvent
ρ_0 density of dialyzed solvent
$(d\rho/dc)_\mu$ density increment at fixed chemical potential of diffusible compounds
Φ constant of integration
ψ Adams computation function
ω angular velocity, radians per second
ϵ defined function for extension of two-species plot to nonideal system
θ sector angle, ultracentrifuge cell
Π product
Σ sum
χ^2 measure of goodness of fit, Visser et al.
Ω Milthorpe et al. function

II. Thermodynamic Preconsiderations: Basic Equations

A. Thermodynamics of Self-Association

1. Equations Describing Chemical Equilibrium

In general, a self-association reaction may be expressed in either of two equivalent forms:

$$\begin{aligned} A_1 + A_1 &\rightleftharpoons A_2 \\ A_1 + A_2 &\rightleftharpoons A_3 \\ &\vdots \\ A_1 + A_{(i-1)} &\rightleftharpoons A_i \\ &\vdots \\ A_1 + A_{n-1} &\rightleftharpoons A_n \end{aligned} \qquad \text{(II-1)}$$

or

$$\begin{aligned} 2A_1 &\rightleftharpoons A_2 \\ 3A_1 &\rightleftharpoons A_3 \\ &\vdots \\ iA_1 &\quad A_i \\ &\vdots \\ nA_1 &\rightleftharpoons A_n \end{aligned} \qquad \text{(II-2)}$$

Herein we have not specified how many species are in equilibrium with each other, nor need the upper limit, n, be finite. In these outlines A_1 represents the monomer, A_2 is the dimer, etc. At present, the term "monomer" is taken to mean the lowest molecular weight species which participates in the equilibria under consideration; however, there are additional considerations, which are discussed in section IV.

Corresponding to the two ways of expressing the association reactions, there are two relationships relating the chemical potentials of the species in equilibrium. For reactions written as in (II-1), one has

$$\hat{\mu}_1 + \hat{\mu}_{i-1} = \hat{\mu}_i \qquad \text{(II-3)}$$

and for reactions expressed as in (II-2), the relationship is

$$i\hat{\mu}_1 = \hat{\mu}_i \qquad \text{(II-4)}$$

The quantities $\hat{\mu}_i$ are the chemical potentials per mole of the ith species, and they are related to the molar concentrations C_i by the expression

$$\hat{\mu}_i = \hat{\mu}_i^0 + RT \ln a_i = \hat{\mu}_i^0 + RT \ln y_i C_i \qquad \text{(II-5)}$$

Here $\hat{\mu}_i^0$ is the chemical potential per mole of the ith species in its standard state, R is the gas constant, T is the absolute temperature, and y_i is the activity coefficient.

The free energy change for the ith reaction listed by eq II-1 is given by

$$-\Delta G_i^{\circ\prime} = \hat{\mu}_1^0 + \hat{\mu}_{(i-1)}^0 - \hat{\mu}_i^0 = RT \ln \frac{C_i}{C_1 C_{(i-1)}} \frac{y_i}{y_1 y_{(i-1)}} \qquad \text{(II-6)}$$

where $\Delta G_i^{\circ\prime}$ is the standard free energy increment for *association*.

Now the equilibrium constant, $K_{(i-1)i}$, for this reaction is defined as

$$K_{(i-1)i} = \frac{C_i}{C_1 C_{(i-1)}} \frac{y_i}{y_1 y_{(i-1)}} \qquad \text{(II-7)}$$

and thus

$$-\Delta G_i^{\circ\prime} = RT \ln K_{(i-1)i} \qquad \text{(II-8)}$$

Corresponding expressions can be obtained when the alternative description (eq II-2) is used:

$$-\Delta G_i^\circ = RT \ln (C_i y_i / C_1^i y_1^i) \qquad \text{(II-9)}$$

or

$$-\Delta G_i^\circ = RT \ln K_i \qquad \text{(II-10)}$$

where ΔG_i° is the standard Gibbs free energy change of association for (II-2). The equilibrium constant for this description of the self-association now takes the form

$$K_i = C_i y_i / C_1^i y_1^i \qquad \text{(II-11)}$$

Comparison of eq II-7 and II-11 indicates that

$$K_i = K_{12} K_{23} \cdots K_{(i-1)i} \qquad \text{(II-12)}$$

Since the sets II-1 and II-2 are equivalent descriptions of the equilibrium system, stipulation of either the set of $K_{(i-1)i}$ or the set of K_i suffices to describe the equilibrium.

Often the concentration of solutes in the ultracentrifuge cell is determined on a mass per volume basis rather than on the molar basis, and it is useful to have expressions in this concentration scale. The development given so far may be repeated for this case, and we obtain the equilibrium constants for reaction 1 (eq II-1).

$$k_{(i-1)i} = \frac{c_i}{c_1 c_{(i-1)}} \frac{y_i}{y_1 y_{(i-1)}} \qquad \text{(II-13)}$$

and for reaction 2 (eq II-2)

$$k_i = c_i y_i / c_1^i y_1^i \qquad \text{(II-14)}$$

In these expressions c_i is the concentration of ith species in g/dL. We may then write

$$k_i = k_{12} k_{23} \ldots k_{(i-1)i} \qquad \text{(II-15)}$$

The relation between K_i and k_i is given by

$$K_i = k_i \left(\frac{M_1}{i \cdot 10} \right)^{i-1} \qquad \text{(II-16)}$$

The determination of the equilibrium (or association) constants by using eq II-14 is very difficult because of the activity coefficient, y_i, which is concentration dependent. If we restrict ourselves to dilute solutions of neutral species ln y_i (i.e., the logarithm of the activity of species i—not of the component) can be expanded[10-12] in a power series in terms of total concentration, c. Making use only of the first of these terms, we have

$$\ln y_i = B_i M_i c \qquad \text{(II-17)}$$

where M_i is the molecular weight of species i, and B_i is the second virial coefficient for that species. If it is assumed that there is a single virial coefficient common to all species

$$B_i = B \qquad (i = 1, 2, \ldots) \qquad \text{(II-17a)}$$

then one obtains

$$\ln y_i = iBM_1 c \qquad \text{(II-18)}$$

This is really an approximation which has come into widespread use because it leads to computational convenience while still providing an adequate description of the solution nonideality. Ogston and Winzor[13] have examined the validity of this procedure at some length and find it justified.

In the treatment of the data presently available, it has been possible to account for the experimental observations in any given situation in terms of either equilibrium constants alone (k_i) or with sets of two quantities: equilibrium constants and a second virial coefficient (k_i, BM_1). It is true that with inclusion of a third virial coefficient in nonideal situations a better fit may be achieved; in other words, whether the concurrence is considered to be sufficient depends upon the criteria used. In the interest of simplification of the equations we shall ordinarily omit any term in the third virial coefficient. When eq II-18 is introduced into eq II-14

$$k_i = \frac{c_i \exp[M_1(iBc + \ldots)]}{c_1^i \exp[iM_1(Bc + \ldots)]} = c_i / c_1^i \qquad \text{(II-19)}$$

Thus, in so far as ln y_i can be expressed in terms of virial coefficients as indicated in eq II-18, the equilibrium constants become simple ratios of concentrations. Likewise

$$K_i = C_i / C_1^i \qquad \text{(II-20)}$$

2. *Relations between the Equilibrium Constants, the Total Concentration, and Weight-Average Molecular Weight*

The primary data from the analytical ultracentrifuge are directly related to the concentration, c, as a function of radius, r. The molecular weight average most easily calculated from these data is the weight-average molecular weight $M_{w(c)}$, where it is explicitly indicated that the weight-average molecular weight will be concentration dependent. When convention II-2 is used, from eq II-19, the concentration can be written as a power series in the variable c_1, the concentration of monomer:

$$c = \sum_{i=1}^{n} k_i c_1^i \qquad \text{(II-21)}$$

where we have made the formal definition $k_1 = 1$. The upper limit n need not be finite.

The weight-average molecular weight is defined

$$M_{w(c)} = \sum_{i=1}^{n} \frac{c_i M_i}{c} \qquad \text{(II-22)}$$

which can be combined with eq II-19 to yield

$$\frac{cM_{w(c)}}{M_1} = \sum_{i=1}^{n} i k_i c_1^i \qquad \text{(II-23)}$$

Equation II-23 is like II-21, except for the coefficient of the terms c_1^i. Both of these expressions provide a general description for $cM_{w(c)}$ for any self-association. There are special cases to be noted, and in terms of eq II-21, they can be stated in terms of conditions on the k_i.

a. $k_i = 0$ for all $i > 1$ except when $i = n$. The resulting expression for c contains two terms, one for monomer and one for n-mer. This is the monomer–n-mer case, which includes the monomer–dimer when $n = 2$.

b. $k_i > 0$ when $1 < i < m$; $k_i = 0$ when $i > m$, where m is some integer. This is the condition for the discrete self-associations—those which have a finite number of species in equilibrium with each other. The monomer–dimer reaction is a special case of this one. The monomer–dimer–trimer would be the simplest example involving more than two species.

c. Returning to the mole per liter scale, and using eq II-1, we could consider the case

$$K = K_{12} = K_{23} = \ldots = K_{(i-1)i} \qquad \text{(II-24)}$$

Such a condition implies a constant free energy increment for each successive step of the association reaction (see eq II-8). This is the indefinite or isodesmic self-association.[14-17] If the monomer has two independent sites for the association, the reaction may proceed indefinitely, leading to very large linear or helical aggregates depending on the relative orientation of these sites. Thermodynamically, this situation corresponds to equal increments of free energy for the addition of one more monomer to any i-mer. The principle of equal reactivity, independent of molecular size, is again involved, just as it was by Kuhn and Flory in their analyses of random degradations and polymerizations, respectively, for the organic high polymer systems.

When eq II-24 applies, it is easy to see that

$$C_i = C_1^i K^{(i-1)} \qquad \text{(II-25)}$$

When the concentrations are expressed in g/dL, eq II-25 becomes

$$c_i = i k^{(i-1)} c_1^i \qquad \text{(II-26)}$$

The quantity $k = 10K/M_1$ is called the intrinsic equilibrium constant. The above expression states that for the indefinite or isodesmic case, $k_i = ik^{(i-1)}$; eq II-21 then is written

$$c = c_1 + 2kc_1^2 + 3k^2 c_1^3 + \ldots \qquad \text{(II-27)}$$

If $kc_1 > 1$, eq II-27 is a diverging function which favors infinitely large aggregates. This will lead to precipitation of the aggregates, as shown by glutamic dehydrogenase at high concentrations.[18] Thus our interest is only with the case where $kc_1 < 1$ and under this condition eq II-27 takes the form

$$c = c_1 / (1 - kc_1)^2 \qquad \text{(II-28)}$$

B. Self-Association Reactions in the Ultracentrifuge: Sedimentation Equilibrium

In the previous section we have laid the thermodynamic groundwork and have related the equilibrium constants to the concentration and the true, weight-average molecular weight. The solution nonideality was represented in a form which cancels

in the relations for the equilibrium constants of the various species (see eq II-19). In this section we consider chemical equilibrium in the presence of the applied centrifugal field, and obtain relationships between the observed quantities (which allow calculation of the apparent weight-average molecular weight, $M^a_{w(c)}$ and the true weight-average molecular weight, $M_{w(c)}$). It will be seen that solution nonideality still must be taken into account, even when the nonideality terms in II-19 cancel.

The basic set of conditions for the sedimentation equilibrium in the ultracentrifuge is given by[19-21]

$$d\hat{\mu}_i - M_i\omega^2 r\, dr = 0 \qquad \text{(II-29)}$$

where ω is the angular velocity and r is the radial distance from the center of the rotor. Here we assume that the self-associating material has been thoroughly dialyzed against simple electrolyte solution and that the partial specific volume and refractive index derivatives have been measured according to the methods suggested by Casassa and Eisenberg[22] so that the system answers the description pseudo-two-component solution. It is also assumed that the rotor speed is sufficiently low that no effective sedimentation of the electrolytes takes place during the performance of the experiments.

When the chemical potential of species i is expressed on the gram basis, we have

$$\mu_i = \mu_i^0 + (RT/M_1) \ln y_i c_i \qquad \text{(II-30)}$$

With it, the condition for sedimentation equilibrium is now written as

$$d\mu_i - \omega^2 r\, dr = 0 \qquad \text{(II-31)}$$

For the monomer species eq II-31 becomes

$$d\mu_1 = \omega^2 r\, dr \qquad \text{(II-32)}$$

As indicated by eq II-3 and II-4, the chemical potentials of all other species are either related to the monomer or to each other. Following the analysis of Adams and Fujita,[10] we obtain

$$\left(\frac{\partial\mu_1}{\partial c_1}\right)_{P,T}\frac{dc_1}{dr} + \left(\frac{\partial\mu_1}{\partial P}\right)_{T,c}\frac{dP}{dr} = \omega^2 r \qquad \text{(II-33)}$$

where P is the pressure. From this equation

$$\left[\sum_{i=1}^{n}\left(\frac{\partial\mu_1}{\partial c_i}\right)_{P,T,c_j}\frac{dc_i}{dc_1}\right]\frac{dc_1}{dr} = \omega^2 r(1 - \bar{v}\rho) \quad (i \neq j) \qquad \text{(II-34)}$$

in which $\bar{v}$ is the partial specific volume of the solute and ρ is the density of the solution. It is generally assumed that these quantities are independent of concentration. By the combination of eq II-18, eq II-30, and eq II-34 there is obtained

$$\left[(1 + c_1(BM_1 + \ldots)\sum_{i=1}^{n}\frac{dc_i}{dc_1}\right]\frac{dc_1}{dr} = AM_1c_1r \qquad \text{(II-35)}$$

where

$$A = \omega^2(1 - \bar{v}\rho)/RT \qquad \text{(II-36)}$$

or

$$\frac{d\ln c_1}{dr} + (BM_1 + \ldots)\frac{dc}{dr} = AM_1 r \qquad \text{(II-37)}$$

The apparent weight-average molecular weight is defined by the statement

$$M^a_{w(c)} = (1/Acr)(dc/dr) \qquad \text{(II-38)}$$

thus, from eq II-37 and eq II-38

$$\frac{M_1}{M^a_{w(c)}} = \frac{d\ln c_1}{d\ln c} + BM_1c + \ldots \qquad \text{(II-39)}$$

From eq II-21 one can obtain

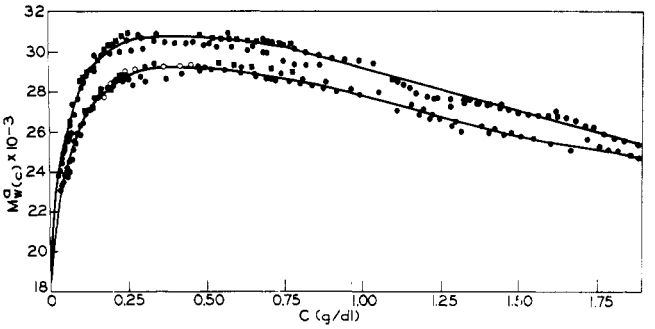

Figure 1. The apparent weight-average molecular weight of β-lactoglobulin B in solution at pH 2.64, $I = 0.16$, plotted against concentration. Upper curve, at 15 °C; lower curve, at 25 °C. Reproduced from Visser et al.[23]

$$dc/dc_1 = \sum_{i=1}^{n} ik_ic_1^{i-1} \qquad \text{(II-40)}$$

And from the definition of weight-average molecular weight

$$M_{w(c)} = \sum_{i=1}^{n} c_iM_i/c = \sum_{i=1}^{n} ik_ic_1^iM_1 \Big/ \sum_{i=1}^{n} k_ic_1^i \qquad \text{(II-41)}$$

we write

$$M_1/M_{w(c)} = c\, dc_1/c_1\, dc = d\ln c_1/d\ln c \qquad \text{(II-42)}$$

Introduction of eq II-42 into eq II-39 gives

$$\frac{M_1}{M^a_{w(c)}} = \frac{M_1}{M_{w(c)}} + BM_1c + \ldots \qquad \text{(II-43)}$$

As has been demonstrated here eq II-43 is a general expression, regardless of the number of species in equilibrium. When applied to the ideal isodesmic reaction, eq II-28 and II-42 now give

$$M_1/M_{w(c)} = (1 - kc_1)/(1 + kc_1) \quad (kc < 1) \qquad \text{(II-44)}$$

Since $M_{w(c)}$ is a function of total concentration only, it follows that $M^a_{w(c)}$ is a unique function of c for every type of stoichiometry. This, in turn, means that the $M^a_{w(c)}$ vs. c curves from different sedimentation equilibrium experiments, obtained by using different initial concentrations and rotor speeds, should lie on a single smooth curve. Such behavior is clearly indicated by the data for the self-association of β-lactoglobulin B in acid solution obtained by Visser et al.[23] (cf. Figure 1).

The uniqueness of this situation arises from the nature of the chemical reaction; the amount of each species in a given self-association system is determined only by the total concentration if T is constant and the equilibrium is not sensitive to pressure. It should be noticed that this behavior is different from that of polydisperse systems such as the synthetic organic high polymers. Although the expression for the apparent molecular weight for such systems is formally similar to that which describes the equilibrium behavior in the self-associating systems (see, for example, eq 12-1 of Chapter V, ref 21), this equation contains terms for the initial concentration and the rotor speed. It is also of the same form as the equation which gives the apparent weight average for the homogeneous, nonideal system. Thus, the apparent molecular weight of a polydisperse system at a given value of c depends on the initial loading concentration as well as on the angular velocity of rotor.

Since eq II-43 depends only on total concentration, other methods which provide apparent weight-average molecular weight vs. c data can be analyzed in the same way. That this is the case may be seen from the expression descriptive of the light scattering (eq 4 of Townend and Timasheff[24]). A single sedimentation equilibrium experiment is thus equivalent to a large number of light-scattering experiments, each with different initial concentrations.

C. Quantities Derived from Experimental Data for $M^a_{w(c)}$

Several additional quantities descriptive of self-associating systems may be derived from the $M^a_{w(c)}$ vs. c data. We start with the apparent number-average molecular weights, $M^a_{n(c)}$.[25] From eq II-43

$$\int_0^c (M_1/M^a_{w(c)})\,dc = \int_0^c (M_1/M_{w(c)})\,dc + BM_1c^2/2 + \ldots \quad \text{(II-45)}$$

From eq II-21 and II-42

$$\int_0^c (M_1/M_{w(c)})\,dc = \int_0^c (c/c_1)\,dc_1 = \sum_{i=1}^{n} (k_i c_1^i/i) \quad \text{(II-46)}$$

The number-average molecular weight, $M_{n(c)}$, is defined as

$$M_{n(c)} = c \Big/ \sum_{i=1}^{n} (c_i/M_i) = M_1 c \Big/ \sum_{i=1}^{n} (k_i c_1^i/i) \quad \text{(II-47)}$$

and

$$cM_1/M_{n(c)} = \sum_{i=1}^{n} (k_i {c_1}^i/i) \quad \text{(II-48)}$$

From eq II-45, II-46, and II-48, one then obtains

$$(1/c)\int_0^c (M_1/M^a_{w(c)})\,dc = M_1/M_{n(c)} + BM_1c/2 + \ldots \quad \text{(II-49)}$$

If we define

$$M_1/M^a_{n(c)} = (1/c)\int_0^c [M_1/M^a_{w(c)}]\,dc \quad \text{(II-50)}$$

then from eq II-49, we have the desired result.[25]

$$M_1/M^a_{n(c)} = M_1/M_{n(c)} + BM_1c/2 + \ldots \quad \text{(II-51)}$$

For the isodesmic reactions, combination of eq II-44, II-46, and II-48 gives

$$M_1/M_{n(c)} = 1 - kc_1 \quad \text{(II-52)}$$

The z-average molecular weight for an ideal system, $M_{z(c)}$, is defined as

$$M_{z(c)} = \sum_{i=1}^{n} c_i M_i^2 \Big/ \sum_{i=1}^{n} c_i M_i = \sum_{i=1}^{n} c_i M_i^2/cM_{w(c)} = \sum_{i=1}^{n} i^2 k_i c_1^i M_1^2 \Big/ \sum_{i=1}^{n} ik_i c_1^i M_1 \quad \text{(II-53)}$$

It is evaluated by using the Wales recursion formula.[26]

$$M_{z(c)} = d[cM_{w(c)}]/dc \quad \text{(II-54)}$$

Adams and Filmer[27] obtained the relationship between $M_{z(c)}$ and its apparent value, $M^a_{z(c)}$, viz.

$$M^a_{z(c)} = d[cM^a_{w(c)}]/dc \quad \text{(II-55)}$$

The operation required is the substitution of eq II-43 into eq II-55. Then

$$M^a_{z(c)} = M_{z(c)}/[1 + BM_{w(c)}c]^2 \quad \text{(II-56)}$$

Theoretically, it is possible to obtain relations similar to eq II-56 from the z + 1, z + 2, etc., average molecular weights, again by using the Wales recursion formula. However, these values involve the second and third derivatives of $M^a_{w(c)}$, and their numerical accuracies are probably too low to warrant their present practical application.

In seeking another experimental quantity, Adams[27] has defined a function

$$-\psi = (1/2M_1^2)\,d[cM_{w(c)}]^2/dc = cM_{w(c)}M_{z(c)}/M_1^2 = \sum_{i=1}^{n} c_i M_i^2/M_1^2 = \sum_{i=1}^{n} i^2 k_i {c_1}^i \quad \text{(II-57)}$$

Again assuming that only the second virial coefficient is involved

$$d\{M_1/cM^a_{w(c)}\}/dc = -M_1M^a_{z(c)}/[cM^a_{w(c)}]^2 = -M_1M_{z(c)}/[cM_{w(c)}]^2 = -\sum_{i=1}^{n} i^2 k_i {c_1}^i \Big/ \sum_{i=1}^{n} (ik_i {c_1}^i)^3 \quad \text{(II-58)}$$

This equation may be rearranged to give

$$\left(\sum_{i=1}^{n} ik_i {c_1}^i\right)^3 d\{M_1/cM^a_{w(c)}\}/dc = \sum_{i=1}^{n} i^2 k_i {c_1}^i \quad \text{(II-59)}$$

Then, combination of eq II-41, II-43, and II-59 yields

$$[d\{M_1/cM^a_{w(c)}\}/dc]/[(M_1/cM^a_{w(c)}) - BM_1]^3 = \psi \quad \text{(II-60)}$$

For the isodesmic reactions, eq II-26 may be introduced into eq II-57 to obtain

$$-\psi = \sum_{i=1}^{\infty} i^3 k^{i-1} {c_1}^i = c_1(1 + 4kc_1 + k^2c_1^2)/(1 - kc_1)^4 \quad \text{(II-61)}$$

which, upon substitution of eq II-28, becomes

$$-\psi = c(1 + 4kc_1 + k^2c_1^2)/(1 - kc_1)^2 \quad \text{(II-62)}$$

In connection with a study of the actual data for selected protein self-association systems (cf. section V), the question of the added value derived from incorporating into the analysis molecular weight averages other than the weight average, and based on the same c vs. r data, will be considered.

Apparent weight-average molecular weight and other experimental quantities derived from it are all functions of the equilibrium constants, the virial coefficient, the monomer concentration, and the total concentration (eq II-43, II-51, II-56, and II-60). A relationship between the monomer concentration, the total concentration, and the virial coefficients, which may be used to reduce by one the number of unknown parameters in the equations cited above, is now derived. If we define

$$f_1 = c_1/c \quad \text{(II-63)}$$

Then

$$dc_1 = f_1 dc + c df_1 \quad \text{(II-64)}$$

From eq II-42

$$dc_1/c_1 = M_1 dc/cM_{w(c)} \quad \text{(II-65)}$$

Combination of eq II-63, II-64, and II-65 and integration gives

$$\ln f_1 = \int_0^c [(M_i/M_{w(c)}) - 1]\,dc/c \quad \text{(II-66)}$$

This equation was first derived by Steiner;[28] it assumed solution ideality. Adams and Williams[11] extended this treatment to include the virial coefficient, defining the apparent monomer weight fraction, f_1^a, by

$$\ln f_1^a = \int_0^c [(M_1/M^a_{w(c)}) - 1]\,dc/c \quad \text{(II-67)}$$

Introduction of eq II-43 into this equation gives

$$\ln f_1^a = \int_0^c \left(\frac{M_1}{M_{w(c)}} - 1\right)\frac{dc}{c} + \int_0^c BM_1\,dc + \ldots \quad \text{(II-68)}$$

and

$$\ln f_1^a = \ln f_1 + BM_1c + \ldots \quad \text{(II-69)}$$

or

$$f_1^a = f_1 \exp(BM_1c + \ldots) \quad \text{(II-70)}$$

Thus, if the value of f_1^a is known experimentally, it is possible to obtain the monomer concentration c_1 from the expression

$$c_1 = f_1c = c_1^a \exp(-BM_1c - \ldots) \quad \text{(II-71)}$$

where c_1^a is its apparent value and expressed as

$$c_1^a = f_1^a c \quad \text{(II-72)}$$

D. Expressions for the Total Concentration as a Function of the Radial Distance

As was pointed out by Haschemeyer and Bowers,[30] an equation equivalent to eq II-35 for the i species may be integrated for c from a reference point in the cell, e.g., the position of the meniscus a, to an arbitrary point r, to obtain for the species i

$$c_i = c_{ia} \exp\{i[A'M_1(r^2 - r_a^2) - BM_1(c - c_a) - \ldots]\} \quad \text{(II-73)}$$

where

$$A' = A/2 \quad \text{(II-74)}$$

The total concentration may then be expressed as

$$c = \sum_{i=1}^{n} c_{ia} \exp\{i[A'M_1(r^2 - r_a^2) - BM_1(c - c_a) - \ldots]\} \quad \text{(II-75)}$$

which is the same as the expression used by Holladay and Sophianopoulos.[31] It is an extended form of a basic equation which has had a long history.

