Determination of Interaction Parameters for a Generalized Associating System from Colligative Data*

R. F. Steiner

ABSTRACT: A method has been developed for analyzing colligative data for an associating system involving any number of different monomers. The method is applicable to number-average molecular weight data and requires number-average molecular weights as a function of concentration for varying ratios of the total concentrations of the different components. From the concentration dependence of molecular weight for a constant composition, a function characterizing the mole fractions of all free monomer units can be computed. From the dependence of the latter function upon the relative concentration of one particular component, at constant relative concentrations of all other components, the mole fraction and the molar concentration of free monomer for that component can be computed. In this way the molarities of all free monomer units can be obtained and from this information the consecutive association constants can be computed. The method applies to systems where the monomers self-associate and requires no assumptions as to the mode of association.

In an earlier publication a method was described for evaluating the association constants for an interacting system involving two different monomer units from colligative data (Steiner, 1968). The present paper is concerned with an extension of this approach to the most general case with any number of different monomer units and no restrictions upon the modes of self-association and heterogeneous association. The generalized model is as follows, for an arbitrary number of monomer units, A, B, C, D, ... etc.

$$\begin{aligned}
&A + A \rightleftarrows A_2;\ K_{A_2} = \frac{[A_2]}{[A]^2}\\
&A_2 + A \rightleftarrows A_3;\ K_{A_3} = \frac{[A_3]}{[A]^3}\\
&\vdots\\
&A_{i-1} + A \rightleftarrows A_i;\ K_{A_i} = \frac{[A_i]}{[A]^i}\\
&B + B \rightleftarrows B_2;\ K_{B_2} = \frac{[B_2]}{[B]^2}\\
&\vdots\\
&B_{i-1} + B \rightleftarrows B_i;\ K_{B_i} = \frac{[B_i]}{[B]^i}\\
&C