Substitution of eq II-19 into eq II-75 yields

$$c = \sum_{i=1}^{n} k_i c_{1a}{}^i \exp\{i[A'M_1(r^2 - r_a^2) - BM_1(c - c_a) - \ldots]\} \quad \text{(II-76)}$$

For the isodesmic reactions, introduction of eq II-73 into eq II-28 gives

$$c = \frac{c_{1a} \exp[A'M_1(r^2 - r_a^2) - BM_1(c - c_a) - \ldots]}{\{[1 - kc_{1a} \exp[A'M_1(r^2 - r_a^2) - BM_1(c - c_a) - \ldots]\}^2} \quad \text{(II-77)}$$

Sometimes it is more convenient to express the concentration of each species in the cell with reference to the initial total concentration, c^0, rather than to the concentration at the meniscus. Integration of eq II-37 now yields[32]

$$c_i = \Phi^i \exp\{i[-M_1\lambda\xi - BM_1(c_0 - c) - \ldots]\} \quad \text{(II-78)}$$

where b is the bottom of the cell, Φ^i are integration constants which are to be determined, and

$$\lambda = (A/2)(b^2 - a^2) \quad \text{(II-79)}$$

$$\xi = (b^2 - r^2)/(b^2 - a^2) \quad \text{(II-80)}$$

The integration constants may generally be related by

$$c^0 = \sum_{i=1}^{n} c_i{}^0 = \sum_{i=1}^{n} \Phi^i \int_0^1 \exp\{i[-M_1\lambda\xi - BM_1(c_0 - c) - \ldots]\} \, d\xi \quad \text{(II-81)}$$

where $c_i{}^0$ is the concentration of the i species in the original solution. The evaluation of the integration constant for the nonideal systems and even for ideal systems of many species is, however, exceedingly complicated. For the simple ideal monomer–dimer case, the integration constant obtained from eq II-81 has the form[32]

$$\Phi = \frac{1}{1 + \exp(-M_1\lambda)} \times \left\{\left[2M_1\lambda\left(\frac{1 + \exp(-M_1\lambda)}{1 - \exp(-M_1\lambda)}\right)c^0k_2 + 1\right]^{1/2} - 1\right\} \quad \text{(II-82)}$$

Reinhardt and Squire,[33] in their analysis of the sedimentation equilibrium data of ovine interstitial cell-stimulating hormone, employed the equation of Rinde

$$c_i = \frac{c_i{}^0\lambda M_i \exp(-M_i\lambda\xi)}{1 - \exp(-M_i\lambda)} \quad \text{(II-83)}$$

However, this treatment is based on the assumption that the mass-conservation statement given below remains valid in application to the evaluation of the data for a self-association system.[34]

$$\frac{2}{(b^2 - a^2)} \int_a^b c_i r \, dr = c_i{}^0$$

Actually, it is the mass of the *component* that is conserved—not the masses of the individual species. The statement should therefore be

$$c^0 = \frac{2}{b^2 - a^2} \int_a^b cr \, dr = \frac{2}{b^2 - a^2} \sum_{i=1}^{n} \int_a^b c_i r \, dr \quad \text{(II-84a)}$$

III. Analysis of Data for Self-Associating Systems

A. General Considerations

The quantities which are the goals of the analysis of a self-associating system are the equilibrium constants k_i, the monomer molecular weight (if it is not known independently), and also a quantitative measure of the influence of the solution nonideality. Knowledge of the collection of the k_i provides a description of the reaction stoichiometry. For example, if for some system one knew that $k_2 \neq 0$, and $k_3 = k_4 = \ldots = k_i = \ldots = 0$, he would then conclude that there was a monomer–dimer reaction occurring in that system. The reliability of the parameters obtained from data analysis will depend on the precision of the experiment. Data may be sufficiently imprecise to permit equally good fit of the data by more than one mechanism, and in later sections, we will show examples of such cases.

There are a number of quantities that are used for data analysis. The more common ones are the concentration, various point molecular weight averages, and the radial position in the cell. Both the Rayleigh interference system and the absorption optical system provide data that are directly related to solute concentration as a function of radius. Mathematical treatment of the c vs. r data may provide the apparent molecular weight averages, which furnish an immediate impression of the nature of the self-association process. In this section we seek to show how such experimental quantities can be analyzed to obtain reaction equilibrium constants and to describe the stoichiometry for several types of protein associations.

There are a number of ways for organizing the discussion of this topic. One might focus attention on the *type* of self-association process. For example, one could divide the self-association reactions into two categories: discrete reactions, in which the reaction proceeds to form a finite number of oligomeric species, and into indefinite or isodesmic reactions. An alternative and mathematically reasonable way of separating self-associations into categories is by the number of parameters required to describe the reaction processes. Physical reasoning suggests that, in general, a term involving BM_1 should be included. Two-parameter representations (assuming that M_1 is known) then have only one k_i (i.e., k) to be assigned. Falling into this category are the monomer–n-mer reactions (with monomer–dimer as a

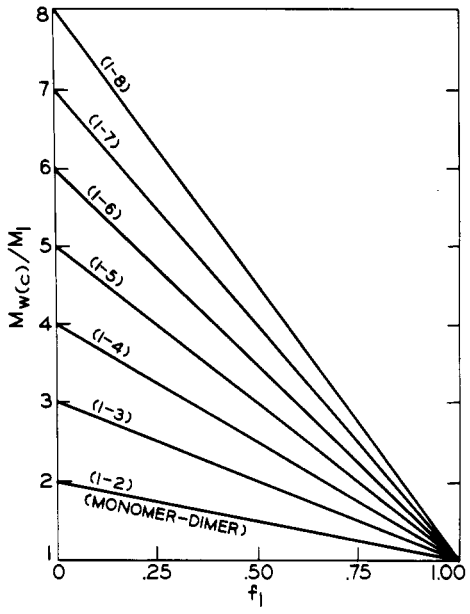

Figure 2. Standard plot of $M_{w(c)}/M_1$ vs. f_1 for monomer–n-mer association reaction. From Chun and Kim.[38]

special case) and the indefinite self-association.

Instead we categorize the procedures according to the type of independent variable used in the analysis. We first discuss diagnostic plots, in which functions of molecular weight averages or of f_1 are plotted to determine the number and the nature of species present. Secondly, we consider the types of analysis in which the radius r is taken as the independent variable, so that the association constants (and virial coefficients) are evaluated from the total solute concentration as a function of r. Finally, we discuss methods in which analysis is based on data for molecular weight averages as functions of total concentration c. In all three cases, we discuss the special case of ideal systems separately, since the analysis for them helps one to visualize the manipulative steps. In so doing, we do not advocate calculations based on the assumption that the system behaves ideally. On the contrary, in the absence of independent data which suggest ideal behavior, we believe that the system should be assumed to be *nonideal.*

B. Diagnostic Plots

Diagnostic plots are used to determine the stoichiometry of a self-association system by appropriate graphing of combinations of two or more of the several point average molecular weights or by using graphs of the c vs. r data. In the past, the practical application of the diagnostic plot to two or three species systems had been largely restricted to ideal solutions. However, even in the case of data for nonideal systems there are now procedures by which correction for, or avoidance of, this solution nonideality may make it possible to employ the diagnostic plot for any arbitrary self-associaton reaction. This provides a much simplified treatment of the data. (Even so, it would be of great advantage to be able to have information concerning the activity coefficients in advance of the analysis.)

1. Ideal Systems

A basic premise of the diagnostic plot for an ideal system is that the experimental quantities (in this instance, average molecular weights) can be expressed as simple functions of the weight fractions, f_i, of the species present. By definition

$$f_i = c_i/c \qquad \left(\sum_{i=1}^{n} f_i = 1\right) \tag{III-1}$$

where i represents the number of monomers present in any arbitrary species. Thus from eq II-41, II-47, and II-53, we obtain

$$M_{w(c)}/M_1 = \sum_{i=1}^{n} if_i \tag{III-2}$$

$$M_1/M_{n(c)} = \sum_{i=1}^{n} f_i/i \tag{III-3}$$

and

$$M_{z(c)}M_{w(c)}/M_1^2 = \sum_{i=1}^{n} i^2 f_i \tag{III-4}$$

It should be remembered that f_1 is an experimental quantity. Thus, if all the f_i except f_1 can be eliminated by a suitable combination of the above equations, it is possible to relate these experimental quantities with the i and thus to evaluate them. This is the basis of the diagnostic plot for ideal two component systems. The plot which has been devised was developed independently by Sophianopoulos and Van Holde[35] and by Elias and Bareiss.[14] The use of such plots has been extended and elaborated, for instance, by Roark and Yphantis,[36] Teller et al.,[37] and Chun and Kim.[38]

For those systems where the monomer associates to form n-mer, with negligible concentration of intermediate species, eq III-2 and III-3 become

$$M_{w(c)}/M_1 = n - (n - 1)f_1 \tag{III-5}$$

and

$$M_1/M_{n(c)} = 1/n + (1 - 1/n)f_1 \tag{III-6}$$

Thus plots of $M_{w(c)}/M_1$ vs. f_1 and of $M_1/M_{n(c)}$ vs. f_1 should give straight lines and, from either, the slope or the intercept of these lines determines the value of n. An example of the use of such plots is given in Figure 2. If the value of M_1 is not known, plots of $M_{w(c)}$ vs. f_1 and of $1/M_{n(c)}$ vs. f_1 should also give straight lines and from the slopes and intercepts of these lines the values of n and M_1 may be also obtained.

It is also possible to combine eq III-5 and III-6 to eliminate f_1 to obtain

$$M_{w(c)}/M_1 = (1 + n) - nM_1/M_{n(c)} \tag{III-7}$$

from which the values of n and M_1 may be obtained as before. Here a plot of $M_{w(c)}/M_1$ vs. $M_1/M_{n(c)}$ is similar to that of $M_{w(c)}/M_1$ vs. f_1. Equation III-5 may be also combined with eq III-4 (for the monomer–n-mer case) to obtain an expression very similar to eq III-7. In general form this equation may be written[36]

$$M_{j(c)}/M_1 = 1 + n - n(M_1/M_{(j-1)(c)}) \tag{III-8}$$

where $M_{j(c)} = M_{n(c)}, M_{w(c)}, M_{z(c)}, M_{(z+1)(c)}$ as $j = 0, 1, 2, 3$, respectively. Equation III-8 suggests that for a monomer–n-mer reaction, all of the plots for $M_{(z+1)(c)}$ vs. $1/M_{z(c)}$, $M_{z(c)}$ vs. $1/M_{w(c)}$, and $M_{w(c)}$ vs. $1/M_{n(c)}$ should superimpose (see Figure 8 of ref 36).

If there is an intermediate species, m, between that of the monomer and the n-mer present, the experimental points should initially lie on the monomer–m-mer line, curve upward, and at higher concentration approach the monomer–n-mer line (see Figure 10 of ref 37.)

For the three-species system where monomer, m-mer, and n-mer are at equilibrium, we obtain from eq III-2–4

$$M_{w(c)}/M_1 = f_1 + mf_m + nf_n \tag{III-9}$$

$$M_1/M_{n(c)} = f_1 + f_m/m + f_n/n \tag{III-10}$$

and

$$M_{w(c)}M_{z(c)}/M_1^2 = f_1 + m^2 f_m + n^2 f_n \tag{III-11}$$

The quantity f_1 can be eliminated by using the identity

$$f_1 + f_m + f_n = 1 \quad \text{(III-12)}$$

Then

$$M_{n(c)}M_{w(c)}M_{z(c)} - (1 + m + n)M_1M_{n(c)}M_{w(c)} + (m + n + mn)M_1^2M_{n(c)} + f(m,n,M_1) = 0 \quad \text{(III-13)}$$

where $f(m,n,M_1)$ is a function of m, n, and M_1. Differentiation of III-13 with respect to $M_{n(c)}$ yields

$$\mathrm{d}[M_{n(c)}M_{w(c)}M_{z(c)}]/\mathrm{d}M_{n(c)} = (1 + m + n)M_1\mathrm{d}[M_{n(c)}M_{w(c)}]/\mathrm{d}M_{n(c)} - (m + n + mn)M_1^2 \quad \text{(III-14)}$$

Therefore a plot of $\mathrm{d}[M_{n(c)}M_{w(c)}M_{z(c)}]/\mathrm{d}M_{n(c)}$ against $\mathrm{d}[M_{n(c)}M_{w(c)}]/\mathrm{d}M_{n(c)}$ should give a straight line if only three species are present. If the value of M_1 is known, the values of m and n may be obtained from the slope and intercept.

However, Chun and Kim[38] have pointed out that the above treatment involves a double differentiation step which introduces a large error; this is especially true at low concentrations. Instead they used only eq III-9 and III-10 to obtain

$$m\{nM_1/M_{n(c)} - 1\} - \{n - M_{w(c)}/M_1\} = (m - 1)(n - 1)f_1 \quad \text{(III-15)}$$

Differentiation of the above equation gives

$$\frac{1}{M_1}\frac{\mathrm{d}M_{w(c)}}{\mathrm{d}f_1} = (m - 1)(n - 1) - nm\,\mathrm{d}\left(\frac{M_1}{M_{n(c)}}\right)\Big/\mathrm{d}f_1 \quad \text{(III-16)}$$

Thus a plot of $\mathrm{d}(M_{w(c)}/M_1)\mathrm{d}f_1$ vs. $\mathrm{d}(M_1/M_{n(c)})/\mathrm{d}f_1$ should give a straight line for the three-species system and from the slope of the intercept one may calculate the values of m and n if the value of M_1 is known. By appropriate rearrangement of eq III-15, one can obtain the expression

$$\frac{M_{w(c)}/M_1 - 1}{M_1/M_{n(c)} - 1} = \frac{(f_1 - 1)}{(M_1/M_{n(c)} - 1)} \times (m - 1)(n - 1) - mn \quad \text{(III-17)}$$

Here, a plot of $(M_{w(c)}/M_1 - 1)/(M_1/M_{n(c)} - 1)$ vs. $(f_1 - 1)/(M_1/M_{n(c)} - 1)$ is indicated. Unlike eq III-14 and III-16, this expression requires no differentiation of any molecular weight averages, so that errors should be comparatively smaller at sufficiently high concentrations.

The independent variable in eq III-17 is constrained by the physical situation to nonzero values. For this reason, the intercept can be determined only by a long extrapolation. The slope can be determined precisely, since data at high concentrations may be used. The value of this slope will limit the number of association types that must be considered.

Diagnostic plots for the ideal isodesmic association have been developed by several authors; we make reference again to Chun and Kim.[38] Combination of eq II-28, II-44, and II-52 gives

$$M_{n(c)}/M_1 = 1/\sqrt{f_1} \quad \text{(III-18)}$$

$$M_{w(c)}/M_1 = \{2M_{n(c)}/M_1\} - 1 = (2/\sqrt{f_1}) - 1 \quad \text{(III-19)}$$

A plot of either $M_{w(c)}/M_1$ or $M_{n(c)}/M_1$ against $1/\sqrt{f_1}$ should give a straight line.

The diagnostic procedures discussed so far make use of the several average molecular weights. It is expected that the development of similar procedures for direct use with c vs. r data will be more difficult. However, Adams and Williams[11] have made the attempt to distinguish between some of the simpler models by this route. For an ideal system ($M^e_{w(c)} = M_{w(c)}$), combination of eq II-38 and II-22 gives

$$\frac{1}{Ar}\frac{\mathrm{d}c}{\mathrm{d}r} = cM_{w(c)} = \sum_{i=1}^{n} c_iM_i \quad \text{(III-20)}$$

Then, the substitution of eq II-73 when written for an ideal system, and eq III-1 into eq III-20 yields

$$\frac{1}{AM_1rc_a}\frac{\mathrm{d}c}{\mathrm{d}r} = \sum_{i=1}^{n} if_{ia}\exp[iA'M_1(r^2 - r_a^2)] \quad \text{(III-21)}$$

where f_{ia} refers to the f_i at the meniscus.

For the case of monomer–dimer–trimer system the introduction of the statement

$$f_{1a} + f_{2a} + f_{3a} = 1$$

serves to convert eq III-21 to the form

$$\left\{\frac{1}{AM_1rc_a}\frac{\mathrm{d}c}{\mathrm{d}r}\Big/\exp[A'M_1(r^2 - r_a^2)]\right\} - 1 = f_{2a}\{2\exp[A'M_1(r^2 - r_a^2)] - 1\} + f_{3a}\{3\exp[2A'M_1(r^2 - r_a^2)] - 1\} \quad \text{(III-22)}$$

Thus, the graph of $[\{(1/AM_1rc_a)/\exp[A'M_1(r^2 - r_a^2)]\} - 1]/\{2\exp[A'M_1(r^2 - r_a^2)] - 1\}$ vs. $\{3\exp[2A'M_1(r^2 - r_a^2)] - 1\}/\{2\exp[A'M_1(r^2 - r_a^2)] - 1\}$ will give an inclined straight line for the monomer–dimer–trimer system and a horizontal straight line for monomer–dimer system. On the other hand, a plot of

$$\frac{1}{AM_1rc_a}\frac{\mathrm{d}c}{\mathrm{d}r}\Big/\exp[A'M_1(r^2 - r_a^2)] - 1/3\exp[2A'M_1(r^2 - r_a^2)]$$

vs.

$$2\exp[A'M_1(r^2 - r_a^2)] - 1/3\exp[A'M_1(r^2 - r_a^2)] - 1$$

will give horizontal straight line if only monomer and trimer are present. As in the case with all of the available diagnostic plots, this procedure is limited in applicability to the simpler models.

2. Nonideal Systems

The type of treatment for ideal systems given thus far should also apply to nonideal systems at sufficiently low concentrations, provided that the nonideality term BM_1 is sufficiently small compared to the k_i. Deviations from the ideal behavior as given in Figure 2 will occur at higher concentrations. In order to extend the useful range of the "two species ideal" diagnostic plots to nonideal systems, Roark and Yphantis[36] obtained the expression

$$2M^e_{(j+1)(c)}/M_1 - M^e_{(j+2)(c)}/M_1 = (1 + n) - nM_1\left\{\frac{2}{M^e_{j(c)}} - \frac{1}{M^e_{(j+1)(c)}}\right\} \quad \text{(III-23)}$$

which may be compared with eq III-8. Here $M^e_{j(c)} = M^e_{n(c)}$, $M^e_{w(c)}$, $M^e_{z(c)}$, and $M^e_{(z+1)(c)}$ as $j = 0, 1, 2, 3$, respectively. It was found that when the apparent average molecular weights are expanded into a Taylor series in concentration with virial coefficients, the linear combinations in eq III-23 eliminate terms linear in concentration. Plots of the left-hand side of eq III-23 against $2/M^e_{j(c)} - 1/M^e_{(j+1)(c)}$ extend the linear portion of the "two species" plot.

An analogous, but perhaps more satisfactory, approach also has been described by Roark and Yphantis. They write a reciprocal of molecular weight for the first "ideal" moment, a quantity which they describe as

$$1/M_{y(c)} = \mathrm{d}[1/cM_{n(c)}]/\mathrm{d}[1/c] = \frac{2}{M_{n(c)}} - \frac{1}{M_{w(c)}} \quad \text{(III-24)}$$

This moment contains no term in B, so it will have a unique value irrespective of whether the solution is ideal or nonideal. The equation is obtained by a combination of eq II-43 and II-51, with only linear terms in c being considered.

Chun et al.[39] have continued this type of development to

eliminate B in the development of standard plots to test for the type and extent of the association, with particular reference to monomer–n-mer systems. Basic to the analysis are two equations which are obtained by combining eq II-43 and eq II-51, and eq II-43 and eq II-69, respectively. They are, in effect, guides to two species plots which can be extended to nonideal systems. The first is

$$\xi = \frac{2M_1}{M^a_{n(c)}} - \frac{M_1}{M^a_{w(c)}} = \frac{2M_1}{M_{n(c)}} - \frac{M_1}{M_{w(c)}} \qquad \text{(III-25)}$$

which is (III-24) multiplied by M_1. The second is

$$\eta = \frac{M_1}{M^a_{w(c)}} - \ln f_1{}^a = \frac{M_1}{M_{w(c)}} - \ln f_1 \qquad \text{(III-26)}$$

These two equations permit the analysis of nonideal systems by using two-species plots, but again, restricted to terms in c. Of course they are also applicable to ideal self-associations as well.

For the nonideal isodesmic case, combination of eq III-25, III-26, and III-19 gives a relationship between ξ and η. A diagnostic plot based on this relation is described by Chun et al. (e.g., see Figure 2 of ref 39).

For monomer–n-mer systems, the introduction of eq III-7 into eq III-25, and eq III-5 into eq III-26 transforms the ξ and η into functions of $M_{w(c)}/M$ and n. A combination diagnostic plot based on this relation is given in Figure 1, also of ref 39.

Both Roark and Yphantis[36] and Chun et al.[39] have also treated the three-species self-association reaction diagnosis by using similar combinations of equations. The equations are now quite involved, and they will not be set down in this report.

C. Determination of Association Constants and Virial Coefficients from the Total Concentration vs. Radial Distance Curves

Of the two general methods to obtain association constants and virial coefficients (in addition to the establishment of the stoichiometry), that which employs directly the total concentration as a function of radial distance has seemingly the greater potential for the improvement of an analytical result. However, to date this method has received far less attention, probably because of the relative complexity of the required calculations. With the present availability of computers, this fact should no longer be a deterrent, and this method may yet become the preferred analysis. Actually the numerical analysis of functions which consist of sums of exponents is a recurrent problem in many areas of science and engineering, and a large number of approaches and procedures have been described for this manipulation. It will be noted that the expressions descriptive of the equilibrium concentration distribution in ordinary polydisperse systems are formally the same as these equations, and therefore that the same computational procedures may be applied even when the solute undergoes a chemical reaction.

Depending on the intricacy of the stoichiometry of the reaction, the description of this general method may be divided into three sections for discussion: (1) simple two-species systems not requiring a computer; (2) more involved ideal systems for which the value of M_1 is known and for which the behavior may be expressed by using linear equations; and (3) ideal systems for which the M_1 is not known and nonideal systems for which the descriptive mathematical expressions are now nonlinear equations.

1. Two-Species Systems

The total concentration versus distance for a two-species ideal system is written (eq II-75)

$$c = c_{1a}\exp[A'M_1(r^2 - r_a{}^2)] + k_n c_{1a}{}^n \exp[nA'M_1(r^2 - r_a{}^2)] \qquad \text{(III-27)}$$

It can be arranged to read

$$c/\exp[A'M_1(r^2 - r_a{}^2)] = c_{1a} + k_n c_{1a}{}^n \exp[(n-1)A'M_1(r^2 - r_a{}^2)] \qquad \text{(III-28)}$$

Thus a plot of the left-hand side of this equation as a function of $\exp[(n-1)A'M_1(r^2 - r_a{}^2)]$ should give an inclined straight line, and from its slope and intercept k_n may be computed.

When the value of B is not negligible, eq II-75 assumes the following form for the monomer–n-mer case

$$c = c_{1a}\exp[A'M_1(r^2 - r_a{}^2) - BM_1(c - c_a)] + c_{na}\exp\{n[A'M_1(r^2 - r_a{}^2) - BM_1(c - c_a)]\} \qquad \text{(III-29)}$$

Noting that

$$c_{na} = c_a - c_{1a}$$

eq III-29 may be rearranged to obtain

$$c_{1a} = \frac{c\exp\{-nM_1[A'(r^2 - r_a{}^2) - B(c - c_a)]\} - c_a}{\exp\{(1-n)M_1[A'(r^2 - r_a{}^2) - B(c - c_a)]\} - 1} \qquad \text{(III-30)}$$

For the evaluation of c_{1a}, a value of B is chosen and then values of the right-hand side of this equation at several r values are plotted against c. This process may be repeated with varying values of B until a horizontal straight line is obtained. From the intercept between this line and the vertical coordinate, the concentration c_{1a} may be obtained, with the B value which gives this horizontal straight line being taken to be correct. Then the c_{na} value is readily obtained, and the use of eq II-19 gives the association constant. This is a straightforward procedure similar to the one employed by Van Holde and Rossetti[16] for analyzing isodesmic self-associating systems with apparent weight-average molecular weight data.

For the indefinite isodesmic self-association there is but one association constant to be determined (it having been assumed that a constant increment of free energy is involved for the addition of 1 mol of monomer to any i-mer); therefore the procedure of the linear plot for two-species systems may be readily adapted. For the ideal isodesmic reaction eq II-77 becomes

$$c = \frac{c_{1a}\exp[A'M_1(r^2 - r_a{}^2)]}{\{1 - kc_{1a}\exp[A'M_1(r^2 - r_a{}^2)]\}^2} \qquad \text{(III-31)}$$

Upon rearrangement we find

$$\{1 - [\sqrt{c_{1a}}\exp[A'M_1(r^2 - r_a{}^2)]/c\}]^{1/2}/c_{1a}\exp[A'M_1(r^2 - r_a{}^2)] = k \qquad \text{(III-32)}$$

Here the value of c_{1a} may be adjusted until a plot of the left-hand side of this equation against c gives a horizontal straight line. From the intercept the value of the intrinsic constant may be obtained. For the nonideal system with one virial coefficient, eq II-77 may be directly applied to give

$$c = \frac{c_{1a}\exp[A'M_1(r^2 - r_a{}^2) - BM_1(c - c_a)]}{\{1 - kc_{1a}\exp[A'M_1(r^2 - r_a{}^2) - BM_1(c - c_a)]\}^2} \qquad \text{(III-33)}$$

A general computer-based curve-fitting procedure may be employed here to evaluate the parameters k, M_1, B, and c_{1a}.

An alternative procedure, recently described by Milthorpe et al.,[40] makes use of the relationship

$$y_1c_1 = y_{1x}c_{1x}\exp[A'M_1(r^2 - r_x{}^2)] \qquad \text{(III-34)}$$

where the r_x refers to an arbitrary reference radial distance. This is a more general form of eq II-73 for the monomer species in that it does not require the use of the approximation eq II-18. This equation may be readily converted to give

$$\Omega = \frac{c\exp[A'M_1(r_x{}^2 - r^2)]}{c_x} = \frac{y_{1x}c_{1x}c}{y_1c_1c_x} \qquad \text{(III-35)}$$

from which one obtains

$$\lim_{c \to 0} \Omega = \frac{y_{1x}c_{1x}}{c_x} \qquad \text{(III-36)}$$

since as $c \to 0$, $y_1 \to 1$, and $c_1/c \to 1$.

Thus by extrapolating the values of Ω to infinite dilution one obtains the value of $y_{1x}c_{1x}$ and once this value is known, eq III-35 may be used to obtain the values of y_1c_1 at any radial distance. At this point eq II-18 can be introduced in the form

$$B = \ln\left[\frac{y_1c_1}{c_1}\right] \Big/ M_1{\cdot}c \qquad \text{(III-37)}$$

In cases for which c_1 can be expressed as an explicit function of c, this function can be combined with eq III-37 at two ($y_1c_1{\cdot}c$) data points to yield an expression in which the equilibrium constant is the only unknown. Otherwise a successive approximation procedure based upon eq II-21 may be employed.

2. Solution of Linear Equations

For the ideal system eq II-75 becomes

$$c = \sum_{i=1}^{n} c_{ia} \exp[iA'M_1(r^2 - r_a^2)] \qquad \text{(III-38)}$$

When M_1 is known, the only unknown parameters are the c_{ia}, and therefore the equation is linear. There are two general methods for solving linear equations: the least-squares method and linear programming.

For the method of least squares, eq III-38 may be re-expressed for each data point, j, as follows:

$$c_j = \sum_{i=1}^{n} c_{ia} \exp[iA'M_1(r_j^2 - r_a^2)] + \delta_j \qquad \text{(III-39)}$$

$$(j = 1, 2, 3, \ldots, m)$$

where δ_j is the difference between the measured and computed value of c_j. The problem of the least-squares method is to find the values of c_{ia} which satisfy the condition

$$\partial\left\{\sum_{j=1}^{m} (\delta_j^2)\right\} \Big/ \partial c_{ia} = 0 \qquad \text{(III-40)}$$

simultaneously for each component i.

Combination of these two statements then gives the set of equations

$$\sum_{j=1}^{m} c_j \exp[iA'M_1(r_j^2 - r_a^2)] = \sum_{k=1}^{n} c_{ka} \sum_{j=1}^{m} \exp[iA'M_1(r_j^2 - r_a^2)] \exp[kA'M_1(r_j^2 - r_a^2)] \qquad \text{(III-41)}$$

A test of this method for the analysis was made by Haschemeyer and Bowers,[30] using synthetic data. The results indicated that even up to five-species systems the result is very good (discrepancy of less than 10%) if the difference in molecular weight is reasonably spread and if the distribution of association constants is such that only one or two species do not predominate. However, we are not aware of any extended description in the literature of the application of this Teller–Haschemeyer–Bowers method in the evaluation of actual experimental data; this is the crucial test.

The method of least squares sometimes yields negative values for the coefficients, a physically untenable situation. Thus some kind of constraint must be imposed or some other method must be employed to alleviate this difficulty. The most common alternative is the method of linear programming which has been successfully applied to the analysis of polydisperse systems.[41,42] This method seeks to minimize the quantity[43]

$$\sum_{j=1}^{m} |\delta_j|$$

with the constraint that

$$c_{ia} > 0$$

for all i. The minimization is carried out by using the modified simplex algorithm. Although this method seems not to have been applied in the analysis of self-associating systems, Scholte[41,42] was able to describe a curve-fitting procedure even for bimodal and trimodal molecular weight distribution of polymers, and with reasonable success.

3. Solution of Nonlinear Equations

There are two types of nonlinear equations to represent c as a function of r. One is the expression for ideal systems (e.g., eq III-39) in which M_1 is also a parameter to be determined. The other is the expression for nonideal systems, and it may be obtained by re-expressing eq II-75 for an arbitrary individual data point, j,

$$c_j = \sum_{i=1}^{n} c_{ia} \exp\{i[A'M_1(r_j^2 - r_a^2) - BM_1(c_j - c_a)]\} + \delta_j \qquad \text{(III-42)}$$

This equation is an implicit function of c and the solution of it is more involved than that for eq III-39. The general approach in the solution of nonlinear equations is to transform them into linear equations by using various approximations and then to solve these linear equations, using the method of least squares. The approximations involved are to fix the parameter values in the exponentials or to expand eq III-39 or III-42 into Taylor series.

There are a large number of procedures which involve a Taylor's series expansion.[43] Since they are basically similar in nature and since the choice of which method to use will be dependent on the nature of the associating systems to be studied, we do not attempt to describe all of the methods available. Two of the more typical methods will be discussed here to illustrate the principle.

One of them, the Newton–Raphson method, seeks to minimize the quantity, $\Sigma\delta_j^2$ from eq III-42. Denoting the primed parameters, α_1', as the initial approximations we assume that the required minimum is at $\alpha_i = \alpha_i' + \Delta\alpha$. Then we can find the minimum value at

$$\left[\left\{\partial\left(\sum_{j=1}^{m} \delta_j^2\right) \Big/ \partial\alpha_i\right\}\right]_{(\alpha_i' + \Delta\alpha_i)} = 0 \qquad \text{(III-43)}$$

Expansion of this equation into Taylor series about α_i', and ignoring the terms involving the third and higher derivatives, gives

$$\left\{\partial\left(\sum_{j=1}^{m} \delta_j^2\right) \Big/ \partial\alpha_i\right\}_{\alpha_i'} = -\sum_{k=1}^{n} \left\{\partial^2\left(\sum_{j=1}^{m} \delta_j^2\right) \Big/ \partial\alpha_i\partial\alpha_k\right\}_{\alpha_i} x_i\Delta\alpha_k \qquad \text{(III-44)}$$

This is a set of simultaneous equations from which $\Delta\alpha_i$ may be evaluated. These values are then added to α_i' to start the second iteration. This operation may be continued until the value of $\Delta\alpha_i$ becomes negligible. To our knowledge this approach has not been applied to self-associating systems.

In the Gauss–Newton method, eq III-42 is directly differentiated to obtain

$$d_j = (c_j - c_j') = \sum_{k=1}^{n} (\partial c_j'/\delta\alpha_k)\Delta\alpha_k \qquad \text{(III-45)}$$

Here, d_j is the difference between observed (c_j) and computed (c_j') values of the concentration. The least-squares determination of the coefficients of this equation which gives the minimum value for Σd_j^2 is now made and the resulting values of $\Delta\alpha$ are added to α_k to obtain the next approximation; the iteration is repeated until these error terms become negligible.

TABLE I. Exponential Analysis of Self-Association: Three-Species System (Taken from Haschemeyer and Bowers[30])[a]

	True value		Linear		Nonlinear	
i	c_{ia}	k_i	c_{ia}	k_i	c_{ia}	k_i
A 1	0.62	1.0	0.62	1.0	0.63	1
2	0.50	1.0	0.48	0.94	0.49	0.94
3	0.49	1.0	0.50	0.98	0.48	0.93
B 1	0.77	1.0	0.78	1.0	0.79	1.0
2	0.0	10^{-6}	0	0	0	0
3	0.83	1.0	0.84	0.98	0.82	0.94
C 1	0.034	1.0	0.041	1.0	0.032	1.0
2	1.49	10^3	1.47	6.6×10^2	1.46	1.06×10^3
3	0.079	10^3	0.089	6.2×10^2	0.104	1.47×10^3
D 1	0.1	1.0	0.11	1.0	0.09	1.0
2	0.011	1.0	−0.016	−1.1	0.07	7.33
3	1.49	10^3	1.51	7×10^2	1.43	1.14×10^3

[a] Physical parameters used for generation of simulated sedimentation equilibrium data: $\bar{v}$, 0.758 cm^3/g; ρ, 1.0 g/cm^3; T 293 K; a, 5.905 cm; b, 6.205 cm; M_1, 60 000.

The application of this approach to the self-associating systems was made by Van Holde et al.[9] Haschemeyer and Bowers[30] investigated the reliability of this approach by computing synthetic c vs. r data for several ideal systems for which the values of M_1 are unknown, generating random errors of 0.03 fringes. The results of the analysis of these synthetic data, using the Gauss–Newton method, are also listed in Table I. For the monomer–dimer–trimer case and when k_2 and k_3 are comparable in value, the association constants as computed agreed to within 7% of the assigned values. This same analysis was also used independently by Rosenthal,[44] who found a similar agreement between the assigned and calculated k for monomer–dimer case as shown in Table II. However, for four- or five-species systems the discrepancies were over 20% (see Table III).

The same general procedure was also tested for nonideal systems by Holladay and Sophianopoulos[31] who assumed that M_1 in eq III-42 is predetermined. However, very good initial values for the parameters are needed for the convergence, because a very ill-conditioned matrix frequently has to be inverted and the range of c_j is often not broad enough for a satisfactory solution. To minimize the computational difficulties the value of B was fixed, so that the resulting linear equations could be solved for c_{ia}, using the least-squares method. By a stepwise search, the B value which minimized $\Sigma\delta_j^2$ was found. These workers limited the unknown parameters in eq III-42 to five: B and four c_{ia}. The concentrations, c_j, were obtained by integration of the concentration gradient.

The initial results have shown, however, that the synthetic data for short column–low speed sedimentation equilibrium experiments with random error of 0.1% in concentration gradient do not give a unique answer in contrast to the exact synthetic data

TABLE II. Analysis of Simulated c vs. r Data (Taken from Rosenthal[44])

M_i	i	c_a(input)[a]	c_a(output)[a]	M_1	k_i(input)	k_i(output)
100 000	1	4.17×10^{-2}	4.03×10^{-2}	101 000		
200 000	2	7.20×10^{-4}	6.13×10^{-4}		0.41	0.38
100 000	2	1.85×10^{-2}	1.82×10^{-2}	50 000		
150 000	3	1.57×10^{-3}	1.38×10^{-3}		0.39	0.32
200 000	4	1.19×10^{-4}	9.78×10^{-5}		0.35	0.30
100 000	2	1.53×10^{-1}	1.57×10^{-1}	50 900		
150 000	3	4.85×10^{-2}	5.20×10^{-2}		0.66	0.64
200 000	4	1.40×10^{-2}	8.76×10^{-3}		0.60	0.36
100 000	4	4.17×10^{-2}	9.8×10^{-2}	28 000		
125 000	5	1.58×10^{-2}	3.0×10^{-1}			
150 000	6	0.0	5.0×10^{-2}			
175 000	7	2.02×10^{-3}	1.2×10^{-3}			
200 000	8	7.20×10^{-4}	7.5×10^{-3}			
25 000	1	3.12×10^{-1}	3.25×10^{-1}	25 000		
50 000	2	7.42×10^{-2}	6.39×10^{-2}		0.76	0.60
100 000	4	2.65×10^{-3}	2.36×10^{-3}		0.28	0.21
150 000	6	0.0	2.57×10^{-8}		0.0	2.2×10^{-5}
200 000	8	1.75×10^{-6}	1.34×10^{-6}		1.98×10^{-2}	1.34×10^{-6}

[a] Concentration expressed in fringe numbers.

TABLE III. Analysis of Simulated c vs. r Data for Four-Species System (Taken from Holladay and Sophianopoulos[31])

$B \times 10$ mol dL g^{-2}	$c_{1a} \times 10^2$ g/dL	$c_{2a} \times 10^2$ g/dL	$c_{3a} \times 10^2$ g/dL	$c_{4a} \times 10^2$ g/dL	δ^2
3.00	1.000	1.900	2.000	1.800	
		A. Exact Data: One Cell			
1.00	1.6334	0.0	3.9925	1.0737	6.31×10^{-9}
2.00	1.4517	0.6255	3.2547	1.3686	4.98×10^{-10}
3.00	0.9960	1.9097	1.9929	1.8018	2.41×10^{-11}
4.00	0.4924	3.3207	0.6169	2.2696	6.74×10^{-10}
		B. Noisy Data: One Cell			
1.0	1.6537	0.0	3.9553	1.0924	1.90×10^{-8}
2.0	1.4282	0.7294	3.1362	1.4082	1.12×10^{-8}
3.0	0.9687	2.0228	1.8664	1.8437	8.46×10^{-9}
4.0	0.4604	3.4468	0.4793	2.3146	6.55×10^{-9}
4.352	0.2901	3.9316	0.0	2.4793	6.11×10^{-9}
		C. Noisy Data: Two Cells			
2.00	1.4809	0.0	4.0805	1.0971	3.84×10^{-5}
3.00	1.0455	1.8254	2.0399	1.7927	1.38×10^{-7}
4.00	0.0	4.1557	0.0	2.4542	2.41×10^{-4}

TABLE IV. Determination of Number of Species Using Two-Cells Data (Taken from Holladay and Sophianopoulos[31])

$B \times 10$ mol dL g^{-2}	$c_{1a} \times 10^2$ g/dL	$c_{2a} \times 10^2$ g/dL	$c_{3a} \times 10^2$ g/dL	$c_{4a} \times 10^2$ g/dL	δ^2
		A. Monomer–Dimer–Trimer			
3.00	1.000	1.900	2.000	0.0	
2.00	0.5558	2.5502	1.7618	0.00	2.29×10^{-5}
2.80	0.9124	2.0326	1.9498	0.00	1.03×10^{-6}
3.00	1.0051	1.8981	1.9985	0.00	5.20×10^{-8}
3.20	0.9564	2.0004	1.9238	0.0202	7.52×10^{-8}
3.40	0.9034	2.1107	1.8441	0.0415	1.27×10^{7}
4.00	0.7267	2.4772	1.5820	0.1101	5.20×10^{-3}
		B. Monomer–Tetramer			
3.00	1.000	0.0	0.0	1.800	
2.00	0.8262	0.00	0.3302	1.6251	4.29×10^{-6}
2.80	0.9661	0.00	0.0679	1.7631	1.91×10^{-7}
3.00	1.0017	0.0007	0.00	1.7986	1.48×10^{-8}
3.20	0.9831	0.00	0.00	1.8074	2.03×10^{-6}
4.00	0.00	0.8622	0.00	1.7408	2.48×10^{-4}

as shown in Table IV. Therefore, a method to "extend" mathematically the column length using two short columns was used. If c' is the concentration in the second cell and the position of the meniscus for this second cell is denoted as r_e, then,

$$c_h' = \sum_{i=1}^{n} c_{ie}' \exp\{i[A'M_1(r_h^2 - r_e^2) - BM_1(c_h' - c_e')]\} + \delta' \quad \text{(III-46)}$$

Also if the position in the second cell where the concentration is the same as that at the bottom, b, of the first cell is measured as d, then

$$c_e' = c_{ia} \exp\{i[A'M_1(r_b^2 - r_a^2 - r_d^2 + r_e^2) - BM_1(c_e' - c_a)]\} \quad \text{(III-47)}$$

Introduction of eq III-47 into eq III-46 then gives

$$c_h' = \sum_{i=1}^{n} c_{ia} \exp\{i[A'M_1(r_b^2 - r_a^2 - r_d^2 + r_e^2) - BM_1(c_e' - c_a)]\} \times \exp\{i[A'M_1(r_h^2 - r_e^2) - BM_1(c_h' - c_e')]\} + \delta_h' \quad \text{(III-48)}$$

Thus, if the position r_d is determined, c_h' can be related to the c_{ia} directly. The least-squares method then involves the equality:

$$\partial\left\{\sum_{j=1}^{m} \delta_j^2 + \sum_{h=1}^{p} \delta_h'^2\right\} \Big/ \partial c_{ia} = 0 \quad \text{(III-49)}$$

For the computations in those cases for which the simpler self-associations are involved, this method, while not much used, is quite successful.

4. Method of Moments: Ideal Systems

Recognizing the difficulties of treating equations with exponential functions, Dyson and Isenberg[45] have proposed the use of moments of the curves for the analysis of sums of exponentials. For the systems with n species with an equilibrium experimental radial distribution, the total concentration (eq II-75) is expressed as a function of the relative position, $z = a^2 - r^2$, as follows:

$$c(z) = \sum_{i=1}^{n} c_i(a) \exp[-A'M_i z] \quad \text{(III-50)}$$

where $c_i(a)$ is the concentration at an arbitrary reference radial position, a. Then the $2n$ moments are defined as

$$\mu_k = \int_0^z c(z) z^k \, dz \qquad (k = 0, 1, \ldots, 2n-1) \quad \text{(III-51)}$$

They are then minimized in order to replace the original data with a smaller number of more precise representations. In this equation

$$Z = a^2 - b^2$$

where b is another arbitrary reference radial position.

Space does not here permit further elaboration of the details; we note simply that these moments are then related to the c_{ia} (or to corresponding association constants) as follows:

$$\mu_k = \int_0^\infty cz^k \, dz - \int_z^\infty cz^k \, dz = \sum_{i=1}^{n} c_{ia}[k!/(iA'M_1)^{k+1} - I_{i,k}] \quad \text{(III-52)}$$

In mensicus depletion experiments, the correction integrals, $I_{i,k}$ are small compared to the nonexponential terms. They may be evaluated recursively from

$$I_{i,k} = (Z^k/iAM_1)\exp(-iZA'M_1) + kI_{i,k-1}/iAM_1 \quad \text{(III-53)}$$

starting from the relation

$$I_{i,0} = \frac{1}{iAM_1}\exp(-iZA'M_1) \quad \text{(III-54)}$$

The c_{ia} are then obtained by linear algebra since they are the only unknowns in eq III-52. The advantage of this procedure over the previous methods was attributed to the data smoothing inherent in the integration used to obtain the moments and its applicability (via iteration) to the case where M_1 is unknown. In spite of its very real promise, the procedure has not been tested to any appreciable extent with protein self-association reaction data.

5. Method of Constant Concentration Increment

An interesting method developed by Chun and Kim[46] makes use of the fact that when r values are obtained for c values which are multiple integers of a constant concentration increment, Δc, then

$$c_{r_j} = c_a + j\Delta c \quad \text{(III-55)}$$

and eq II-75 simplifies to

$$c_{r_j} = \sum_{i=1}^{n} c_{ia}\phi_{r_j}{}^i\beta \quad \text{(III-56)}$$

where

$$\phi_{r_j} = \exp[A'M_1(r_j^2 - r_a^2)] \quad \text{(III-57)}$$

and

$$\beta = \exp[-BM_1\Delta c] \quad \text{(III-58)}$$

For an n-species system, eq III-56 has $n + 1$ unknowns, i.e., n values of the c_{ia} and β. If $j = n + 1$ measurements are made here, one obtains a sufficient number of equations which may be solved for the unknowns, and from these, k_i and B may be evaluated. This procedure, although seemingly simple, does not employ all the data obtainable from the Rayleigh patterns and is thus subject to more uncertainties than the general curve fitting methods described earlier.

D. Analysis of Self-Association by Use of Apparent Molecular Weight Averages Expressed as Functions of Concentration c

The equations which relate $M^e_{n(c)}$, $M^e_{w(c)}$, and $M^e_{z(c)}$ with concentration through the nonideality parameters are similar in form. However, the relationships between these apparent molecular weight averages and the association constants are all different; furthermore, when it is possible to express a molecular weight average like $M^e_{w(c)}$ as an explicit function of c with the association constant(s) and the virial coefficient(s) as parameters, the different functionalities are found, in general, for each mechanism.

Equation II-43 and eq II-51, II-56, and II-57 are not in themselves particularly suitable for evaluating association constants because writing them in terms of the k_i usually requires introduction of c_i. They must be converted into functions only of c by using either eq II-21 or II-63. Except for two-species systems, the application of eq II-21 is quite complicated, and therefore eq II-63 is generally employed for the purpose at hand.

We describe those procedures which employ appropriate derivatives of molecular weight averages to obtain the various association constants. Then we consider curve-fitting procedures which can be employed with expressions which relate $M^e_{w(c)}$, k_i, and BM_1, with c as the independent variable. Such procedures have been limited up to now to the nonideal monomer–n-mer case and to the isodesmic case. Finally we describe how several different molecular weight averages may be combined to provide for analysis of more general cases.

1. Derivative Procedures

Steiner was apparently the first to use the weight-average molecular weight vs. concentration data of ideal systems to obtain association constants. The procedure, originally developed for the light-scattering experiments, was later applied to the data from the approach to sedimentation equilibrium[47] and sedimentation equilibrium experiments[8,48] with self-associating protein systems. By this time there has been a considerable extension of the analytical methods which involve first of all, the transformation of the original c (or dc/dr) vs. r data. A description of one which is basically simple follows.

When eq II-63 is introduced into eq II-23 we obtain

$$cM_{w(c)}/M_1 = \sum_{i=1}^{n} ik_i(f_1c)^i = f_1c + 2k_2(f_1c)^2 + 3k_3(f_1c)^3 + \ldots \quad \text{(III-59)}$$

Since eq III-59 is a simple polynomial in which (f_1c) is the independent variable and the association constants are the coefficients, the association constants may be evaluated by successive differentiation. For example, the limiting slope of a plot of $[cM_{w(c)}/M_1 - 1 - f_1c]$ vs. $(f_1c)^2$ curve yields the value of $2k_2$. Analogous plots may be constructed to obtain all of the k_i values. It is obvious that a similar procedure may be used for the number average molecular weight which may be defined from eq II-48 and eq II-63, as

$$cM_1/M_{n(c)} = \sum_{i=1}^{n} k_i(f_1c)^i/i = f_1c + k_2(f_1c)^2/2 + k_3(f_1c)^3/3 + \ldots \quad \text{(III-60)}$$

In this derivative process, the evaluation of each of the k_i requires the use of values of all the previous association constants; therefore, the error in its calculation is cumulative. Also, the method relies on limiting slopes, which employ some of the least precise data.

A somewhat different approach was taken by Derechin[49-51] who first expanded eq II-49 for the ideal solution case in powers of c and then applied the multinomial theorem to obtain, finally, the expressions

$$k_2 = -\left[\mathrm{d}\left(\frac{M_1}{M_{w(c)}}\right)\Big/\mathrm{d}c\right]_{c=0} \quad \text{(III-61)}$$

$$k_3 = \frac{3}{2}\left[\mathrm{d}\left(\frac{M_1}{M_{w(c)}}\right)\Big/\mathrm{d}c^2\right]_{c=0} - \frac{1}{4}\left[\mathrm{d}^2\left(\frac{M_1}{M_{w(c)}}\right)\Big/\mathrm{d}c^2\right]_{c=0} \quad \text{(III-62)}$$

$$k_4 = -\frac{8}{3}\left[\mathrm{d}\left(\frac{M_1}{M_{w(c)}}\right)\Big/\mathrm{d}c^3\right]_{c=0} + \left[\mathrm{d}\left(\frac{M_1}{M_{w(c)}}\right)\Big/\mathrm{d}c\right]_{c=0} \times \left[\mathrm{d}^2\left(\frac{M_1}{M_{w(c)}}\right)\Big/\mathrm{d}c^2\right]_{c=0} - \frac{1}{18}\left[\mathrm{d}^3\left(\frac{M_1}{M_{w(c)}}\right)\Big/\mathrm{d}c^3\right]_{c=0} \quad \text{(III-63)}$$

Similar expressions for the k_i in terms of $M_1/M_{n(c)}$ were also obtained. Although this method eliminates the necessity of the determinations of f_1, there is a serious problem inherent in it, namely that of the uncertainty in obtaining successive derivatives of $M_1/M_{w(c)}$ at the limiting concentration of $c = 0$. The above treatment was extended to $M_{z(c)}$ data, but an additional differentiation is now required. Thus, there appears to be no gain in advantage with its application.

Derechin has extended his original treatment to nonideal systems. These procedures were tested by him, for the most part, with synthetic data. A difficulty with this approach is that it requires determination of the slope, curvature, and higher derivations of the $M^e_{w(c)}$ vs. c curves at $c = 0$, quantities which cannot be evaluated with accuracy from experimental data. The fundamental problem is that even though these quantities may be known with some precision at finite concentrations, they become increasingly inaccurate as $c \rightarrow 0$.

2. General Curve-Fitting Procedures

Since eq III-59 and eq III-60 are linear expressions in the unknown quantities k_i, they are amenable to the application of the least-squares method and to the method of linear programming discussed in the previous section. These procedures are also applicable to the quantity $-\psi$ (eq II-57), but the situation with respect to $M_{z(c)}$ (eq II-53) is more complicated. In any event, there is no reason why quantities other than $M_{w(c)}$ should be used for this purpose. Certain advantages of using stepwise polynomial regression analysis have been described by Chun et al.[52] and by Efroymson.[53] Here it should be noted that the value of f_1 must be also evaluated.

The simplest case that may be assigned to this general category is the ideal two-species system in which monomer is in equilibrium with an n-mer. For this case eq II-21 and II-23 may be simplified to read

$$c = c_1 + k_nc_1^n \quad \text{(III-64)}$$

$$cM_{w(c)}/M_1 = c_1 + nk_nc_1^n \quad \text{(III-65)}$$

By combination of these two equations Kegeles and Rao[47] obtained the following general expression

$$\frac{M_{w(c)} - M_1}{[nM_1 - M_{w(c)}]^n} = k_n\left(\frac{c}{(n-1)M_1}\right)^{n-1} \quad \text{(III-66)}$$

When this equation is appropriately plotted, one obtains the

value of k_n from the slope of a straight line. Such a linear plot provides the simplest form of curve fitting. Alternatively, c_1 may be obtained from an expression which is obtained by rearranging eq III-65 to obtain

$$c_1 = c[n - M_{w(c)}/M_1]/(n - 1) \quad \text{(III-67)}$$

Once c_1 values are available as a function of c, then k_n may be evaluated graphically by using either of the following two equations

$$c/c_1 - 1 = k_n c_1^{n-1} \quad \text{(III-68)}$$

or

$$\ln (c - c_1) = \ln k_n + n \ln c_1 \quad \text{(III-69)}$$

These are different forms of eq III-64. Plots of the left-hand side of these equations against either c_1^{n-1} or in $n \ln c_1$ give the value of k_n.[37]

It is, of course, possible to evaluate f_1 by using eq II-66 and to obtain the value of k_n from either of two expressions.[39]

$$(M_{w(c)}/M_1) - (f_1/f_1^n) = nk_n c^{n-1} \quad \text{(III-70)}$$

or

$$(1/f_1 - 1) = k_n(cf_1)^{n-1} \quad \text{(III-71)}$$

These equations may be obtained from eq III-64 and III-65 after the introduction of eq II-63. The method of using eq III-66 and III-67 may be preferred because the uncertain extrapolation to low concentration required for the numerical integration used to obtain f_1 will be avoided.

The situation becomes more complicated when the system is nonideal. Combination of eq II-41 and II-43 gives

$$\frac{cM^a_{w(c)}}{M_1[1 - cM^a_{w(c)}B]} = \sum_{i=1}^{n} ik_1 c_1^{\,i} \quad \text{(III-72)}$$

This equation, upon substitution of eq II-71, may be transformed to give

$$\frac{M^a_{w(c)}}{M_1[1 - cM^a_{w(c)}B]} = \sum_{j=1}^{n} ik_j c^{j-1}[f_1{}^a \exp(-BM_1 c)]^i \quad \text{(III-73)}$$

The procedures of solving nonlinear equations described in connection with the analysis of c vs. r data may be employed here to determine k_i and B_1 simultaneously. For this purpose the B value may be assigned first and then the resulting linear equation may be solved, using the method of least squares. This process may be repeated with varying values of B until the best results are obtained. This general procedure, using a slightly different expression, was first proposed by Van Holde et al.[9] With eq III-73 an additional quantity $f_1{}^a$ must be evaluated for each of the c values where $M^a_{w(c)}$ is known. For this general curve-fitting procedure, over-determined data points are usually used.

Inspection of eq II-43 suggests that if B is first determined then the $M_1/M_{w(c)}$ are immediately available, and the procedures which are applicable to ideal two-species systems already discussed may be used to calculate the association constant. Substitution of eq II-41 and II-71 into eq II-43 yields for the monomer–n-mer case,

$$\frac{M_1}{M^a_{w(c)}} = \frac{1}{n + (1 - n)f_1{}^a \exp(-BM_1 c)} + BM_1 c \quad \text{(III-74)}$$

In eq III-74 the only unknown is B; it may be obtained by successive approximation[25] or by a general curve-fitting procedure.[43,54] Once the value of B is known, the quantities c_1 and f_1 can be calculated by using eq II-71. The quantity $M_{w(c)}$ may be obtained from eq II-43 and the association constant computed.

As is the case for ideal systems (eq III-66), it is possible to convert eq III-72 for a simple monomer–n-mer reaction into a function of one variable, c, using the expressions of mass conservation (eq II-21) without going through a determination of $f_1{}^a$. Combination of eq II-21, II-41, and II-43, when applied to a monomer–dimer reaction gives[10]

$$[2M_1/M^a_{w(c)}] - 1 = (1 + 4k_2 c)^{-1/2} + 2BM_1 c \quad \text{(III-75)}$$

Visser et al.[23] used a nonlinear least-squares fit for the simultaneous estimation of k_2 and B. The "goodness of fit" criteria used was

$$\chi^2 \equiv \sum \frac{1}{\sigma_j^2}[y_j - y(x_j)]^2 \quad \text{(III-76)}$$

where the σ_j are the uncertainties in the data points $y_j = (2M_1/M^a_{w(c)} - 1)$. They used a grid search technique, in which the values of k_2 and BM_1 are systematically changed to minimize χ^2 until a final minimum was reached. Although this procedure may be superior to the general curve-fitting method described earlier (eq III-72), extension of it to the general case does not seem to be feasible. To develop an extrapolation method for the analysis of monomer–dimer self-associations, one may integrate eq III-75 to obtain[55]

$$\int_0^c (2M_1/M^a_{w(c)} - 1)\,dc = \frac{1}{2k_2}(\sqrt{1 + 4k_2 c} - 1) + BM_1 c^2 \quad \text{(III-77)}$$

This expression may be used together with III-75 to eliminate the radical. The result is

$$\int_0^c \left(\frac{2M_1}{M^a_{w(c)}} - 1\right) dc - BM_1 c^2 = \frac{1}{2k_2}\left(\frac{2M_1}{M^a_{w(c)}} - 1 - 2MB_1 c\right)^{-1} - \frac{1}{2k_2} \quad \text{(III-78)}$$

This expression was rearranged to give an equation of the type

$$U = 2/a + bV - W \quad \text{(III-79)}$$

in which U and V are experimentally obtainable quantities, and W is a term which approaches zero as the system approaches ideality, and which is related to the parameters a and b. This equation can be solved by an iterative procedure based on a linear plot. The reader is referred to ref 55 for an exact description of the procedure, and to ref 43 for an evaluation of it.

An alternative procedure which employs eq III-75 in another form has been developed by Deonier and Williams.[55] For this purpose eq III-75 is rearranged to obtain

$$R_a{}^2/[2(1 - BM_1 R_a c) - R_a]^2 - 1 = 4k_2 c \quad \text{(III-80)}$$

where

$$R_a = M^a_{w(c)}/M_1 \quad \text{(III-81)}$$

The quantity on the left-hand side of eq III-80 is divided by c and then plotted against c with varying values of BM_1, until a horizontal straight line is obtained. The intercept between this line and the vertical coordinate divided by 4 yields the value of k_2. This procedure is a direct adaptation of a method originally used for the isodesmic reactions by Van Holde and Rossetti,[16] which will be discussed shortly. For the general nonideal two-species systems, we may obtain

$$\frac{1 - R_a - BM_1 R_a c}{[R_a - n(1 - BM_1 R_a c)]^n}[(1 - n) \times (1 - BM_1 c R_a)]^{n-1} = k_n c^{n-1} \quad \text{(III-82)}$$

Here again the quantity on the left-hand side of eq III-82 may be divided, this time by c^{n-1}, and then plotted against c with varying values of B until a horizontal straight line is obtained.

For ideal isodesmic reactions, eq II-28 and II-44 may be combined to obtain a simple relation[9]

$$(M_{w(c)}/M_1)^2 - 1 = 4kc \qquad \text{(III-83)}$$

A plot of $[(M_{w(c)}/M_1)^2 - 1]$ against c should give a straight line which passes through the origin. From the slope of this line the value of k may be calculated. Although it is possible to substitute eq II-63 into eq II-44 to obtain an alternative equation which may be used for the graphical evaluation of k, as in the case of discrete reactions, it does not provide an advantage over the use of eq III-83.

For the nonideal systems, Van Holde and Rossetti[16] combined eq II-28, II-43, and II-44 to obtain

$$[R_a/[(1 - BM_1R_ac)]^2 - 1 = 4kc \qquad \text{(III-84)}$$

Here again the quantity on the left-hand side of eq III-84, divided by c, may be plotted against c, and B may be adjusted until a horizontal line is obtained. The intercept between this line and the vertical coordinate gives the value of $4k$.

The general curve-fitting procedure adopted by Deonier and Williams for the monomer–dimer case may also be used here after eq III-84 is converted into another form, viz.

$$M_1/M^a_{w(c)} = (1 + 4kc)^{-1/2} + BM_1c \qquad \text{(III-85)}$$

As in eq III-77, an integration is performed to allow elimination of the radical, thus leading to an analogue of eq III-78 and a corresponding variant of eq III-79.

3. Combinations of Molecular Weight Averages

The curve-fitting procedures described in the previous section are adequate when two parameters suffice to describe the self-association, but they become unwieldy when the molecular weights of monomer and n-mer are not known as additional parameters or other additional species are present. By combining different molecular weight averages it is possible to obtain expressions for the more complicated cases in which only one of the unknowns is expressed in terms of experimental quantities. This approach has been extensively developed by Adams.[56]

a. Ideal Systems

We set down expressions relating experimental quantities to the equilibrium constants for an ideal three-component system (monomer–m-mer, n-mer). Equations II-21, II-23, II-48, and II-57 then take the form

$$c = c_1 + k_mc_1^m + k_nc_1^n \qquad \text{(III-86)}$$

$$cM_{w(c)}/M_1 = c_1 + mk_mc_1^m + nk_nc_1^n \qquad \text{(III-87)}$$

$$cM_1/M_{n(c)} = c_1 + k_mc_1^m/m + k_nc_1^n/n \qquad \text{(III-88)}$$

$$-\psi = c_1 + m^2k_mc_1^m + n^2k_nc_1^n \qquad \text{(III-89)}$$

By using Steiner's method[28] one may obtain c_1 which can be substituted into the above set of equations to give a linear system that may be solved for the k_j by simple least-squares procedures. The four-species systems can be treated in a similar fashion.

For the monomer–dimer reactions, with the molecular weights of the species unknown, Dyson[57] combined eq II-21, II-41, II-47, and II-53 to obtain a unique solution:

$$M_i = \frac{1}{2}(\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}) \qquad \text{(III-90)}$$

where $i = 1$ or n, depending on the sign chosen for the above. The quantities α are defined as

$$\alpha_1 = M_{w(c)}[M_{z(c)} - M_{n(c)}]/[M_{w(c)} - M_{n(c)}]$$

$$\alpha_2 = M_{w(c)}M_{n(c)}[M_{w(c)} - M_{z(c)}]/[M_{w(c)} - M_{n(c)}]$$

Once the molecular weights of the two species are found, the concentration of monomer may be calculated by using the expression

$$c_1 = c[M_{2(c)} - M_{w(c)}]/[M_{2(c)} - M_1] \qquad \text{(III-91)}$$

With this quantity, together with c_2 values obtained using eq II-21, the association constant is made available (eq II-19).

b. Nonideal Systems

In the procedures for analyzing nonideal two-species systems (eq III-73), equations have been combined to obtain an expression for B as a function of certain experimental quantities. In an alternative procedure developed by Chun et al.,[39] however, equations are combined, first to eliminate B and then to obtain an expression of f_1 in terms of experimental quantities. Thus eq III-5 and III-6 are substituted into eq III-25 to give

$$\xi = 2f_1\left(1 - \frac{1}{n}\right) + \frac{2}{n} - \frac{1}{n - f_1(n-1)} \qquad \text{(III-92)}$$

This equation is quadratic in f_1; it may be readily solved for f_1. With f_1 known eq III-70 may be used to obtain the value of k_n.

For the nonideal three-species system eq II-43 and II-51 assume the forms

$$M_1/cM^a_{w(c)} = (c_1 + mk_mc_1^m + nk_nc_1^n)^{-1} + BM_1 \qquad \text{(III-93)}$$

and

$$cM_1/M^a_{n(c)} = c_1 + k_mc_1^m/m + k_nc_1^n/n + BM_1c^2/2 \qquad \text{(III-94)}$$

The use of these equations, along with eq III-86 and II-71, to estimate the quantities k_2, k_3, and B has been described in some detail by Adams.[56]

Alternatively eq III-86 and III-93 may be combined to eliminate k_n; the resulting equation reads

$$1/[(M_1/cM^a_{w(c)}) - BM_1] - nc = (1 - n)c_1 + (m - n)k_mc_1^m \qquad \text{(III-95)}$$

Equation III-86 may also be combined with eq III-94 to give

$$n\left(\frac{cM_1}{M^a_{n(c)}} - \frac{BM_1c^2}{2}\right) - c = (n-1)c_1 + (n/m - 1)k_mc_1^m \qquad \text{(III-96)}$$

Equations III-95 and III-96 may then, in turn, be combined to eliminate k_m. When eq II-71 is introduced into the resulting equation, the following expression for B in terms of experimental quantities is obtained

$$1\Big/\left(\frac{M_1}{M^a_{w(c)}} - BM_1\right) - nm\frac{cM_1}{M^a_{n(c)}} - \frac{BM_1c^2}{2} + (m-n)c = (n-1)(1-m)f_1^a c\exp(-BM_1c) \qquad \text{(III-97)}$$

This equation may be solved for B either by using a successive approximation procedure or a general curve-fitting method. Once B is determined c_1 may be calculated from eq II-71. Introduction of B and c_1 into eq III-95 then provides for the evaluation of k_m, and the introduction of k_m and c_1 into eq III-86 should give the value of k_n.

As in the case of the nonideal two-species system, it is also possible to use eq III-92 for the calculation of f_1, thus to provide means to evaluate k_n and k_m without prior determination of b. Equation III-1 may be combined with eq III-89 to obtain

$$-\psi/c = f_1 + m^2f_m + n^2f_n \qquad \text{(III-98)}$$

From eq III-9, III-10, III-12, III-25, and III-98 it is possible to evaluate f_1, from which c_1 may be obtained. With the availability of c_1, eq III-86 and III-9 provide the k_m and k_n data.

A procedure of analysis using both weight- and number-average molecular weights was first developed for isodesmic

systems by Adams and Lewis,[17] who combined eq II-43, II-44, II-51, and II-52 to obtain

$$\frac{M_1}{M^a_{w(c)}} = \left(\frac{M_1}{M^a_{n(c)}} - \frac{BM_1c}{2}\right) \Big/ \left(2 - \frac{BM_1c}{2} + \frac{M_1}{M^a_{n(c)}}\right) + BM_1c \quad \text{(III-99)}$$

This equation may be solved for B either by using successive approximation or general curve-fitting. With the substitution of the B into eq II-43, the value of $M_1/M_{w(c)}$ is readily obtainable and from eq II-84 the value of the intrinsic constant, k, may be computed. It is also possible to obtain first the value of f_1, instead of B, using either eq III-25 or III-26; this method was developed by Chun et al.[39] From eq III-18, III-19, and III-25

$$\xi = 2\sqrt{f_1} - \sqrt{f_1}/(2 - \sqrt{f_1}) \quad \text{(III-100)}$$

This is a quadratic equation in $\sqrt{f_1}$. It has the solution

$$\sqrt{f_1} = [(\xi + 3) - \sqrt{(\xi + 3)^2 - 16\xi}]/4 \quad \text{(III-101)}$$

Alternatively, using the successive approximation, the value of f_1 may be obtained from eq III-19 and III-26 since

$$\eta = [\sqrt{f_1}/(2 - \sqrt{f_1})] - \ln f_1 \quad \text{(III-102)}$$

It follows from the combination of eq II-28 and II-63 that

$$1 - \sqrt{f_1} = kcf_1 \quad \text{(III-103)}$$

Thus, once the value of f_1 is known, a plot of $1 - \sqrt{f_1}$ against cf_1 should give a straight line, with the value of k being obtained from its slope.

In the analysis of nonideal, discrete self-associating systems, it has proven useful to employ several molecular weight averages. In the case of the nonideal three-species system which is discussed here, introduction of these averages makes it possible to reduce nth degree polynomials in several unknowns to forms in one unknown, which can readily be solved. The molecular weight averages other than $M^a_{w(c)}$ were introduced for computational convenience. More than one molecular weight average could also be used for analyzing two-parameter systems (e.g., monomer–n-mer and nonideal isodesmic).

Teller[58] has considered at some length the question of whether such combinations of data, all taken from the *same* actual c vs. r information, are advantageous over those procedures which involve only one average, usually the weight-average molecular weight, as it varies with c along the sedimentation equilibrium column in the cell.

The question is complicated by the different methods of calculating the molecular weight averages and by experimental error. We believe that *independent* measures of $M^a_{n(c)}$, $M^a_{w(c)}$, and $M^a_{z(c)}$, for example, all of the same precision, would provide additional information, since each molecular weight average weights the various species differently (e.g., $M^a_{z(c)}$ gives relatively greater weighting to higher molecular weight species). Suppose, however, that $M^a_{z(c)}$ were calculated from $M^a_{w(c)}$ data as $d(cM^a_{w(c)})/dc$: does this provide additional information? Potentially, one could generate in this way a function ($M^a_{z(c)}$) which more strongly represents the higher molecular weight species. But this information is not independent, since it is already contained in the $cM^a_{w(c)}$ function. Actually, the differentiation step applied to imprecise data yields an even less precise function, and for this reason we believe that there is no advantage to applying $M^a_{z(c)}$ and still higher molecular weight averages to two-species or isodemic systems. If the extrapolation of $M_1/M^a_{w(c)}$ to zero concentration can be made with accuracy, there may be practical advantages associated with using $M^a_{n(c)}$ since random errors tend to cancel when an integration is performed. We believe that the possible advantages accruing from the independent value of $M^a_{n(c)}$ are practical and useful, but that no new information is provided when $M^a_{n(c)}$ values are calculated from $M^a_{w(c)}$ data.

There have appeared in the most recent edition of "The Proteins" [59] two lucid and well-organized short accounts of our subject, one of Van Holde[60] and the other by Klotz, Darnall, and Langerman,[61] related largely to the material in our section II. Both of these reports form excellent introductions to the subject. In addition, more detailed treatments are to be found in the Fujita[21] and Magar[54] books and in the Teller review.[58]

IV. Complications and Experimental Limitations

A. General Considerations

With certain simplifying assumptions, derivations of some basic equations and a description of the analytical procedures which are based upon them now have been presented. While in most cases our assumptions have been shown to possess a reasonable degree of validity, for others corrections and extensions may be required in order to obtain more accurate association constant data. Three of the more pronounced difficulties which require consideration are (a) solution nonideality (already treated by using an approximation which has been demonstrated to be highly acceptable, (b) pressure effects, and (c) the presence of one or more additional components, either inert or not in rapid chemical equilibrium with the main component. In one way or another, properly chosen analytical and experimental procedures may serve to minimize them. Overall, it is felt that for the present the main problem in an analysis resides more in the accuracy of the data themselves and not so much in basic theory or in the several methods which may be used to analyze them.

To this point no estimates of the effect of experimental inaccuracies on the final conclusions have been attempted. It is essential to do this in order to judge the significance of the thermodynamic data which result from the computations; this is especially true in those cases where an appreciable number of species is present.

To provide a broader understanding of the use of the sedimentation equilibrium experiment for the study of protein self-association reactions, mention should be also made of the fact that with it there are two distinct operational procedures: the Wales–Yphantis[62,63] "meniscus depletion" method and what we shall term (for want of a better description) the low-speed, short-column experiment. We note here only that each type of experiment has its advantages and disadvantages. They have been considered at length for the meniscus depletion method by Yphantis. For the second procedure we note that with the Rayleigh optical system to record the equilibrium redistribution of the components, the data provided are the *relative* concentrations of solute species over the cell, a disadvantage. However, there are well-established procedures to obtain c vs. r data (the loading concentration, c_0, must be determined by independent means), and this type of investigation can be conducted over a much wider range of initial solute concentrations, a distinct benefit. Furthermore, using it, there is little if any need for concern about possible pressure effects.

The purpose of this section is to present methods of determining whether one or more complicating effects are present in the system under study and to give procedures of correcting for them. As already mentioned, the requirement of correction for the solution nonideality is generally present. This particular topic is included in the general description of the analytical procedures in sections II and III. But the question remains of how experimental inaccuracies in both primary and ancillary data limit the conclusions which may be drawn about the reaction mechanisms.

B. Thermodynamic Nonideality

It is proposed here to present brief mention of the causes for solution nonideality; it is obvious that if it were possible to predetermine the magnitude of the second virial coefficient B,

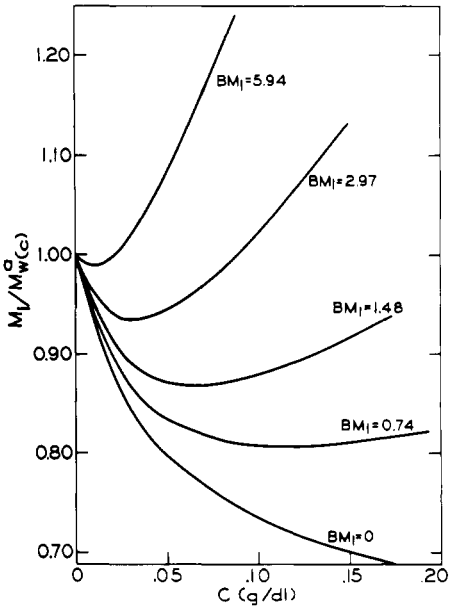

Figure 3. The effect of thermodynamic nonideality on simulated $M_1/M^a_{w(c)}$ vs. c curves for monomer–dimer association: k_2 = 9.8 dL/g. From Godfrey and Harrington.[64]

the whole problem would be greatly simplified. As an example we reproduce in Figure 3 such curves from an article by Godfrey and Harrington.[64] The presence of appreciable nonideality is often indicated at once by plots of $M_1/M^a_{w(c)}$ vs. c data for a simulated monomer–dimer association, with increasing values of the coefficient BM_1. As the coefficient BM_1 becomes larger, the curves start to bend upward with the upward slope increasing; at the same time the concentration at which the minimum value of $M_1/M^a_{w(c)}$ appears is shown to decrease. When the value of BM_1 is very large, the c value for the minimum in $M_1/M^a_{w(c)}$ may become so small that it will be difficult to detect the existence of any self-association reaction at all. At the intermediate values of BM_1 the nonideality effects might pass unnoticed if the experiments are performed only at solute concentrations below the point where the minimum value of $M_1/M^a_{w(c)}$ occurs. For this reason, a cursory examination of a $M_1/M^a_{w(c)}$ vs. c (or $M^a_{w(c)}$ vs. c) curve alone is not always sufficient to determine the presence of the nonideality. Further, the advantage of using data taken over a wide concentration range is at once made apparent.

If the self-association reaction has been otherwise established to be of the simple monomer–n-mer type, the ideal two-species diagnostic plot may be a sensitive method for detecting the presence of solution nonideality.[16] Another practical approach is to make use of the procedures given in the preceding section, allowing for one virial coefficient and determining whether the value of BM_1 so obtained is significant.

Similar methods may be used to determine whether an additional (third) virial coefficient, B_2, is required. As an example, Visser et al.[23] plotted the values of $(1 - f_1)/f_1$ against cf_1 (eq III-71) for data for the system β-lactoglobulin B in acid solutions. For the data from 15 °C or below, the straight-line portion of the plot was either very limited or nonexistent, while data for the higher temperatures gave longer straight lines. Since plots of the data for the experiments at the lower temperatures by using the f_1 values obtained after correction for both B_1 and B_2 gave much longer straight-line portions, it was concluded that the use of a third virial coefficient might be advantageous for the interpretation of the data.

Previously, Albright and Williams[65] used an extended form of the Steiner–Adams equation to evaluate their data for certain acidic systems which contained β-lactoglobulin B as the solute and under certain solution conditions. With the computational methods available at the time, it was concluded that the experimental observations could be represented by three parameters, a dimerization constant and two virial coefficients.

After reexamining these data with the aid of newer procedures for their interpretation, Visser et al.[23] concluded that for all values of the solution ionic strength and pH which had been used it was possible to find a set of two parameters, a k_2 and a BM_1, which provided a good fit of the earlier data, without recourse to a third virial coefficient. Clearly, the adequacy of any given fit depends on the goodness of fit criterion, and one may question whether, at this stage of development, the use of a third virial coefficient can be justified. Actually in the three-component system with 1–1 type supporting electrolyte present, there is reason to believe that the third virial coefficient will be very small.

Under certain very favorable circumstances it may become possible to predict an approximate value for the second virial coefficient. It is generally believed that the main sources of the virial coefficient in the usual aqueous protein systems at sedimentation equilibrium are the excluded volume contributions and the charge effects.

1. Excluded Volume

The effect of excluded volume is relatively small and will not be considered at length.[66] It is obviously a function of the size and shape of the solute molecule; it depends upon the distance of closest approach of the centers of mass of two such units. The problem of computing the magnitude of the effect for a simple globular molecule is straightforward, but when rigid rod-like or flexible polymers are involved the situation is quite different. Too, if the solute units include bound or entrapped solvent, the partial specific volume must be a composite factor, really the effective hydrodynamic quantity. In general, a simplified theoretical treatment is required and more often the precision of a computed excluded volume is hardly sufficient for application to the problem at hand.

2. Charge Effects

The literature is not without its record of attempts to calculate the second virial coefficient which is caused by charge effects. Again because of the necessity of fitting the mathematical argument to an explicit model rather than to real systems, the computed values of BM_1 do not often fall within the range of those which are seemingly required for the interpretation of the self-association equilibrium data.

When a charged protein is under the influence of a centrifugal force, because of the difference in mass between the protein and the counterion, an electrical potential gradient is developed in the ultracentrifuge cell; the net result is a reduction of the effect of the centrifugal force on the protein species. (This is just opposite to the situation where the counterions "pull" or "drag" the charged proteins along during the diffusion process, to make the diffusion coefficient of the charged protein larger than that when the protein unit bears no net charge.[67,68]) Thus when the charged protein in water is subject to a sedimentation equilibrium experiment, the charge effect brings about a redistribution of the protein such that the apparent molecular weight which results is smaller than the true quantity. When sufficient electrolyte is added to this system, charge effects are largely suppressed and the protein is no longer constrained to be transported together with its counterions in order to maintain macroscopic electroneutrality, and both sedimentation and diffusion coefficients approach those of the neutral protein. Thus the apparent molecular weight obtained under this condition should be close to the true value.

This mutual influence of electrolytes on the sedimentation equilibrium in ternary systems (water, protein (PX_z), supporting electrolyte) has been studied at length. Following the very early observations of Svedberg,[69] Tiselius,[70] and Pedersen,[71] a number of investigators have addressed themselves to the problem. Of them we mention Lamm,[72] Mijnlieff,[73] Johnson et al.,[74] Williams et al.,[20] Vrij and Overbeek,[75] Casassa and Eisenberg,[22] and Roark and Yphantis.[76]

In a succession of recent advances due largely to Scatchard, Vrij and Overbeck, Casassa and Eisenberg, and Roark and Yphantis, the study of the effect of the binding of solvent molecules and ions has been given most careful scrutiny. They have established that the sedimentation equilibrium equations for use with multicomponent systems—written with $(\partial\rho/\partial c)_\mu$ to replace the conventional $(1 - \bar{v}\rho)$ in the basic equation for molecular weight—reduce in form to that of the much simpler equations for two-component systems to yield directly the true molecular weights. The point-average molecular weights which are used in the interpretation of the self-association reactions are no exception; they are just as reliable as the more common variety, taken over the whole cell. In those cases where the Vrij and Casassa–Eisenberg procedures are not possible, reference may be made to Williams et al.[20]

In the Roark and Yphantis development, the Scatchard definition of components has been extended to polydispersed systems of macromolecules, and to self-association systems, in which each species possesses identical charge-to-mass ratios. The algebra involved is rather cumbersome so that here only the final equation is reproduced. It is:

$$\frac{1}{M^e_{w(c)}} = \frac{1}{M_{w(c)}} + 2B_1c + 3B_2c^2 + 4B_3c^3 + \ldots \quad \text{(IV-1)}$$

with the coefficients B_1, B_2, and B_3 being given in terms of available significant quantities.

It is an interesting consequence of the Roark–Yphantis development that when a 1–1 type salt is added as the supporting electrolyte, the third virial coefficient vanishes, and when a 1–2 or 2–1 salt is added, the fourth virial coefficient approaches zero. The maximum protein concentrations below which the several virial coefficients become negligible were also estimated.

C. Pressure Effect

The theoretical relationships which have been considered to this point are based upon the assumption that the interacting system is incompressible during the ultracentrifugation process. For the low-speed, short-column case this approximation is valid; this point has been generally accepted. For the meniscus depletion method, where the angular velocity often reaches as high as 40 000 rpm, this assumption may not be justified. The general subject of corrections for pressure has been investigated by Young et al.[77] For the self-association system, if the value of the molar volume, V, changes during the reaction, the total volume change is appreciable, and the principle of Le Chatelier dictates that the equilibrium constant will be shifted in the direction of decrease in the total volume; this effect should increase with increasing pressure. Therefore, the association "constants" are no longer constants, but are functions of total pressure and are thus functions of the radial distance, r. That such a pressure effect may be appreciable during the high-speed ultracentrifugation of a reacting system with 10-mm solution columns has been gradually realized.[78–83] For instance, a dramatic pressure effect has been described by Kegeles et al. (Table I of ref 78). Although some of the values of ΔV listed in their discussion may be regarded as being extreme, values of the order of a few hundred ml/mol are not altogether unreasonable.

An interesting study of the possible pressure effect on the sedimentation equilibrium distribution of interacting solutes was carried out by Kegeles et al. for the heterogeneous reactions. Howlett et al.[81] have described comparable studies on self-association reactions. Any appreciable pressure effect will cause the failure of the apparent average molecular weight data taken at several different initial loading concentrations and at different rotor speeds to superimpose to form a single smooth curve of $M^e_{w(c)}$ vs. c. This arises since at different initial concentrations or speeds, a given concentration will appear at different r values. Thus, association constants corresponding to these points will be apparent quantities which are functions of the pressure.

The effect of pressure on the equilibrium constant may be described by the following equation[82]

$$K(r) = K_{r_a} \exp\left(-\int_{P_{r_a}}^{P} r\frac{\Delta V}{RT}\,dP\right) \quad \text{(IV-2)}$$

Upon integration, with the assumption that ΔV is independent of P, and remembering that $P_r - P_{r_a} \cong 1/2[\rho\omega^2(r^2 - r_a^2)]$, this statement becomes

$$K(r) = K_{r_a} \exp\left\{-\frac{\Delta V}{RT}\frac{\rho\omega^2}{2}(r^2 - r_a^2)\right\} \quad \text{(IV-3)}$$

In these equations, $K(r)$ is the association constant at r where the pressure is P_r; $K(r_a)$ is the association constant at the meniscus. The pressure at this point, P_{r_a}, is equal to the atmospheric pressure; K_{r_a} is equivalent to the association constant in the absence of the pressure effect. When eq II-19 is introduced into eq IV-3,

$$\frac{c_n}{c_1^2} = \frac{c_{na}}{c_{1a}^2} \exp\left\{-\frac{\Delta V}{RT}\frac{\rho\omega^2}{2}(r^2 - r_a^2)\right\} \quad \text{(IV-4)}$$

For a monomer–n-mer system, the law of the conservation of mass must hold, so that

$$Q = Q_1 + Q_n \quad \text{(IV-5)}$$

where Q is the total amount of the solute in the cell, and Q_i is the total amount of species i in the cell, which is given by

$$Q_i = h'\theta \int_a^b c_i r\,dr \quad \text{(IV-6)}$$

Integration of this equation yields

$$Q_i' = c_{ia}h'\theta[1 + \exp\{iA_iM_1(a^2 - b^2)\}]/2iA_i'M_1 \quad \text{(IV-7)}$$

In these equations h' is the cell thickness and θ is the cell sector angle in radians. Now

$$\frac{Q_n}{Q_1^n} = \frac{c_{nb}[1 - \exp\{nA_nM_1(r_a^2 - r_b^2)\}](h'\theta)^{1-n}(2A_1M_1)^n}{c_{1b}^n[1 - \exp A_1M_1(r_a^2 - r_b^2)]^n 2nA_nM_1} \equiv \chi \quad \text{(IV-8)}$$

The ratio c_{nb}/c_{1b}^n in eq IV-8 may be obtained by using eq IV-4 with $r = b$, thus permitting the evaluation of the factor χ. From eq IV-5 and IV-8 it follows that

$$\chi Q_1^n + Q_1 - Q = 0$$

which may be solved for Q_1. From eq IV-7 the c_{ia} may be evaluated; thus the application of eq II-69 should give c_1. From eq II-73, IV-5 and IV-7 one may obtain c_n, and the use of eq II-41 gives the value of $M_{w(c)}$. The result of such a calculation for the monomer–dimer case (Figure 2 of ref 83) indicates that the failure to form a single continuous curve of $M_{w(c)}$ vs. c data by several experiments at different rotor speeds may be caused by the pressure effect.

Howlett et al.[83] have extended their earlier treatment to nonideal systems; certain of the results are presented in Figure 4. Here the computed values are compared with their own experimental values which were obtained for lysozyme solutions. Experiments with different initial loading concentration and rotor speed did indeed give discrete curves.

Harrington and Kegeles[79] pointed out that for the meniscus depletion method, the region where the concentration gradient is large, thus leading to large errors in the data, is also the region where an appreciable pressure effect may be present. Thus any deviation from the behavior expected for no pressure effect may very well arise from experimental errors. Even though a method of correction for the pressure effect is available, it may not always be clear when the method should be applied. Perhaps the surest way of detecting a pressure effect is to perform the experiments at several different rotor speeds, and without dissembling the cell.

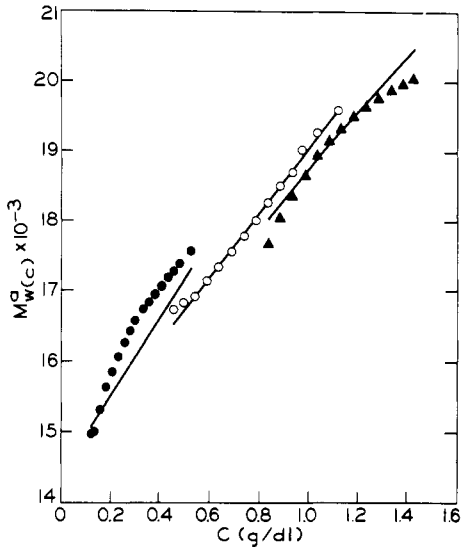

Figure 4. The apparent weight-average molecular weight of lysozyme in diethyl barbiturate buffer at pH 8.0, $I = 0.15$ and at 15 °C: ●, initial concentration 0.2804 g/dL, 20 000 rpm; ○, initial concentration 0.7237 g/dL, 15 000 rpm; ▲, initial concentration 1.1068 g/dL, 11 000 rpm. The solid lines were computed employing values of $k_2(a) = 0.24$ dL/g, $B = -4 \times 10^{-6}$ dL/mol of monomer of molecular weight 14 400, and $\Delta V = -0.03$ ml/g. From Howlett et al.[83]

D. Effect of Solute Impurity

As has already been indicated one of the characteristic properties of a self-association system is that the apparent weight-average molecular weights, obtained from different experiments at varying initial concentrations and rotor speeds, when plotted against c, must lie on a single smooth curve. However, discontinuous $M^a_{w(c)}$ vs. c curves have been reported in a number of cases even where pressure effects are presumably absent. It has been suggested, and in several instances demonstrated, that solute heterogeneities can also give rise to this type of discontinuity. This conclusion stems from the finding that certain unpurified protein solutions which initially gave discontinuous $M^a_{w(c)}$ vs. c curves later have exhibited continuous smooth curves following rigid purification procedures. The term "impurity" is here defined as any minor solute which is inert with respect to the main self-association reaction.

Teller[43] and Deonier[32] have presented provisional thermodynamic treatments of the situation by considering an ideal dimerizing system where a fraction of the solute is irreversibly aggregated monomer. The essence of these treatments is to obtain the concentration of the self-associating component and of the impurity in terms of the original total concentration and the radial distance. From them, the $M_{w(c)}$ values as a function of c are computed. The concentration of the dimerizing component may be obtained from eq II-78 and eq II-82 after the quantity c^0 in the latter equation is multiplied by the fraction of the associating component which is present. The equation of Rinde (eq II-83) gives the concentration of the impurity when the quantity c_i^0 in this equation is substituted by the total concentration of the original solution times the fraction of the impurity.

Deonier calculated the values of $M_{w(c)}/M_1$ as a function of k_2c for several cases, each of which contained different types of inert homogeneous impurities. A representative example is provided by Figure 5. This diagram provides clear evidence that the presence of impurity does indeed give discontinuous $M_{w(c)}/M_1$ vs. c curves when results from experiments performed at different initial concentrations are plotted. He also observed that the curves for individual experiments may be concave upward, concave downward, and even may show inflection points depending on the molecular weight, fraction of the impurity etc.; all of those have been observed with real systems. He noted as

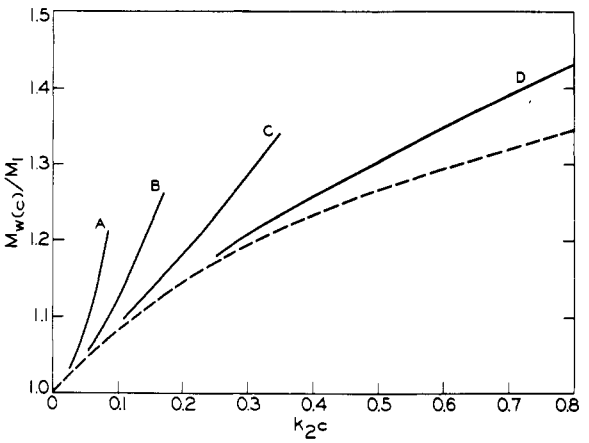

Figure 5. The effect of an impurity to the ratio $M_{w(c)}/M_1$ plotted against k_2c; 2% of the protein is assumed to be present as the impurity with molecular weight of $4M_1$. Here $M_1\lambda = 1$. In each curve, the values of k_2c_0 are: (A) 0.05, (B) 0.10, (C) 0.20 and (D) 0.50. The broken line represents the value of $M_{w(c)}/M_1$ that would be observed if the solute has no impurity. From Deonier.[32]

well that the several experimental curves lie *above* the $M_{w(c)}/M_1$ vs. k_2c curve that has been obtained for the corresponding system from which the impurity had been removed. Thus he has questioned the propriety of analyzing data on the basis of a polynomial fitted *through* the actual data points in some manner as has been done repeatedly in the literature.[84,85] Clearly, the pattern of such discrepancies will be more evident if the experiments are performed at a single rotor speed and at constant solution depth.

E. Some Observations about Obtaining the Primary Data

It has been already mentioned that the primary data from the equilibrium ultracentrifugal study of a self-associating system are, according to eq II-38 and II-73, either n or dn/dr as a function of r.

Since there is no direct way to determine the concentration of the several solute species following the attainment of equilibrium, indirect methods must be employed. Of the refractometric techniques, the most commonly used ones are the Schlieren and the Rayleigh optical systems. The much more sensitive Rayleigh procedure was first successfully applied to diffusion studies,[86-88] and it was later adapted to ultracentrifugation,[89-92] leading to a dramatic increase in the precision of the sedimentation equilibrium experiment. It is this increase in the precision, along with the decrease in time required to approach sedimentation equilibrium by using short solution columns, which has brought about a strong resurgence in the use of this method in general. To a large extent, it has been these two factors which have made possible a quite good analysis for self-association systems.

The evaluation of the Rayleigh pattern to provide solute concentration vs. radial distance in the cell (c vs. r) data has been described in several places. A particularly lucid description of the subject is to be found in a treatment of the sedimentation equilibrium methods by Van Holde.[93] In another useful source Schachman[94] has described it, along with several other optical systems, with directions for their use. (For instance, an early account of the automatic direct recording, scanner absorption device is here included.)

F. Determination of Monomer Molecular Weight

In the study of self-association reactions, a predetermination of the monomer molecular weight, while not actually required, is an important consideration. This quantity may be obtained separately or it may evolve during the course of the analysis of

sedimentation equilibrium data for a self-associating system. In principle, a plot of the $M^{a}_{w(c)}$ vs. c curve should yield the value of M_1 at infinite dilution. However, the customary extrapolation procedures are often hazardous, particularly for the systems with strong association tendencies, because the error in $M^{a}_{w(c)}$ at low concentrations is large, and can lead to errors in M_1 as large as 100%.

The monomer molecular weight, M_1, may be obtained during the curve fitting process with c vs. r plots. An analysis by using synthetic data has indicated that the estimated value of M_1 may agree to within about 1% of the true value.

For some proteins, solution conditions can be found in which the protein exists entirely in the monomer form. In this case the value of M_1 can be obtained in the conventional way by determining several $M^{a}_{w(c)}$ values at different concentrations and then extrapolating these data to infinite dilution in a separate and carefully controlled series of experiments.

Another, but less common, procedure is to perform the sedimentation equilibrium experiment in 4 M guanidine hydrochloride or in a 8 M urea system.[95-96] If an appreciable amount of second solvent is specifically bound to the solute in a ratio which is different from that of the bulk solution, the apparent change in density may introduce a significant error unless suitable experimental and computational methods are applied.

For chymotrypsinogen A, by using a new and ingenious experimental approach, La Bar[97] did observe an agreement to within 1% between the value of M_1 from the amino acid content and from his sedimentation equilibrium experiment.

G. Propagation of Errors: Computation of Average Molecular Weights and Other Quantities

1. Weight-Average Molecular Weight

The value of the quantity $M^{a}_{w(c)}$, a point average, may be taken directly from sedimentation equilibrium experiments performed at several different initial solute concentrations. The $M^{a}_{w(c)}$ data as a function of concentration are calculated by using the basic sedimentation equilibrium equation, eq II-38. Before the Rayleigh optical system came into general use, the molecular weight data were often taken from the slopes of ln c vs. r^2 plots, but computed in this fashion, they are hardly accurate enough for the study of self-association behaviors when Schlieren patterns are used. In terms of Rayleigh fringe data, eq II-38 may be rewritten

$$M^{a}_{w(c)} = \mathrm{d}[\ln (J_a + \Delta J)]/2A\mathrm{d}\left(r_a + \frac{\Delta r}{G}\right)^2$$

so instead of the ln c vs. r^2 plots, Rayleigh fringe numbers vs. radial distance data are used directly for the evaluation of $M^{a}_{w(c)}$. In this equation

$$c = (J_a + \Delta J)\lambda/hR$$

and

$$r = r_a + \Delta X/G$$

In these two definitions, J_a is the fringe number at the meniscus, ΔJ is the total fringe number, h is the thickness of the solution column, λ is the wavelength of the light used, R (the refractometric increment) = $\mathrm{d}n/\mathrm{d}c$, r_a = radial distance at the meniscus, and $\Delta r = \Delta X/G$, with G being the magnification factor. For the evaluation of the derivative $\mathrm{d} \ln (J_a + \Delta J)/\mathrm{d}(r_a + \Delta X/G)^2$ several different procedures have been described. First of all, this quantity may be obtained directly from the slopes of the ln $(J_a + \Delta J)$ vs. $(r_a + \Delta X/G)^2$ plots; they correspond to the earlier ln c vs. r^2 diagrams. Such plots usually show very little curvature, so Hancock and Williams[98] have sought to increase the accuracy by plotting $\Delta \ln (J_a + \Delta J)$, the difference between the ln $(J_a + \Delta J)$ at a fixed $(r_a + \Delta X/G)^2$ value which lies half-way between the first and last data points, against $(r_a + \Delta X/G)^2$. The derivative of this difference function is then taken at a number of points.

In an alternative approach Van Holde et al.[9] and Teller et al.[37] represented the data by a three of four term polynomial, which could then be differentiated analytically. Cassman and Schachman[99] used the Dyson[57] computer programs to analyze the molecular weight distributions in making a selection of a self-association model for the beef liver glutamic dehydrogenase system. Also, instead of using all the data points from an experiment, Roark and Yphantis[36] have taken five adjacent data points and obtained the slope of the line through these points by using a least-squares process. They found that this slope is equal to the slope at the central point, as calculated from the first derivative of a quadratic representation of the curve through the five points.

As already mentioned, the random error in the determination of J is about ±0.02 fringe. In order to discuss the propagation of this error to the determination of $M^{a}_{w(c)}$, it is convenient to conduct the analysis by using the expression of $(\mathrm{d}J/\mathrm{d}r^2)/J$ rather than its equivalent form of $\mathrm{d} \ln J/\mathrm{d}r^2$. Since the magnitude of scatter in J is independent of the solute concentration at moderate concentrations, it is expected that the extent of the scatter in $\mathrm{d}J/\mathrm{d}r^2$ will be about the same at all concentrations. However, since each measure of this random error is divided by J, the scatter in the value of $(\mathrm{d}J/\mathrm{d}r^2)/J = \mathrm{d} \ln J/\mathrm{d}r^2$ may be expected to be concentration dependent. Thus the random error in $M^{a}_{w(c)}$ at a low concentration is larger than that at a higher concentration. This was demonstrated in a model calculation by Roark and Yphantis, using data which had been obtained by the meniscus depletion method. Roark and Yphantis estimate the random errors in $M^{a}_{w(c)}$ to be about 1% except at very low concentrations, where they may become as large as 6%.

It was mentioned earlier that the error in $(1 - \bar{v}\rho)$ may be as high as 1%. This may be expected to be a major source of the systematic error in $M^{a}_{w(c)}$. Therefore we may conclude that beyond the first several fringes the precision and the accuracy for the weight-average molecular weight data may be as low as 2%, but in regions of high dilution the error will be appreciably enhanced. Added to this will be errors in the initial concentration, when experiments from several different initial concentrations are used to produce a curve of $M^{a}_{w(c)}$ vs. c.

2. Number-Average Molecular Weight

According to eq II-50, the evaluation of the apparent number-average molecular weight involves the computation of the quantity:

$$\frac{1}{c}\int_0^c (1/M^{a}_{w(c)})\,\mathrm{d}c$$

The process of numerical integration may increase the error; it causes a greatly magnified effect at the lower concentrations, as shown by model calculations of Roark and Yphantis.[36] These workers estimated that the random errors beyond several fringes to be about 2%, but at the lower concentrations the error could be even greater than 15%. If experiments from several different initial concentrations are pieced together, errors can arise from the manner in which slight discontinuities are treated. If initial concentration errors are random and a relatively large number of experiments have been performed, such errors will tend to cancel at high concentrations.

3. z-Average Molecular Weight

The evaluation of z-average molecular weight involves the differentiation of the product of $cM^{a}_{w(c)}$ with respect to c. The process of differentiation itself increases the error. The multiplication of $M^{a}_{w(c)}$ by c increases the error, which is roughly proportional to the total solute concentration. Since the error

TABLE V. Sedimentation Equilibrium (and Approach to Sedimentation Equilibrium) Studies of Some Selected Self-Association Reactions

System	Solvent	t, °C	M_1	Stoichiometry	Ref
Purine		4.9, 9.9, 14.9, 19.9, 24.9	120	Indefinite	16
Cytidine		25	243.2	Indefinite	9
ATP	NaCl (0.154 M)	10, 16, 20, 25, 30	492.7	1–2–3	111
Glucagon	0.2 M phosphate, pH 10	25	3 500	1–2–6	84
Hemerythrin	Tris–cacodylate, pH 7.0, I = 0.15	5, 25	13 500	1–8	109
Lysozyme	0.15 M NaCl–0.005 M NaH_2PO_4, pH 6.7	15, 25	14 300	1–2	27a
	Na cacodylate, I = 0.20, pH 6.7	25	14 300	1–2 or isodesmic	55
	0.005 M NaH_2PO_4, 0.005 M Na_2HPO_4, 0.15 M NaCl, I = 0.17, pH 6.7	15	14 400	1–2 or 1–2–3	40
β-Lactoglobulin A	0.1 M NaCl + 10^{-3} DTT	5, 9.8, 20	18 422[a]	1–2	104
β-Lactoglobulin B	NaCl–HCl, pH 2.58, 2.20, I = 0.1, 0.15	25	18 400	1–2	65
	NaCl–glycine, pH 2.64, I = 0.16	5, 10, 15, 25, 35.5	18 336	1–2	23
β-Lactoglobulin C	0.2 M glycine, pH 2.46, I = 0.1, 0.2	10, 16, 20, 25	18 344	1–2	107
	0.1 M, 0.2 M acetate, pH 4.65, I = 0.1, 0.2	23	18 344	1–2	108
α-Chymotrypsin	Na_2HPO_4–NaH_2PO_4, pH 6.2, I = 0.2	20–25	(23 000)	1–2–3	47, 27
	Tris·HCl, KCl, pH 8.3, I = 0.05	25		Indefinite	129
	NaCl–NaAc, pH 2.8–5.5	25	25 600[a]	1–2	120
	0.178 M NaCl–0.01 M HAc, pH 4.12	0, 5, 10, 15, 20, 25, 30, 35		1–2	123
Chymotrypsinogen A	Vernal buffer, pH 7.9, I = 0.03	25	25 600	1–2–3 or Indefinite	98
Aldolase[b] (rabbit muscle)	0.2 M Tris·HCl, 1.2 M $MgCl_2$, pH 7.2	20	40 500	1–2–4	130
D-Glyceraldehyde 3-phosphate dehydrogenase (rabbit muscle)	0.1 M Tris, pH 7.0	5	72 000	2–4	131
Glutamic dehydrogenase (bovine liver)	0.05 M Tris–0.1 M NaCl, 0.128 M EtSH, EDTA, pH 6.65, 7.2, 8.0	4	324 000	1, 3, 1–2–3	99
Myosin	0.5 M KCl–0.2 M K_2HPO_4, pH 7.3	6	458 000	1–2	64

in $M^a_{w(c)}$ at low concentrations is larger than it is at higher concentrations, the net result is that the scatter in $M^a_{z(c)}$ at low concentrations as well as at high concentrations is larger than the scatter at the intermediate ones. This effect is also apparent in the calculation of Roark and Yphantis who estimated the error in the intermediate concentration range to be about 5%. It is well to note, however, that these error estimates are based on simulated data, and that errors may be enhanced for actual experiments.

4. Other Experimental Quantities

Inspection of eq II-67 indicates that the scatter in the quantity $-\psi$ is about the same as that of the z-average molecular weight, but the scatter at very high concentration should be much larger. Also eq II-67 suggests that the scatter in the quantity in $f_1{}^a$ will be about the same as it is in the case of the number-average molecular weight.

V. Tests of the Several Computational Methods with Experimental Data

In testing the application of the several computational methods with experimental data some arbitrariness of choice of system is inevitable. The approach we have adopted is to consider in reasonable detail some selected records of experiments which illustrate different types of systems and instances where special problems are presented. For example, hemerythrin is quite unique in its mode of association; ATP presents an unusual situation because of its high charge; and lysozyme, purine, and chymotrypsinogen A are substances which are subject to weak associations. The genetic variants of β-lactoglobulin have been popular subjects of investigation, largely by light-scattering and sedimentation equilibrium data, but often with conflicting interpretations.

All of the substances here considered are biological materials which are amenable to a high degree of purification, an advantage for quantitative studies. Some possible systems have been eliminated from the survey since their stoichiometry is too complicated to decipher in view of the present limitations in the accuracy of the data or the purity of the solute; insulin is a case in point—others are not included simply because of space limitation.

In Table V we present a list of selected substances for which relatively complete studies have been made. (Certain information which has been obtained by using the approach to sedimentation equilibrium method is also included in this table.) It is to be understood that they represent but a very small portion of all the substances which are known to undergo dissociation–association reactions. And even in the case of those systems listed, the experiments were performed under a limited set of conditions, and in order to understand better the nature of the self-association reactions, the investigation may have to be carried out over wider ranges of solution compositions and temperatures.

To this point we have considered the general procedures by which the stoichiometry and association constants of self-associating systems are obtained. In addition to this, the errors involved in the procurement of the primary and secondary data have been the subjects of a cursory analysis. It is thus a purpose of this section to test the application of certain of the procedures which have been described to study the analysis of actual experimental data and to estimate the errors in the stoichiometry and the association constants which have been obtained.

Before beginning brief descriptions for the self-association behaviors of each of the selected biological materials, it is important to remind ourselves that a goal of the experiments is to provide some degree of thermodynamic and stoichiometric understanding of the reaction. As indicated, the equilibrium constants and second virial coefficients which characterize the behavior in any system are dependent upon the conditions, pH, ionic strength, and temperature of the aqueous solutions in which they are dissolved; even the chemical nature of the buffer requires specification. Then, there are accessible criteria to ascertain whether the associating material is pure, whether pressure effects are small enough to be neglected, and whether the reaction is a rapidly reversible process; all must be applied

TABLE VI. Association Constants and Virial Coefficients of β-Lactoglobulin B at Different Temperatures

	AFVR method[a]		Least-squares procedure[b]		Chun et al.[c]
t, °C	BM_1	k_2, dL/g	BM_1	k_2	k_2
10	0.105	53.5 ± 15.1	0.105	52	42.3
15	0.109	38.9 ± 10.1	0.110	40	
25	0.110	23.1 ± 2.5	0.108	22.0	21.4
35.5	0.111	10.4 ± 0.8	0.111	10.2	

[a] Equation III-80. [b] Equation III-75. [c] Equation III-70.

before any mathematical interpretation of the data can be meaningful.

The order of presentation is arbitrary, with the simpler cases receiving the earlier attention.

A. The β-Lactoglobulins A, B, and C

The self-association behavior of solutions of three of the genetic variants (A, B, and C) of β-lactoglobulin has been rather extensively studied by the sedimentation equilibrium method. In three subsections below we survey briefly representative accounts of researches with each one of these variants.

1. β-Lactoglobulin A (β-A)

As is the case for the B and C genetic variants, the association–dissociation behavior of β-A is responsive to the pH of the solution. Below pH 3.5, the nearly spherical monomer, approximately 18 400 in molecular weight, is in rapid chemical equilibrium with the dimer. The sedimentation equilibrium studies of Tang and Adams[100] of β-A at pH 2.46 (in 0.2 mol/L glycine) and at several temperatures are consistent with a monomer–dimer association mechanism. Using this model, they obtained values of the association constants and the second virial coefficients at 11, 16, 20, and 25 °C.

At higher pH values the self-association behavior for this protein becomes more complicated. From the earlier light-scattering and sedimentation transport studies, Timasheff and Townend[101] had concluded that 90% of β-A dimer is in equilibrium with octamer between pH 3.7 and 5.2 and at several temperatures (4.5, 8, 15, and 25 °C). More careful light-scattering studies in this pH region were later carried out under the same solution conditions by Kumosinski and Timasheff.[102] They concluded that there is no inert component in β-A as had been earlier suggested and that the data can be best described in terms of a monomer–dimer–tetramer–hexamer–octamer reaction. However, the results of sedimentation equilibrium studies by Adams and Lewis[17] at pH 4.6 in acetate buffer at 16 °C have indicated that the monomer apparently undergoes indefinite self-association. Thus decision as to the mode of association of β-A between pH 3.7 and 5.2 remains unresolved, a perhaps not surprising conclusion because to differentiate between a 1–2–4–6–8 reaction and an indefinite association requires exceedingly precise data. A pertinent discussion of this matter, with additional references, has been presented by McKenzie and Nichol.[103]

Above pH 5.2, the β-A species larger than dimer disappear. The monomer–dimer reaction was studied in 0.1 M NaCl–10^{-3} M dithiothreitol solution (pH 5.2) and in Na_2HPO_4–KH_2PO_4 buffer solution (pH 7.0, $I = 0.1$) at 5, 9, 8, and 20 °C by Kelly and Reithel,[104] again with the sedimentation equilibrium method.

2. β-Lactoglobulin B (β-B)

Although we shall be mainly concerned here with a discussion of the data (in NaCl–glycine buffer, pH 2.64, $I = 0.16$) obtained by Visser et al.,[23] earlier sedimentation equilibrium experiments by Albright and Williams[65] at the single temperature of 25 °C and under slightly different solution conditions (in NaCl–HCl, pH 2.2 and 2.58, $I = 0.1$ and 0.15), had established that (1) this is a self-associating system from the fact that the values of $M^a_{w(c)}$ show a rapid initial increase with concentration, (2) the solution nonideality is substantial because the $M^a_{w(c)}$ values at the higher concentrations reach a maximum value which is much smaller than the molecular weight of the dimer and subsequently decrease with increasing concentration, (3) the reaction is rapid, and (4) the effects of solute impurity and pressure are negligible. Their analyses[65] also supported the stoichiometry of monomer–dimer reaction which was originally deduced by Townend[105] et al., using light-scattering determination in acid solutions.

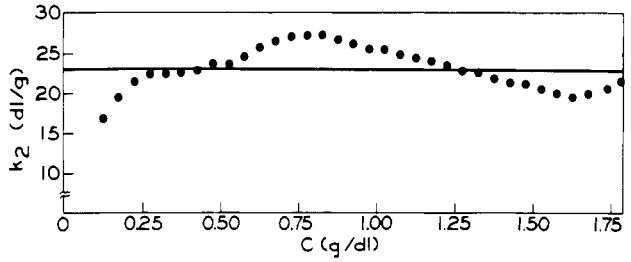

Figure 6. The association constant k_2 of β-lactoglobulin B (pH 2.64, $I = 0.16$) as a function of concentration. These values were calculated from the interpolation curve at 25 °C by use of the AFVR method (eq III-72). From Visser et al.[23]

The experiments of Visser et al.[23] to study the self-association of β-B were conducted at 5, 10, 15, 25, and 35.5 °C; the actual $M^a_{w(c)}$ vs. c data at these temperatures are presented in the original report. Plots of these data at 15 and 25 °C have been reproduced in Figure 1. The monomer molecular weight, M_1, was taken to be 18 336 g/mol on the basis of amino acid composition.[106] Although diagnostic plots such as proposed by Chun et al.[39] were not yet available for this system, the earlier studies by light-scattering and sedimentation equilibrium had strongly indicated that a monomer–dimer process is involved. If one assumes that an uncomplicated nonideal monomer–dimer reaction is involved, there are two quite simple general methods available for the evaluation of k_2 (cf. section III). One of these methods is based on the use of $M^a_{w(c)}$ vs. c data with the application of the Adams and Fujita equation,[10] in either of two different forms (eq III-75 and III-80), and a second general procedure combines the quantity $M^a_{w(c)}$ with another quantity which is derived from it, either $M^a_{n(c)}$ or $f_1{}^a$. The quantity BM_1 may be first evaluated by using eq III-74, or this may be eliminated at once by using eq III-25 (or eq III-26).

Of these four specific analyses, three were tested by Visser et al. In the first one in which eq III-75 was employed with eq III-76 as the "goodness of fit" criterion, a least-squares procedure was used to minimize χ^2 with respect to each of the parameters, k_2 and BM_1, simultaneously. The value of σ is taken to be 350 g/mol. The final results are summarized in Table VI. For the method which utilizes eq III-80 (AFVR method), data taken from the interpolation curves of Figure 1 were employed for the calculation of k_2. The value of the quantity BM_1 was adjusted until a least-squares fit of k_2 vs. c to a straight line yields the value of slope zero. In Figure 6 is presented an example of this type of plot.

Some of the results which were obtained from these calculations are also included in Table VI. An estimate of errors given in this table may be obtained from Figure 6. It is to be noted that the agreement between the results from the two analytical methods is very good, certainly well within the estimated errors. This is to be expected since the two methods make use of variants of the same fundamental equation. For the same reason the errors of these two methods may be expected to be about the same. The random errors in k_2 given in Table VI are generally somewhat larger than 10%. This is the value to be expected from an error of 2% in $M^a_{w(c)}$.

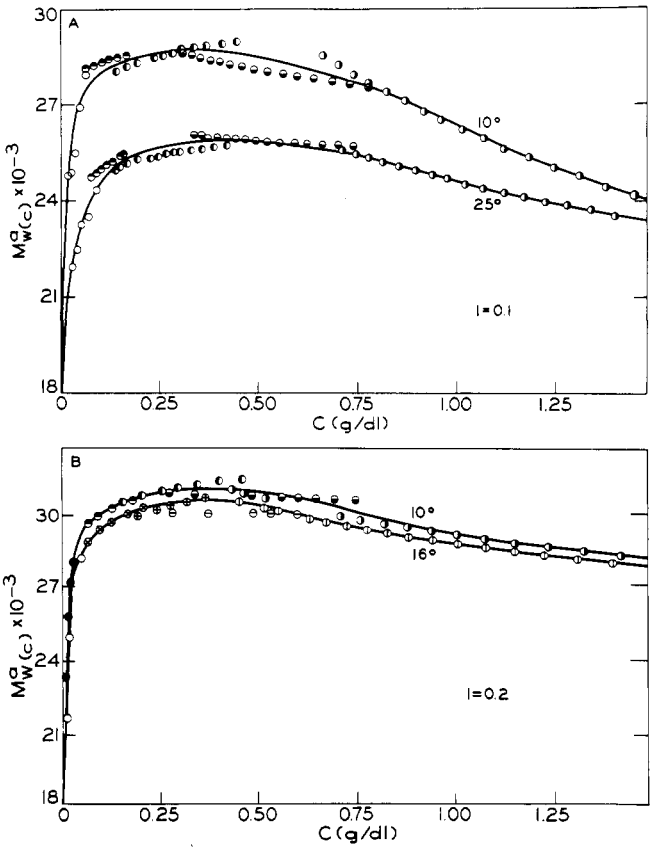

Figure 7. The apparent weight-average molecular weight of β-lactoglobulin C in glycine buffer (pH 2.45) plotted against concentration data at 10, 16, and 25 °C. Curves A: I = 0.1, 10 and 25 °C; Curves B: I = 0.2, 10 and 16 °C. From Sarquis and Adams.[107]

The third method, the graphical one which was used by Visser et al.[23] is a procedure developed by Chun et al.,[39] and in which $M^a_{w(c)}$ and $M^a_{n(c)}$ data are combined (eq III-25).[36,37] First of all, eq III-25 and eq III-92 are used to evaluate f_1 as a function of c; then the quantity $(1 - f_1)/f_1$ is plotted against cf_1 (eq III-71, n = 2) to obtain a straight line which should pass through the origin. The value of k_2 is then obtained from the slope of this straight line. Visser et al.[23] observed that the $(1 - f_1)/f_1$ values for the lowest temperature provided a linear plot over a restricted domain, and that the curves did not pass exactly through the origin. At the higher temperatures, the linear portions were more extended but still they did not show the proper limiting behavior. It was found that the reconstructed $M^a_{w(c)}$ vs. c data, taken from values of k_2 and BM_1 obtained from the linear portion of a given curve, deviated appreciably from the experimental curve in the region of lower concentrations. The values of $M^a_{w(c)}$ were then readjusted by a correction to the number of fringes within this low concentration range, and new values of $M^a_{n(c)}$ were calculated by using these new apparent weight-average molecular weights. When the newly calculated f_1 values were used for the plots of $(1 - f_1)/f_1$ against cf_1, the curves so obtained were almost linear over the whole concentration range and they did pass through the origin. The association constants obtained in this way are also presented in Table VI. It is seen that these values agree with those obtained by using the AFVR method, within the estimated errors of the latter method. Since the random error in obtaining $\ln f_1^a$ is about the same as that for $M^a_{n(c)}$, the procedure of combining $M^a_{w(c)}$ and $\ln f_1^a$ should give about the same error in k_2 as is the case for those analyses in which both $M^a_{w(c)}$ and $M^a_{(c)}$ data are utilized. Also it is expected that the procedure of combining $M^a_{w(c)}$ and f_1^a to obtain the value of BM_1 by successive approximation (eq III-74)[25] should provide for about the same accuracy as in the procedures just discussed.

This discussion suggests that the procedures which derive from the $M^a_{w(c)}$ vs. c data alone are probably at least as precise

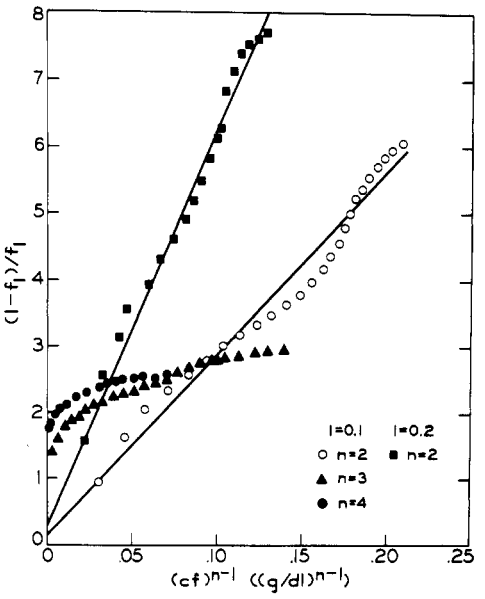

Figure 8. Test of monomer-n-mer association of β-lactoglobulin C in 0.2 mol/L glycine buffer (pH 2.46, I = 0.1 and 0.2) at 10 °C. Evaluation of the association constant, k_n, for monomer-n-mer associations (n = 2, 3, 4) from a plot based on eq III-64. From Sarquis and Adams.[107]

as those which employ either $M^a_{n(c)}$ or $\ln f_1^a$ vs. c data in combination with $M^a_{w(c)}$. Actually the latter methods seem to suffer from the cumulative errors (arising from integration of the quantity $M_1/M^a_{w(c)}$ in the very low concentration region).

The curve-fitting method and AFVR method give comparable results, and the basis of any choice between them might be convenience. For systems where the mode of association is rather uncertain, the AFVR method may be preferable because graphical displays such as Figure 6 may quickly reveal unsuspected trends in the data.

3. β-Lactoglobulin C (β-c)

Sarquis and Adams[107] have reported sedimentation equilibrium experiments for solutions of the β-C variant in 0.1 and 0.2 mol/L glycine buffers (pH 2.45) at 10, 16, 20, and 25 °C, using both the meniscus depletion and the low-speed-short-column methods. The $M^a_{w(c)}$ vs. c curves for solutions at both ionic strengths and at two of the temperatures are presented in Figure 7. The monomer molecular weight was taken to be 18 344 g/mol, again from the amino acid composition.[106,107] Although the maximum value of $M^a_{w(c)}$ for β-C at 25 °C is much smaller than that of β-B at the same temperature, the plots for these two proteins are very similar in general form. This, in turn, suggests that the β-C system is one of rapid dimerization, with an appreciable second virial coefficient.

The $M^a_{w(c)}$ vs. c data were analyzed only by the method which makes use of eq III-25, III-71, and III-92. Plots of $(1 - f_1)/f_1$ against $(cf_1)^{n-1}$ are given in Figure 8 for n = 2, 3, and 4. It is noted that for both ionic strengths the plot for n = 2 gives the best fit straight line and one which passes through the origin, suggesting that this is a monomer–dimer reaction. In Table VII are presented the association constants and the virial coefficients thus obtained. It is to be noted that the linear plots of Figure 8 pass nearly through the origin without prior adjustment of $M^a_{w(c)}$ in the low concentration range. This behavior is in contrast to that for β-B. The results are summarized in Table VII.

The analysis performed by Sarquis and Adams was based on a smooth interpolation curve drawn through the data points. An advantage of the AFVR method (eq III-80) is that k_2 can be calculated point by point, without the need of an interpolation curve. When we used this method to analyze the original data of Sarquis and Adams, discontinuous curves were obtained, reflecting deviation of portions of the actual data for several concentration

TABLE VII. Association Constants and Second Virial Coefficients of β-Lactoglobulin C in 0.2 M Glycine Buffer (pH 2.45)

t, °C	BM_1, dL/g	k_2, dL/g	K_2, L/mol
	$I = 0.1$		
10	0.146	27.2	24.9
16	0.128	14.3	13.1
20	0.137	11.4	10.5
25	0.123	5.9	5.4
	$I = 0.2$		
10	0.080	60.1	55.1
16	0.080	45.7	41.9
20	0.081	40.1	36.8
25	0.070	27.2	24.9

intervals from the monomer–dimer curve. For ionic strength 0.2 and 10 °C, we find k_2 ranging from 53 to 111 dL/g for the experiment having the lowest initial concentration. At intermediate concentration ranges, k_2 values ranged from 39 to 103 dL/g in one case, and from 62 to 116 dL/g in another. We obtained $k_2 = 67 \pm 20$ dL/g when all available data points were used, together with $BM_1 = 0.080$. The average value for k_2 obtained in this way agrees quite well with the value obtained by Sarquis and Adams (60.1 ± 0.2). The standard deviation is much larger, because prior smoothing of data was not performed.

There are two points to be emphasized. First, the plot of k_2 vs. c is very sensitive to slight discontinuities in the M_w^a vs. c plot. In this case, when experiments at adjacent concentration intervals fail to overlap by 1000 g/mol (about 3% error), the k_2 values at some concentrations can differ by almost a factor of 2. This indicates that very accurate M_w^a data are required even to obtain k_2 with an accuracy of 25–50%. Second, experiments should be performed over a wide concentration range. In this case, the average k_2 calculated from the experiment at the lowest concentration is 88 ± 25 dL/g, while k_2 for the highest concentration range is 50 ± 13 dL/g. Thus, failure to perform experiments over a wide concentration range could lead to serious errors. The AFVR method allows easy and direct calculation of errors in k_2 and provides for direct checks for internal consistency in individual experiments.

At an ionic strength of 0.1 and at 25 °C, the effects of nonoverlap were smaller, and we obtained $BM_1 = 0.143$ dL/g and $k_2 = 6.0 \pm 0.3$ dL/g, which agrees well with the results from Sarquis and Adams.[107] When the smoothed data for 10 °C and ionic strength 0.1 were used for this calculation, a plot similar to that in Figure 6 was obtained, and from this a value of $k_2 = 23 \pm 2$ dL/g resulted, which again agrees well with the value reported.

The association constant for the dimerization was also obtained by us from c vs. r data which were provided by Professor Adams, using eq III-30. The results for ionic strength 0.1 and 25 °C are presented in Figure 9. The upper line represents the behavior for the assumed ideal case, and the lower line is for the case of $BM_1 = 0.123$ dL/g. This value, taken from Sarquis and Adams for β-C under the same conditions, apparently gives a horizontal straight line, again suggesting that the dimerization model is correct. The quantity A used here was also taken from the article by Sarquis and Adams.[107] From this plot one obtains $c_{1a} = 0.155 \pm 0.002$ g/dL. The k_2 value follows directly by using the expression $k_2 = (c_a - c_{1a})/(c_{1a})^2$. The final result is $k_2 = 5.6 \pm 0.2$ dL/g which agrees very well with the value obtained by Sarquis and Adams. The data obtained from other experiments under the same conditions but at different initial loading concentrations, however, gave somewhat different k_2 values. The discrepancies between them were larger than the expected errors. These calculations were repeated for the system at ionic strengths 0.1 and 10 °C. Unlike the results from the AFVR calculation of the data obtained under the same conditions, the c_{1a} vs. r plot gave a reasonably straight horizontal line with $BM_1 = 0.146$. The k_2 value thus obtained from this plot is 20.4 ± 0.5

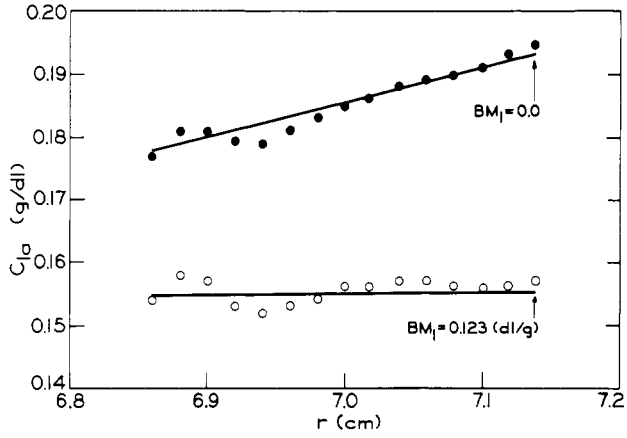

Figure 9. Graphical determination of association constant, k_2, of β-lactoglobulin C in 0.2 mol/L glycine buffer (pH 2.46, $I = 0.1$) at 25 °C, using c vs. r data. This plot was obtained by using eq III-28; and the association constant is readily obtained from the relation $k_2 = (c_a - c_{1a})/c_{1a}^2$.

dL/g. Although this value is somewhat different from the value given by Sarquis and Adams, it agrees reasonably well with our value which was obtained by using the AFVR method.

In comparison, it is to be noted that the direct use of c vs. r data gives, as expected, better internal consistency. However, the discrepancies between the k_2 values obtained under any given condition but at different initial concentrations are larger than can be accounted for by the experimental errors. For the procedures which involve values for $M_{n(c)}^a$ or f_1^a, which are obtained after an integration step, a smoothing of $M_{w(c)}^a$ vs. c data is required. Since one might prefer not to adjust the c vs. r data, the inherently superior method based on such data would be sensitive to even small complications such as the presence of a small amount of impurity, and this would affect the whole analysis. Any tendencies toward discontinuity in the $M_{w(c)}^a$ vs. c curve would be seen immediately and further purification could be undertaken.

In an interesting sequel Sarquis and Adams[108] have described the self-association of β-C in acetate buffer at pH 4.65 at the same four temperatures, 10, 16, 20, and 25 °C, and at ionic strengths 0.1 and 0.2. It was noted that the association of β-C did not proceed beyond the dimer stage, in marked contrast to that of β-A under substantially the same conditions; further, its self-association constant, $k_2 = 2.1 \times 10^3$ dL/g, is independent of temperature and ionic strength. The self-association tendency is found to be much stronger in the acetate buffer as compared to that in the glycine buffers.

B. Hemerythrin

For a two-species association system, it might have been expected that the errors in k_n would increase with increasing values of n, the number of monomers per oligomer, but it seems not to be the case. In order to examine this problem we have elected to consider the self-association studies of hemerythrin, as reported by Langerman and Klotz.[109] In Figure 10 we reproduce their plot of $M_{w(c)}^a$ against c for certain of its solutions. The molecular weight of the monomer is 13 500.[110] From Figure 10 it appears that the monomer associates into an oligomer, or in oligomers of up to octamer. From the initial steep increase in $M_{w(c)}^a$ with increasing c, the presumption is that the association constant is large. (This diagram also indicates that the monomer molecular weight must be obtained by methods other than the extrapolation of the $M_{w(c)}^a$ vs. c curve to $c = 0$. This is primarily because for a monomer–n-mer reaction with large n, the M_w vs. c curve is of sigmoidal shape; if n is very large a "critical micelle" concentration is found.)

The analysis of the weight-average molecular weight data was performed by using the following equation (it is to be noted that the units of C are mol/L and those of k_8 are (L/mol^7))

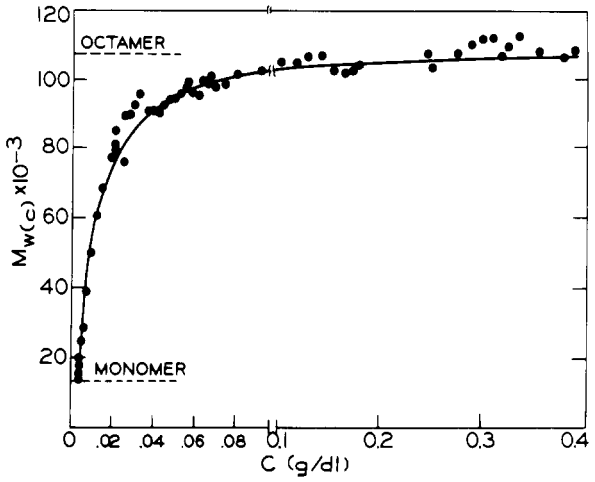

Figure 10. Weight-average molecular weight of azidehemerythrin in Tris–cacodylate buffer (pH 7.0, I = 0.15, T = 5 °C) plotted against concentration. The solid curve was calculated by assuming that only monomer and octamer are in equilibrium. From Langerman and Klotz.[109]

$$M^{a}_{w(c)}/M_1C_1 = 1 + 64K_8C_1^7 \qquad \text{(V-1)}$$

where C is the total molar concentration in monomeric units. This equation is equivalent to eq III-65 for n = 8, with the difference in the concentration units being taken into account. Earlier studies had suggested that the hemerythrin system undergoes self-association according to the scheme monomer–octamer. The system was assumed to be ideal. Thus, a plot of $M_{w(c)}C/M_1C_1$ against f_1C should give a straight line, with the slope of the line being $64K_8$. The C_1 may be calculated using eq II-66 and the relation ln (C_1/C) = ln f_1. A straight line was indeed observed under these conditions, and the value of K_8 from this slope was found to be 3×10^{36} (L/mol). The $M^{a}_{w(c)}$ vs. c values, when calculated by using this value, agreed very well with the experimental values which had been observed at the several solution concentrations. There will always be some difficulty in performing the integration involved in obtaining either $M_{n(c)}$ (eq II-49) or ln f_1 (eq II-69), especially when the initial increase in $M_{w(c)}$ is very sharp.

This problem may be avoided by another analytical approach, in which an initial K_8 value is taken and the values of $M_{w(c)}$ are estimated over the entire concentration range by using the statements

$$C = C_1 + 8K_8C_1^8 \qquad \text{(V-2)}$$

and

$$M_{w(c)} = (M_1/C)(C_1 + 64K_8C_1^8) \qquad \text{(V-3)}$$

Such calculations by Langerman and Klotz were repeated until the best fit was obtained. The K_8 value obtained in this way was 3.4×10^{36} (L/mol)^7. This procedure is essentially the same as the general curve-fitting procedure which was employed by Visser et al.[23] Also substantially the same result might have been obtained by application of the simpler linear plotting procedure developed by Kegeles and Rao.[47]

Langerman and Klotz have estimated that the error in K_8 caused by a 2% error in C_1 (or Cf_1) is, in this situation, about 100%. Calculation of the association constant, using eq III-66, gives about 50% error if the error in $M^{a}_{w(c)}$ is assumed to be about 1%. This great sensitivity arises, of course, from the high power of C which is involved. Since the actual raw data for this system were not available, it was not possible for us to examine the uniqueness of the model of monomer–octamer reaction. However, introduction of either n = 7 or n = 9 into eq II-66 gave values for the association constants which differ from the value for n = 8 by much more than the estimated errors.

C. Adenosine 5′-Triphosphate

As an example of a more complicated three-species system, we consider the self-association reaction of adenosine 5′-triphosphate (ATP). Ferguson et al.[111] studied self-association of its disodium salt (Na_2ATP) in isotonic saline solutions (0.154 mol/L) at 10, 16, 20, 25, and 30 °C. Their apparent weight-average molecular weight vs. concentration curves for the two temperatures, 10 and 20 °C, show discontinuities which are somewhat more pronounced than for those for the other systems which have been discussed to this point. The $M^{a}_{w(c)}$ vs. c data indicate that the self-association process goes well beyond the dimerization stage (M_1 = 492.7).

The analysis for the stoichiometry was thus made first for a two-species system. This was carried out by plotting the values of $(1 - f_1)$ against $c^{n-1}f_1^{n}$ (eq III-71), with $n > 2$ after the value of f_1 had been obtained by using eq III-92. From these plots (Figure 2 of ref 111), a monomer–n-mer stoichiometry was excluded. An isodesmic or indefinite association process was also seemingly eliminated.

Thus, tests for a three-species system and the determination of the association constants were conducted by using eq III-86. It may be rearranged for such a system, to read

$$(c - c_1)/c_1^{m} = k_m + k_nc_1^{(n-m)} \qquad \text{(V-4)}$$

If this stoichiometry is correct, the plot of $(c - c_1)/c_1^{m}$ vs. $c_1^{(n-m)}$ should give a straight line, and from its intercept and slope the two association constants may be obtained. The necessary c_1 values may be obtained by combining the data of $M^{a}_{w(c)}$, $M^{a}_{n(c)}$, and ln f^{a}_1 (eq III-97). First BM_1 is evaluated by using eq III-97, and then c_1 may be obtained from eq II-63 and II-71. (Of course, this does not preclude the possibility of some more general indefinite expression.)

Ferguson et al. recognized the difficulty in the integration process to obtain f_1^{a} (eq II-67), and the integrations were performed from a finite concentration, c^*, rather than from c = 0. The resulting quantity is then

$$\ln (f_1^{a}/f_1^{a*}) = \int_{c^*}^{c} (M_1/M^{a}_{w(c)} - 1)\frac{dc}{c} \qquad \text{(V-5)}$$

The evaluation of BM_1 may be made after the introduction of eq III-97 into eq V-5 above. Evaluations of the association constants were made for monomer–dimer–trimer and monomer–dimer–tetramer cases, using a Monte Carlo method based on eq V-5. From the fact that the

$$\sum[(f_1^{a}/f_1^{a*})_{obsd} - (f_1^{a}/f_1^{a*})_{calcd}]$$

value for the monomer–dimer–trimer was much smaller than that for the monomer–dimer–tetramer reaction, it was concluded that the former mode of association is more nearly the correct one. The association constant and the virial coefficients are given in Table VI.

The general curve-fitting procedures for either $M^{a}_{w(c)}$ vs. c data (eq III-73) or of c vs. r data (eq III-39) have not been applied to this system, and the comparison of several diverse methods is not possible at the moment.

D. Purine

A seemingly complicated, but none the less mathematically simple case is the so-called indefinite or isodesmic self-association, provided that equilibrium constants for all of the successive steps can be taken to be equal. From osmotic coefficient determinations, Ts'o et al.[112] had concluded that in aqueous systems purine and certain other nucleotides associate according to a "stacking" process in which equal increments of free energy are involved in the addition of 1 mol of monomer to any n-mer, a random reaction which involves the principle of equal reactivity of each molecular species.

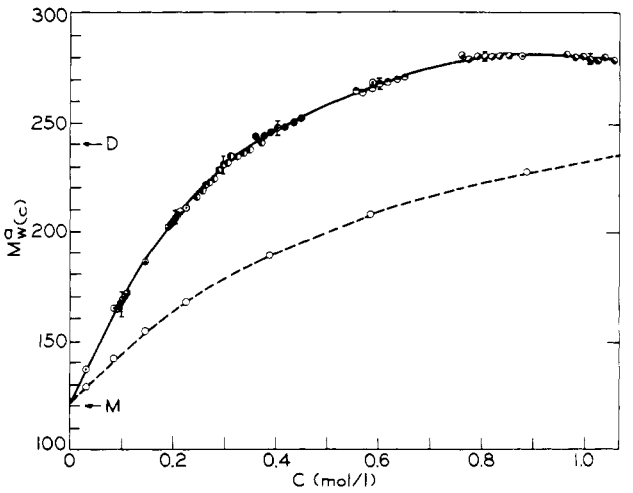

Figure 11. Apparent average molecular weights plotted against molar concentration for purine in aqueous solution at 24.9 °C. Upper curve; apparent weight-average molecular weight. The vertical bars show estimated maximum error at each concentration. The symbols ⊙ are $M^a_{w(c)}$ values calculated from the deduced equilibrium constant (k) and virial coefficient (B). The dashed line shows $M^a_{n(c)}$ vs. c, as calculated from the smoothed curve (solid line) through the $M^a_{w(c)}$ vs. c data. The open circles (O) are $M^a_{n(c)}$ values calculated from k and B. The arrows show theoretical molecular weights for monomer and dimer, respectively. From Van Holde and Rossetti.[16]

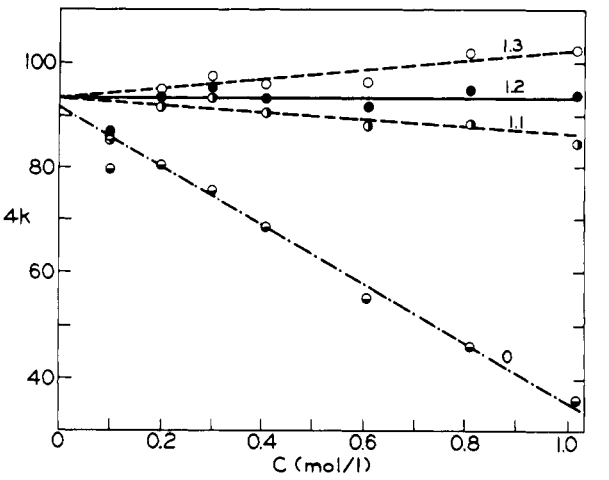

Figure 12. Graphical determination of k and BM_1 for purine in aqueous solution at 24.9 °C. Equation IV-7 was used with the following assumed values of BM_1: ◒, 0.0; ◑, 1.1; ●, 1.2; O, 1.3. From Van Holde and Rossetti.[16]

The subject was further investigated by Van Holde and Rossetti,[16] using the sedimentation equilibrium method for purine self-association in solution at the temperatures 4.9, 9.9, 14.9, 19.9, and 24.9 °C. There are presented in Figure 11 representative plots of $M^a_{w(c)}$ vs. concentration for the purine solutions at different initial loading concentrations and at 24.9 °C. From these and other data it was concluded that (1) a reversible association takes place since data from different experiments lie upon a single curve, (2) the reaction goes beyond dimerization stage, and (3) the system shows nonideality as evidenced by the downward curvature at very high concentrations. Also the curve extrapolates well to $c = 0$ to provide the known monomer molecular weight of 120.

Since any analysis of these data for a monomer–n-mer reaction ($n > 2$) was found to be unsuccessful, it was assumed that this is a case of indefinite reaction, as had indeed been already suggested by Ts'o et al.[112] Therefore eq III-84 was used for the final numerical analysis. This method involves adjustments of BM_1 until a horizontal line is obtained, as is shown by Figure 12. The value of 1.2 for BM_1 was taken to be the correct one, and the intrinsic constants found are $k = (2.33 \pm 0.05) \times 10^3$ dL/g and $K = 2.80 \pm 0.06$ L/mol. The values for k and BM_1 were introduced into eq II-43, II-44, II-51, and II-52 to back-calculate the values of $M^a_{w(c)}$ and $M^a_{n(c)}$. It can be seen from Figure 12 that the experimental and calculated values agree well within experimental error.

Again, eq II-43 (in conjunction with eq II-44) may be used directly to obtain the value of the intrinsic constant k by a general curve-fitting procedure. The precision of this quantity so obtained may be expected to be of the same order as that provided by the application of the Van Holde and Rosetti method (eq III-84). The procedures developed by Adams and Lewis[17] (eq III-99) and Chun et al.[39] (eq III-103) have not been tested in the indefinite association cases, but they may be expected to give a result of near-equivalent accuracy. Further, a direct use of c vs. r data (eq III-33) may enhance the accuracy.

Although Van Holde and Rosetti did not test for other complicated reaction mechanisms such as monomer–dimer–trimer–tetramer, they point out that their data might be equally well explained in terms of such reaction mechanisms (cf. also Van Holde, Rossetti, and Dyson[9]).

E. Lysozyme (Muramidase)

For obvious reasons this protein has been a popular one to evaluate the several computational methods by which the self-association constants are estimated. We cite several of them.[27,35-37,40] Assuming for the present that the predominant reaction in near-neutral systems is that of a monomer–dimer, the several values obtained for k_2 are in quite good overall agreement, after making allowances for temperature and ionic strength differences. But, with the present resolving power of the ultracentrifuge itself and even with the high sensitivity of the Raleigh optical system, the association seems not to be sufficiently pronounced as to make it possible to differentiate between several possible stoichiometries: monomer–dimer, monomer–dimer–trimer, and the isodesmic reactions. In the reports to which reference has just been made, one usually finds that a preference has been expressed, with reasons being given to favor a choice as presented.

Considering the monomer–dimer and isodesmic cases, Teller[43] has made the attempt to set a lower limit for the magnitude of k_2 below which their differentiation will not be possible. It turns out that for lysozyme the k_2 value is just about at the threshold, but he suggests that a valid choice can be made. The function $M_{w(r)}$ is available, and more information is to be obtained from it by giving the derived M_z (Wales recursion formula) as well.

A more detailed summary of several selected individual researchers is now presented. Deonier and Williams[55] used solution conditions of sodium cacodylate buffer at pH 7.0, $I = 0.20$, and 25 °C for their study. In all, data from 15 experiments at different initial concentration are involved. When integration was required, $M^a_{w(c)}$ values were taken from an interpolation curve (Figure 1 of ref 55), drawn in such a way that it is continuous and passes through the average ordinate at those concentrations for which there were more than two experimental values. Since the data at very low concentrations were considered to be unreliable, the interpolation curve was drawn in this concentration region in such a way that the ordinate intercept fell at 14 300, the known monomer molecular weight of the enzyme. The average deviation of the data from the interpolation curve was within 1%.

The maximum value of the apparent weight-average molecular weight is appreciably smaller than the molecular weight of dimer, and it was assumed that only the cases of indefinite reaction and the monomer–dimer association required consideration. The procedure of Van Holde and Rosetti[76] was employed for the analysis of the supposed indefinite reaction case, using eq III-84. The values of k and BM_1 thus found were 0.200 ±

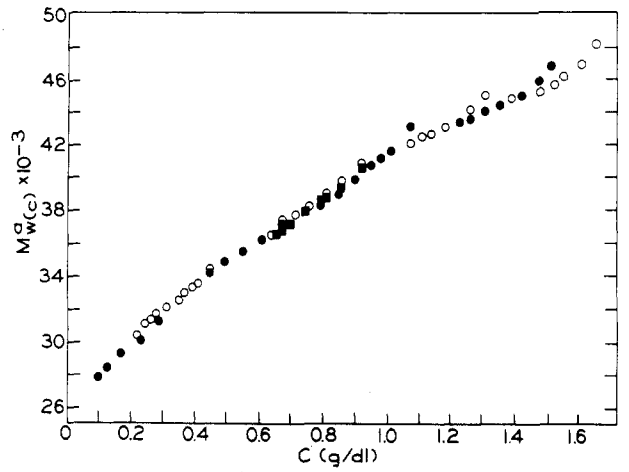

Figure 13. Apparent weight-average molecular weight of chymotrypsinogen A in veronal buffer (pH 7.9, I = 0.03) at 25 °C plotted against concentration. From Hancock and Williams.[98]

0.015 and 0.091 dL/g, respectively. The values of $M^a_{w(c)}$ were then calculated by using these parameters, and it was found that the average deviation between these and the experimental values was about 1%.

For testing the case of the monomer–dimer stoichiometry, both eq III-78 and III-80 were applied. The two methods gave the same values for k_2 (0.347 ± 0.036 dL/g) and for BM_1 (0.020 dL/g). The values of $M^a_{w(c)}$ were back-calculated, and the agreement between the experimental and calculated values was found again to within about 1%.

Although the calculated $M^a_{w(c)}$ values obtained for these two modes of self-association agree with the experimental data equally well, it should be noted that the virial coefficients obtained for these two cases are quite different. Deonier and Williams also pointed out that, while the agreement between the values of $2k$ and k_2 at low concentrations is to be expected, it was difficult to account for the fact that this condition appears to hold over the entire concentration range.

Milthorpe et al.[40] used lysozyme at two initial loading concentrations, in solution at pH 6.7, I, 0.17 (in phosphate buffer), T = 15 °C, for their experiments to test their method for the analysis of sedimentation equilibrium data in a self-association system (cf. p 175). A least-squares estimate gave k_2 = 0.44 dL/g and the conclusion that a nonideal monomer–dimer system is consistent with the experimental data. These results are also consistent with a monomer–dimer–trimer system, but their tests showed that tetramers and higher oligmers are not formed in significant amounts. For reasons given, these authors seem to favor the monomer–dimer–trimer mechanism.

In their analysis of near-neutral systems of lysozyme at pH 6.7, comparable buffer, and T = 15 °C, Adams and Filmer[27] give k_2 = 0.44 dL/g, with BM_1 being taken as −0.04 dL/g. It is, however, true that there is an appreciable amount of scatter in the data points of the $1/M^a_{w(c)}$ vs. c plots. Also the method of computation which was used does not provide a simple means of fitting the data to the model with simultaneous minimization of residuals over the complete concentration range. But, even allowing for the somewhat different set of experimental conditions, the datum of Deonier and Williams at 25 °C, k_2 = 0.35 dL/g, is not at all inconsistent with the Milthorpe et al. value when adjustment for the temperature difference is made. In general, we have reason to feel that some of the earlier data, even those obtained by using the early Steiner–Adams approaches (where an integration step is involved), compare favorably in precision with the newer ones of Milthorpe et al.

F. Chymotrypsinogen A

The investigation of the self-association behavior of chymotrypsinogen A by Hancock and Williams[98] was conducted in a veronal buffer system, pH 7.9, I = 0.03 at 25 °C. The more recent study of Tung and Steiner[113] utilized a barbital buffer of pH 8.1, I = 0.021 at 7 °C. The apparent weight-average molecular weight data for the former are presented in graphical form in Figure 13 and will be considered to be representative in the following discussion. Since there is no indication of a leveling-off of this curve at the high concentrations, it was assumed that the association must proceed beyond the dimerization stage. Two tests were made, one for the indefinite reaction and the other for monomer–dimer–trimer stoichiometry.

For the indefinite reaction mechanism, eq III-99 was used first to evaluate the quantity BM_1 by using a successive approximation method.[17] The use of eq II-51 and eq II-52 then provides the value of k. The numerical quantities thus obtained were k = 0.50 ± 0.061 dL/g and BM_1 = 4.0 dL/g. It is to be noted that the scatter in k is about 1.2%.

For the monomer–dimer–trimer association, the value of BM_1 was first calculated by a successive approximation method, eq III-97. Then, use of eq II-71, III-86, and III-95 (m = 2, n = 3) gave data for k_2 and k_3. The final values so obtained were k_2 = 0.908 ± 0.030 dL/g, k_3 = 0.844 ± 0.032 dL/g, and BM_1 = −0.01 dL/g. The random errors are 3.3% for k_2 and 3.8% for k_3.

When the two distinct sets of equilibrium constant and virial coefficient data were applied in the computations required to give values of $M_1/M^a_{w(c)}$ as a function of concentration which correspond to the two mechanisms, a satisfactory agreement between the experimentally observed and calculated values of $M_1/M^a_{w(c)}$ was obtained in both instances. The monomer molecular weight was assumed to be 25 600.

Again the values of BM_1 for two different modes of association are quite different; actually a small negative value is found to account for the monomer–dimer–trimer reaction data. It may also be noticed that the agreement between calculated and experimental values for $M_1/M^a_{w(c)}$ is slightly better for the random stoichiometry. Furthermore, there is a smaller random error in the intrinsic association constant.

The emphasis of the Tung and Steiner article[113] is on the presentation of a general method to analyze nonideal self-association system which does not require a prior assumption as to model, with this reaction of chymotrypsinogen A in solution having been selected as the test system. The solution conditions were pH 8.14, I = 0.021, and T = 7 °C. Because of the appreciable difference in the temperature used in the two laboratories, a simple direct comparison between the k and BM_1 data from the two laboratories is not possible.

The three cases (purine, lysozyme, and chymotrypsinogen A) are alike in that a clear-cut differentiation between two (or more) mechanisms of self-association is difficult, but the preponderance of the present data of the literature seem to favor the isodesmic reaction. In such situations the cause of the uncertainty seems to be more a matter of the resolving power of the ultracentrifuge than limitations of either theory or computational methods. The sensitivity of the optical system is another important consideration. If the amounts of any of the successively higher oligomers are beyond detection, $k_2 \simeq 2k$. This fact is well illustrated by Chun and Kim (Figure 2 of ref 38) in which it appears that in $M_{w(c)}/M_1$ vs. f_1 plots, the curves for the monomer–dimer and the indefinite stoichiometries (lysozyme) approach each other tangentially as $f_1 \rightarrow 1$. For the chymotrypsinogen A it will be noted that the relationships between the numerical values of k, k_2, and k_3 are all consistent with the random mechanism, a necessary condition for the establishment of the isodesmic reaction mechanism.

In general, if the association is weak and monomer and dimer are the dominant species over the accessible concentration range, it will be difficult at best to distinguish between a limited discrete association and one of the isodesmic type. Van Holde et al.[9] have shown that, if at the highest solute concentration used species larger than dimer constitute but a few per cent of the equilibrium mixture, their contribution to the data will be so

small that any one of the simpler mechanisms will appear to be indicated.

VI. Approach to a Physical Picture: The Thermodynamic Interpretation of the Equilibrium Constants

A. General Considerations

So far we have been concerned with obtaining accurate information about the stoichiometry of self-association reactions and the corresponding association constants. The stoichiometry is a foundation upon which knowledge from other studies may be combined in order to elucidate the overall configuration of the oligomer. The association constants, under favorable circumstances, will give information as to the nature of the association reaction. It is this situation which remains for brief consideration.

The first step toward this goal is to obtain the Gibbs free energy for the reaction by using eq II-10 after proper adjustment for concentration scale is made (eq II-20). The free energy thus provided has several components which derive from different sources. They can be described as being work done to bring the charged particles together, the result of direct interaction between the particle surfaces themselves, changes in monomer structure, reduction of particle numbers, etc. Thus a number of complex problems are involved. Found in the literature of today are accounts only of beginning attempts to achieve quantitative descriptions of the component parts, and it is felt to be premature in time to attempt any definite comments about them.

For attempts to precalculate a numerical value of BM_1, the comprehensive McMillan and Mayer[114] thermodynamical treatment of multicomponent systems had provided the means for the Stigter and Hill[115] treatment of the osmotic pressure of an electrolyte system which contains added charged macromolecules and at Donnan equilibrium. The large molecules were assumed to have the form of uniformly charged spheres. To compute the effect of the electrostatic interaction between two colloid particles in the system, expressions first written down by Vervey and Overbeek[116] were applied.

In their application of the Stigter–Hill result to the self-association reaction of chymotrypsinogen A, Tung and Steiner[113] have indicated how one goes about to compute the BM_1 value and have compared their result with that taken from their sedimentation equilibrium data for the protein in solution. The two values, theory and experiment, differ by a factor of approximately 2, with the latter figure being the larger.

B. Effect of Solvent Composition on the Self-Association Reactions

Many types of protein reactions, including the self-associations, depend on variations in solvent composition such as pH and the concentration of added solutes. A general thermodynamic formalism of such linked functions has been given by Wyman.[117] The dependence of equilibrium constant, or the free energy change, ΔG^a, on the solvent composition may be expressed in general form as

$$d(\Delta G^a) = \left\{\frac{\partial(\Delta G^a)}{\partial \mu_H}\right\}_{a_w,a_x} d\mu_H + \left\{\frac{\partial(\Delta G^a)}{\partial \mu_x}\right\}_{a_w,a_H} d\mu_x + \left\{\frac{\partial(\Delta G^a)}{\partial \mu_w}\right\}_{a_x,a_H} d\mu_w \quad \text{(VI-1)}$$

or

$$d \ln K = \left(\frac{\partial \ln K}{\partial \ln a_H}\right)_{a_w,a_x} d \ln a_H + \left(\frac{\partial \ln K}{\partial \ln a_x}\right)_{a_w,a_H} \times d \ln a_x + \left(\frac{\partial \ln K}{\partial \ln a_w}\right)_{a_x,a_H} d \ln a_w \quad \text{(VI-2)}$$

In these equations ΔG^a is the molar free energy change due to the direct interaction between particle surfaces, μ is the chemical potential, and a represents the activity of the several components. The quantities H, w, and x are proton, water, and added solute, respectively. This added solute is commonly a low molecular weight compound. Wyman has shown that

$$\left\{\frac{\partial(\Delta G^a)}{\partial \mu_i}\right\}_{a_j,a_k} = \left(\frac{\partial \ln K}{\partial \ln a_i}\right)_{a_i,a_x} = \Delta\tilde{\nu}_i \quad \text{(VI-3)}$$

where $\Delta\tilde{\nu}_i$ is the difference in the number of bound components i per monomer between the n-mer and the monomer.

If the association constant is obtained as a function of pH with the concentration of added solute remaining fixed,

$$\frac{d \ln K}{d \ln a_H} = \left(\frac{\partial \ln K}{\partial \ln a_H}\right)_{a_w,a_x} = \Delta\tilde{\nu}_H \quad \text{(VI-4)}$$

since it is expected that

$$d \ln a_w / d \ln a_H \simeq 0$$

On the other hand, if the concentration of added solute is changed while maintaining the pH constant

$$\frac{d \ln K}{d \ln a_x} = \left(\frac{\partial \ln K}{\partial \ln a_x}\right)_{a_w,a_H} + \left(\frac{\partial \ln K}{\partial \ln a_w}\right)_{a_x,a_H} \frac{d \ln a_w}{d \ln a_x} \quad \text{(VI-5)}$$

For dilute solutions, the second term of the right-hand side of eq VI-5 is negligible and

$$\frac{d \ln K}{d \ln a_x} = \left(\frac{\partial \ln K}{\partial \ln a_x}\right)_{a_w,a_H} = \Delta\tilde{\nu}_x \quad \text{(VI-6)}$$

For a system with high concentration of added solute, this approximation is not valid; Tanford[118] has extended Wyman's treatment to this case. The relation between d ln a_w and d ln a_x for the system under consideration may be obtained by using the Gibbs–Duhem equation

$$d \ln a_w = -(n_x/n_w)\, d \ln a_x - (n_A/n_w)\, d \ln a_A \quad \text{(VI-7)}$$

where n_A, n_x, and n_w designate the total number of moles of macromolecule (A), ligand (x), and water (w) in the solution. When the molar concentration of protein is small, eq VI-7 may be approximated to obtain

$$d \ln a_w = -(n_x/n_w)\, d \ln a_x \quad \text{(VI-8)}$$

Introduction of eq VI-3 and VI-8 into eq VI-5 gives

$$\frac{d \ln K}{d \ln a_x} = \Delta\tilde{\nu}_x - \frac{n_x}{n_w}\Delta\tilde{\nu}_w$$

and from this statement

$$d \ln K / d \ln a_x = \Delta\tilde{\nu}_x - (m_x/m_w)\Delta\tilde{\nu}_w \quad \text{(VI-9)}$$

where m_i is the molarity of component i. Thus eq VI-9 suggests that the dependence of association constant on the concentration of added solute may be caused by the binding of the solute to protein as well as by its influence on the hydration of the protein.

Since the value of $\Delta\tilde{\nu}_i$ is due to the change in the equilibrium constants between a ligand and the protein, in principle it should be possible to relate them. However, since there is generally a large number of groups on the protein molecule which form complexes with the ligand, a general expression to account for this effect is not practical. Thus it is usually assumed that the $\Delta\tilde{\nu}_i$ are predominantly determined by a very small number of groups which are directly involved in the association reaction.[119,120] If there are only two such groups, then for a monomer–n-mer system:

$$\Delta\tilde{\nu}_i = \frac{na_i}{a_i + \hat{k}_{1,n}} + \frac{na_i}{a_i + \hat{k}_{2,n}} - \frac{na_i}{a_i + \hat{k}_{1,1}} - \frac{na_i}{a_i + \hat{k}_{2,1}} \quad \text{(VI-10)}$$

Here $k_{1,n}$ and $k_{2,n}$ represent the equilibrium constants between the ligand and the groups 1 and 2 on the n-mer, respectively, and

$\hat{k}_{1,1}$ and $\hat{k}_{2,1}$ designate the corresponding quantities for the monomer. Thus, from eq VI-4, VI-6, and V-10, it is possible to obtain the association constant between ligand and protein. By comparing these values with known values from other studies, it may be possible to identify the amino acid groups involved in the self-association of the protein. This is most successful for the case where the ligand is a proton; this will be the subject of the following discussion.

1. Dependence of Self-Association on pH: Dimerization of α-Chymotrypsin

Egan et al.[121] studied the dimerization reaction of α-chymotrypsin as a function of pH at 20 °C, using the sedimentation transport method. It was found that the pH dependence of $s_{20,w}$ follows roughly a bell-shaped curve and that the maximum value of the sedimentation coefficient appears at about pH 4. Later, Aune and Timasheff[120] have made extensive sedimentation equilibrium observations as a function of pH to obtain association constant data for the dimerization process (cf. Figure 3 of ref 120). Here again a bell-shaped curve was found with the maximum value of ln k_2 near pH 4. This type of pH dependence suggests that the self-association is at least in part due to the interaction of two ionizing groups for which the maximum value of their association constants is located near the mean pK value of these ionizing groups. Noting that the maximum absolute value of $\Delta\tilde{\nu}_H$ deduced from the negative slope in their Figure 3 is 1.5, these workers assumed that only two ionizing groups are involved. The pK values of these two groups were then obtained by the application of a curve-fitting process to the data into the expression

$$\ln K = \ln K(\mathrm{pH} = 0) + 2 \ln \frac{(1 + a_H/\hat{k}_{1,2})(1 + a_H/\hat{k}_{2,2})}{(1\ a_H/\hat{k}_{1,1})(1 + a_H/\hat{k}_{2,1})} \qquad \text{(VI-11)}$$

Equation VI-11 was obtained by integrating eq VI-4, following the insertion of eq VI-10 for the case of a dimerization reaction. The results are given in Table VIII. The k_i^0 in this table are the values corrected for the long-range electrostatic effect.

As was pointed out by these authors, these results, Table VIII, are based on rather drastic approximations. However, the values presented are considered to be adequate to assign the reaction site to particular ionizing groups of the protein. They concluded that (a) the shift in p$\hat{k}$ values shown in the table suggests that a cationic group interacts directly with an anionic group; (b) considering the pH range, the most likely groups involved are a terminal or side-chain carboxyl and the imidazole group of histidine side chain; and (c) since the calculated values, using eq VI-11, agree well with the experimental points as shown in Figure 3 of ref 120, there is no need to consider other groups to be involved in the dimerization mechanism. From the study of inter- and intracharge interactions and using the atomic coordinates (x ray) of α-chymotrypsin as obtained by Birktoft et al.,[122] Aune and Timasheff suggested that interaction between the histidine 57 and the α-carboxyl group of tyrosine 146 is directly involved in the dimerization. This is perhaps the first time in which the results from sedimentation equilibrium and x-ray crystallographic studies for a self-associating protein have been combined to assign the particular amino acid groups which are directly involved.

2. Dependence of Self-Association Reactions on Electrolytes at High Concentrations

Unlike the studies of the pH dependence of a self-association reaction, the work on the dependence of the reaction on the salt (or other solute) concentration cannot be used for the identification of the amino acid groups participating in the reaction. For one thing, eq VI-9 contains two unknowns and therefore it is not possible to obtain either $\Delta\tilde{\nu}_x$ or $\Delta\tilde{\nu}_w$ with accuracy. Even if these quantities were known, it would be very difficult to identify the sites which contribute. Apparently, only some qualitative information may be derived from this kind of study. The most detailed investigation for the effect of salt concentration on an association reaction has been carried out by Aune et al.[123] They had observed that the self-association of this protein is enhanced by an increase in ionic strength, other things being equal. Steiner[124] found that the calculated value of the work done to bring the charged particles together decreased with increasing ionic strength. He then suggested that the increase in the association constant is due to the nonspecific effect of decreasing the unfavorable electrostatic repulsions. It was observed by Aune et al. that the ln K are all dependent on the nature of the salt which is present. Plots of ln K against the logarithm of the ion-pair activity gave curves which approximate a straight line.

TABLE VIII. pK's of Groups Involved in Dimerization of α-Chymotrypsin

pK_2	pK_1	pK_2^0	pK_1^0
6.2	5.0	6.2	5.2
2.4	3.6	2.2	4.5

Aune et al. noted that the values of (d ln K/d ln a_x) at pH 4.1 are about 1.0 in NaCl solution and 0.5 in $CaCl_2$ solution. They suggested that a change in ion binding to the protein should make these values larger than two, because of the symmetrical structure of the dimer, unless the value of $\Delta\tilde{\nu}_w$ is positive. From the fact that the sedimentation coefficient of this protein in D_2O is larger than in H_2O they concluded that $\Delta\tilde{\nu}_w$ is negative. In their judgment, then, no change in specific ion binding is involved but rather the observed values of (d ln K/d ln a_x) are suggestive of the release of hydrated water upon dimerization which, in turn, indicates that a hydrophobic interaction takes part in the association of α-chymotrypsin monomers. These workers then performed sedimentation transport and sedimentation equilibrium experiments in solutions of $NaClO_4$ and Na_2SO_4 which are near the extremes of the Hofmeister series. It is believed that perchlorate ion salts out hydrophobic groups less effectively than do sulfate ion salts.[125,126] From this fact it is to be expected that the sedimentation coefficient of α-chymotrypsin in an Na_2SO_4 system should be larger than that in $NaClO_4$ solution. Actual experiment shows that the reverse order obtains, indicating that these electrolytes cannot act only in the hydrophobic regions. It was suggested that these electrolytes influence the interaction between charged groups in protein and water molecules as well.

There is no other system of which we are aware for which a record of extensive studies of the effect of salt concentration on the self-association is available. It is interesting to note that the association of chymotrypsinogen A is enhanced by a decrease in ionic strength near the isoelectric point.[127] This was interpreted to mean that electrostatic interaction is an important factor in the dimerization.

C. Effects of Temperature: Enthalpy and Entropy of Self-Association

If the values of association constants are determined as a function of temperature, one may be able to calculate the entropy change, $\Delta S°$, and the enthalpy change, $\Delta H°$, from the following expression

$$\Delta G° = -RT \ln K_i = \Delta H° - T\Delta S° \qquad \text{(VI-12)}$$

which upon rearrangement becomes

$$\ln K_i = \frac{\Delta H°}{RT} + \frac{\Delta S°}{R} \qquad \text{(VI-13)}$$

It should be noted that both $\Delta S°$ and $\Delta H°$ are sums of several

contributions, as is the case for ΔG°. A plot of ln K against $1/T$ should give a straight line if the entropy and enthalpy of an association reaction are constants within the temperature range studied. It is from the intercept and the slope of this plot that the two thermodynamic parameters are calculated. For certain systems such as α-chymotrypsin the plot of ln K vs. $1/T$ curve is not strictly linear and thus neither ΔH° nor ΔS° can be considered as being independent of temperature. In such situations, then, the association reaction has a significant effect on the partial molal heat capacity of the protein ΔC_p, as shown by the expression

$$\delta \Delta H^\circ / \delta T = T \delta \Delta S^\circ / \delta T = \Delta C_p \qquad \text{(VI-14)}$$

1. Theoretical Values of ΔH° and ΔS°

The most important interactions involved in protein self-association reactions are (a) electrostatic interactions between negatively charged side chains on one protein with positively charged side chains on the others, (b) hydrogen bonds, and (c) hydrophobic interactions. The enthalpies characteristic of electrostatic interactions are negative and large. The enthalpy for the side chain hydrogen bond surrounded entirely by nonpolar groups is estimated to be about −6 kcal/mol and for the groups exposed to water it has a value of −1.5 kcal/mol.[128] The enthalpy for the interaction between two nonpolar side chains ranges between +0.3 to +1.8 kcal/mol.

2. Experimental Values

We have assembled in Table IX a few of the experimental values of these thermodynamic parameters for the three β-lactoglobulins and ATP. The interpretation of these parameters in terms of the association mechanism is difficult because corrections for the charge repulsion have not been made. It is of interest to note that for β-A at the isoelectric point (pH 5.2) the ΔS° is positive, indicating the participation of the hydrophobic interactions in the association reaction. Unlike the systems for which data are presented in this table, for α-chymotrypsin, the plot of ln K vs. $1/T$ is not linear. Since a positive change in enthalpy and entropy is expected when a hydrophobic interaction is an important factor for a self-association reaction, Aune et al.[123] have suggested that this factor must be also considered in the dimerization of α-chymotrypsin (Table I of ref 123); electrostatic interaction alone is insufficient to account for the observations. The interpretation of the thermodynamic parameters for any protein system is still a hazardous task because of a large number of uncertainties in the isolation of a local effect from the overall effect of association. Since the association seems to involve the cooperative effect of several factors, an understanding of the mechanism will presumably require experiments under a very large number of solvent conditions.

TABLE IX. Thermodynamic Parameters of Some Self-Association Reactions

System	Reaction	pH	I	ΔH°, kcal/mol	ΔS°, eu	Ref
β-Lactoglobulin A	1-2	2.46	0.1	−17.1	−42.1	107
β-Lactoglobulin B	1-2	2.64	0.16	−10.4	−15.0	23
β-Lactoglobulin C	1-2	2.46	0.1	−16.5	−38.3	107
		2.46	0.2	−8.6	−8.5	107
ATP	1-2			−14.9	−41.6	111
	1-3			−10.8	−13.0	111

VII. Concluding Remarks

In bringing this survey to a close, we may note that the theoretical development of the analytical procedures for use with the sedimentation equilibrium data in a protein self-associating system has reached a reasonably high degree of sophistication. Of the general methods to obtain both the association constants and the virial coefficient (usually necessary) those based upon the direct use of the concentration as a function of radial distance in the cell at equilibrium seem to be preferred, though they have not as yet been fully exploited. However, there are situations where it is apparently essential to use data for the average molecular weight at fixed points vs. the corresponding concentrations along the cell to describe the stoichiometry of the reaction, for instance, to learn whether a pressure effect on the equilibrium constant is present.

With proper care and under favorable circumstances quite satisfactory values for the thermodynamic parameters of interest can be obtained. However, the interpretation of these quantities in terms of a detailed and unambiguous association mechanism remains a difficult task. Success in such an operation depends first of all on the quality of the experimental data. Involved are the resolving power of the ultracentrifuge and the sensitivity of its optical system. It is not likely that the machine itself will be much improved in resolving power in the immediate future but it is obvious that the adaptation of the Rayleigh optical system to observe the redistribution of the reaction system at equilibrium was a step which made possible the recent explosive development of our subject. It appears that later on a real advantage may accrue to the use of laser light sources for they produce improved fringe quality and they allow the performance of experiments at higher solute concentrations. Further, the use of averaging techniques is continually being advanced.

Although it is perhaps an ancillary subject for a report of this kind, very brief consideration has been given to the study of the forces involved in the self-association process. It is a subject of considerable significance in the overall picture, but it is one which is hardly ready for definitive survey. The relative contributions of the several types of forces involved in the reactions are difficult to separate and evaluate. In addition to the use of the traditional methods of exploitation, e.g., changes in the conditions of solution, resort may be taken to the use of model compounds and of local chemical modifications of the protein which is to be the subject of the experimental study. With the experience and information gained in their application, ways may be indicated by which the nature and quantitative measure of these several forces can be delineated so to aid in the eventual more definitive description of the reaction mechanisms themselves.

Acknowledgments. It had been our original plan that Drs. Kensal Van Holde and Robert D. Dyson collaborate with us in the preparation of this review. Other duties prevented joint authorship, but they did give detailed and critical study to the manuscript and both provided many valuable suggestions for its improvement, rearrangement of subjects, etc., practically all of which are included. We are deeply indebted to them for their aid.

This effort has been supported by (and, in fact, would have been impossible without) a grant from the National Science Foundation, No. GB-29,100. Grateful recognition is made.

VIII. References

(1) J. R. Cann, "Interacting Macromolecules", Academic Press, New York, N.Y., 1970.
(2) L. W. Nichol and D. J. Winzor, "Migration of Interacting Systems", Clarendon Press, Oxford, 1972.
(3) T. Svedberg and K. O. Pedersen, "The Ultracentrifuge", Clarendon Press, Oxford, 1940.
(4) G. S. Adair, *Proc. R. Soc. London, Ser. A*, **108**, 627 (1925).
(5) L. S. Moody, Ph.D. Thesis, University of Wisconsin—Madison, 1944.
(6) H. Gutfreund, Jr., *Biochem. J.*, **42**, 156, 544 (1948); **50**, 564 (1952).
(7) A. Tiselius, *Z. Phys. Chem.*, **124**, 449 (1925).
(8) P. G. Squire and C. H. Li, *J. Am. Chem. Soc.*, **83**, 3521 (1961).
(9) K. E. Van Holde, G. P. Rossetti, and R. D. Dyson, *Ann. N.Y. Acad. Sci.*, **164**, 279 (1969).

(10) E. T. Adams, Jr., and H. Fujita in "Ultracentrifugal Analysis in Theory and Experiment", J. W. Williams, Ed., Academic Press, New York, N.Y., 1962, p 119.
(11) E. T. Adams, Jr., and J. W. Williams, *J. Am. Chem. Soc.*, **86,** 3454 (1964).
(12) J. G. Kirkwood and I. Oppenheim, "Chemical Thermodynamics", McGraw-Hill, New York, N.Y., 1961.
(13) A. G. Ogston and D. J. Winzor, *J. Phys. Chem.*, **79,** 2496 (1975).
(14) H.-G. Elias and R. Bareiss, *Chimica*, **21,** 53 (1967).
(15) H.-G. Elias and H. Lys, *Makromol. Chem.*, **96,** 64 (1966).
(16) K. E. Van Holde and G. P. Rossetti, *Biochemistry*, **6,** 2189 (1967).
(17) E. T. Adams, Jr., and M. S. Lewis, *Biochemistry*, **7,** 1044 (1968).
(18) R. Josephs, H. Eisenberg and E. Reisler in R. Jaenick and E. Helmreich, Ed., "Protein–Protein Interactions", Springer-Verlag, Berlin, 1972.
(19) R. J. Goldberg, *J. Phys. Chem.*, **57,** 194 (1953).
(20) J. W. Williams, K. E. Van Holde, R. L. Baldwin, and H. Fujita, *Chem. Rev.*, **58,** 715 (1958).
(21) H. Fujita, "Foundations of Ultracentrifugal Analysis", Wiley-Interscience, New York, N.Y., 1975.
(22) E. F. Casassa and H. Eisenberg, *Adv. Protein Chem.*, **19,** 287 (1964).
(23) J. Visser, R. C. Deonier, E. T. Adams, Jr., and J. W. Williams, *Biochemistry*, **11,** 2634 (1972).
(24) R. Townend and S. N. Timasheff, *J. Am. Chem. Soc.*, **82,** 3168 (1960).
(25) E. T. Adams, Jr., *Biochemistry*, **4,** 1646 (1965).
(26) M. Wales, *J. Phys. Colloid Chem.*, **52,** 235 (1948).
(27) (a) E. T. Adams, Jr., and D. L. Filmer, *Biochemistry*, **5,** 2971 (1966); (b) E. T. Adams, Jr., *ibid.*, **6,** 1864 (1967).
(28) R. F. Steiner, *Arch. Biochem. Biophys.*, **39,** 333 (1952).
(29) H. Rinde, Dissertation, Uppsala, 1928.
(30) R. H. Haschemeyer and W. F. Bowers, *Biochemistry*, **9,** 435 (1970).
(31) L. A. Holladay and A. J. Sophianopoulos, *J. Biol. Chem.*, **247,** 427 (1972).
(32) R. C. Deonier, Ph.D. Thesis, University of Wisconsin—Madison, 1970.
(33) W. P. Reinhardt and P. G. Squire, *Biochim. Biophys. Acta*, **94,** 566 (1965).
(34) E. T. Adams, Jr., *Proc. Nat. Acad. Sci. U.S.*, **51,** 509 (1964).
(35) A. J. Sophianopoulos and K. E. Van Holde, *J. Biol. Chem.*, **239,** 2516 (1964).
(36) D. E. Roark and D. A. Yphantis, *Ann. N.Y. Acad. Sci.*, **164,** 245 (1969).
(37) D. C. Teller, T. A. Horbett, E. G. Richards, and H. K. Schachman, *Ann. N.Y. Acad. Sci.*, **164,** 66 (1969).
(38) P. W. Chun and S. J. Kim, *Biochemistry*, **9,** 1957 (1970).
(39) P. W. Chun, S. J. Kim, J. D. Williams, W. T. Cope, L. H. Tang, and E. T. Adams, Jr., *Biopolymers*, **11,** 197 (1972).
(40) B. K. Milthorpe, P. D. Jeffrey, and L. W. Nichol, *Biophys. Chem.*, **3,** 169 (1975).
(41) Th. G. Scholte, *J. Polym. Sci., Part A-2*, **6,** 111 (1968).
(42) Th. G. Scholte, *Ann. N.Y. Acad. Sci.*, **164,** 156 (1969).
(43) D. C. Teller, Dissertation, University of Washington, 1965.
(44) A. Rosenthal, *Macromolecules*, **4,** 35 (1971).
(45) R. D. Dyson and I. Isenberg, *Biochemistry*, **10,** 3233 (1971).
(46) P. W. Chun and S. J. Kim, *Biophys. Chem.*, **3,** 46 (1975).
(47) G. Kegeles and M. S. N. Rao, *J. Am. Chem. Soc.*, **80,** 5721, 5724 (1958).
(48) P. D. Jeffrey and J. H. Coates, *Biochemistry*, **5,** 489 (1966).
(49) M. Derechin, *Biochemistry*, **7,** 3253 (1968).
(50) M. Derechin, *Biochemistry*, **8,** 921, 927 (1969).
(51) M. Derechin, *Biochemistry*, **10,** 4981 (1971).
(52) P. W. Chun, M. Fried, and K. S. Yee, *J. Theoret. Biol.*, **19,** 147 (1968).
(53) M. A. Efroymson in "Mathematical Methods for Digital Computing", Part V, A. Ralston and H. S. Wilf, Ed., Wiley, New York, N.Y., 1960, p 17.
(54) M. E. Magar, "Data Analysis in Biochemistry and Biophysics", Academic Press, New York, N.Y., 1972.
(55) R. C. Deonier and J. W. Williams, *Biochemistry*, **9,** 4260 (1970).
(56) E. T. Adams, Jr., Fractions No. 3, Spinco Division of Beckman Instruments, Inc., Palo Alto, Calif., 1967.
(57) R. D. Dyson, personal communication.
(58) D. C. Teller in "Methods in Enzymology", Vol. 27, Part D, C. W. Hirs and S. N. Timasheff, Ed., Academic Press, New York, N.Y., 1973, p 346.
(59) H. Neurath and R. L. Hill, Ed., "The Proteins", Vol. 1, 3rd ed, Academic Press, New York, N.Y., 1975.
(60) K. E. Van Holde, in ref 59, p 268.
(61) I. M. Klotz, D. W. Darnall, and N. R. Langerman, in ref 59, p 352.
(62) M. Wales, F. T. Adler, and K. E. Van Holde, *J. Phys. Colloid Chem.*, **55,** 145 (1951).
(63) D. A. Yphantis, *Biochemistry*, **3,** 297 (1964).
(64) J. E. Godfrey and W. F. Harrington, *Biochemistry*, **9,** 894 (1970).
(65) D. A. Albright and J. W. Williams, *Biochemistry*, **7,** 67 (1968).
(66) C. Tanford, "Physical Chemistry of Macromolecules", Wiley, New York, N.Y., 1961.
(67) L. J. Gosting, *Adv. Protein Chem.*, **11,** 430 (1956).
(68) H. kim, G. Reinfelds, and L. J. Gosting, *J. Phys. Chem.*, **77,** 934 (1973).
(69) T. Svedberg, *Kolloid-Z.*, **36** (Erg.-Bd.), 58 (1926).
(70) A. Tiselius, *Z. Phys. Chem. (Leipzig)*, **124,** 449 (1926); *Kolloid-Z.*, **59,** 306 (1932).
(71) K. O. Pedersen, *Z. Phys. Chem., Abt. A*, **170,** 41 (1934); cf. also *J. Phys. Chem.*, **62,** 1282 (1968).
(72) O. Lamm, *Ark. Kemi, Mineral. Geol.*, **17A,** No. 25 (1944).
(73) P. F. Mijnlieff in "Ultracentrifugal Analysis", J. W. Williams, Ed., Academic Press, New York, N.Y., 1963.
(74) J. S. Johnson, K. A. Kraus, and G. Scatchard, *J. Phys. Chem.*, **58,** 1034 (1954).
(75) A. Vrij and J. Th. G. Overbeek, *J. Colloid Sci.*, **17,** 570 (1962).
(76) D. E. Roark and D. A. Yphantis, *Biochemistry*, **10,** 3241 (1971).
(77) T. F. Young, K. A. Kraus, and J. S. Johnson, *J. Chem. Phys.*, **22,** 878 (1954).
(78) G. Kegeles, G. L. Rhodes, and J. L. Bethune, *Proc. Nat. Acad. Sci. U.S.*, **58,** 45 (1967).
(79) W. F. Harrington and G. Kegeles in ref 58, p 306.
(80) R. Josephs and W. F. Harrington, *Biochemistry*, **5,** 3474 (1966); *Fed. Proc., Fed. Am. Soc. Exp. Biol.*, **26,** 728 (1967); *Proc. Nat. Acad. Sci. U.S.*, **58,** 1587 (1967).
(81) G. L. Howlett, P. D. Jeffrey, and L. W. Nichol, *J. Phys. Chem.*, **74,** 3607 (1970).
(82) L. F. TenEyck and W. Kauzmann, *Proc. Nat. Acad. Sci. U.S.*, **58,** 888 (1967).
(83) G. L. Howlett, P. D. Jeffrey, and L. W. Nichol, *J. Phys. Chem.*, **76,** 777 (1972).
(84) J. C. Swan and G. Hammes, *Biochemistry*, **8,** 1 (1969).
(85) P. D. Jeffrey and J. H. Coates, *Biochemistry*, **5,** 489, 3820 (1966).
(86) L. J. Gosting and H. Fujita, *J. Am. Chem. Soc.*, **79,** 1359 (1957).
(87) L. G. Longsworth, *J. Am. Chem. Soc.*, **74,** 4155 (1952).
(88) J. M. Creeth, *J. Am. Chem. Soc.*, **77,** 6428 (1955).
(89) J. S. Johnson, K. A. Kraus, and G. Scatchard, *J. Phys. Chem.*, **63,** 787 (1959).
(90) E. G. Richards and H. K. Schachman, *J. Phys. Chem.*, **63,** 1578 (1959).
(91) S. Claesson and I. Moring-Claesson, *Anal. Methods Protein Chem.*, **3,** 121 (1961).
(92) H. K. Schachman in "Ultracentrifugal Analysis", Academic Press, New York, N.Y., 1969.
(93) K. E. Van Holde, Fractions, No. 1, Spinco Division of Beckman Instruments, Inc., Palo Alto, Calif., 1967.
(94) H. K. Schachman in "Ultracentrifugal Analysis in Theory and Experiment", J. W. Williams, Ed., Academic Press, New York, N.Y., 1963, p 171.
(95) K. Kawahara, A. G. Kirschner, and C. Tanford, *Biochemistry*, **4,** 1203 (1965).
(96) J. M. Creeth and R. H. Pain, *Prog. Biophys. Mol. Biol.*, **17,** 127 (1967).
(97) F. E. LaBar, *Proc. Nat. Acad. Sci. U.S.*, **54,** 31 (1965).
(98) D. K. Hancock and J. W. Williams, *Biochemistry*, **8,** 2598 (1969).
(99) M. Cassman and H. K. Schachman, *Biochemistry*, **10,** 1015 (1971).
(100) L. H. Tang and E. T. Adams, Jr., *Arch. Biochem. Biophys.*, **157,** 520 (1973).
(101) S. N. Timasheff and R. Townend, *J. Am. Chem. Soc.*, **83,** 464 (1961).
(102) T. F. Kumosinski and S. N. Timasheff, *J. Am. Chem. Soc.*, **88,** 5635 (1966).
(103) H. A. McKenzie and L. W. Nichol in "The Milk Proteins", Vol. I, H. A. McKenzie, Ed., Academic Press, New York, N.Y., 1970.
(104) M. J. Kelly and F. J. Reithel, *Biochemistry*, **10,** 2639 (1971).
(105) R. Townend, L. Weinberger, and S. N. Timasheff, *J. Am. Chem. Soc.*, **82,** 3175 (1960).
(106) V. G. Frank and G. Braunitzer, *Hoppe-Seyler's Z. Physiol. Chem.*, **348,** 1691 (1967).
(107) J. L. Sarquis and E. T. Adams, Jr., *Arch. Biochem. Biophys.*, **163,** 442 (1974).
(108) J. L. Sarquis and E. T. Adams, Jr., *Biophys. Chem.*, **4,** 181 (1976).
(109) N. R. Langerman and I. M. Klotz, *Biochemistry*, **8,** 4746 (1969).
(110) I. M. Klotz and S. Keresztes-Nagy, *Biochemistry*, **2,** 455 (1963).
(111) W. E. Ferguson, C. M. Smith, E. T. Adams, Jr., and G. H. Barlow, *Biophys. Chem.*, **1,** 325 (1974).
(112) P. O. P. Ts'o, I. S. Melvin, and A. C. Olson, *J. Am. Chem. Soc.*, **85,** 1289 (1963).
(113) M. S. Tung and R. F. Steiner, *Eur. J. Biochem.*, **44,** 52 (1974).
(114) W. G. McMillan and J. E. Mayer, *J. Chem. Phys.*, **13,** 276 (1954).
(115) D. Stigter and T. L. Hill, *J. Phys. Chem.*, **63,** 551 (1959).
(116) E. J. Vervey and J. Th. G. Overbeek, "Theory of the Stability of Lyophobic Colloids", Elsevier Publishing Co., Amsterdam, 1948.
(117) J. Wyman, Jr., *Adv. Protein Chem.*, **19,** 223 (1964).
(118) C. Tanford, *J. Mol. Biol.*, **39,** 539 (1969).
(119) K. C. Aune and C. Tanford, *Biochemistry*, **8,** 4579 (1969).
(120) K. C. Aune and S. N. Timasheff, *Biochemistry*, **10,** 1609 (1971).
(121) R. Egan, H. O. Michel, R. Schlueter, and B. J. Jandorf, *Arch. Biochem. Biophys.*, **36,** 366 (1957).
(122) J. J. Birktoft, B. W. Mathews, and D. M. Blow, *Biochem. Biophys. Res. Commun.*, **36,** 131 (1969).
(123) K. C. Aune, L. C. Goldsmith, and S. N. Timasheff, *Biochemistry*, **10,** 1617 (1971).
(124) R. F. Steiner, *Arch. Biochem. Biophys.*, **49,** 400 (1954).
(125) W. F. McDevit and F. A. Long, *J. Am. Chem. Soc.*, **74,** 1773 (1952).
(126) E. E. Schrier and E. B. Schrier, *J. Phys. Chem.*, **71,** 1851 (1967).
(127) J. C. Nichol, *J. Biol. Chem.*, **243,** 4065 (1968).
(128) G. Nemethy, I. Z. Steinberg, and H. A. Scheraga, *Biopolymers*, **1,** 43 (1963).
(129) M. W. Pandit and M. S. N. Rao, *Biochemistry*, **13,** 1048 (1974).
(130) L-S Hsu and K. E. Neet, *Biochemistry*, **12,** 586 (1973).
(131) V. D. Hoagland, Jr., and D. C. Teller, *Biochemistry*, **8,** 594 (1969).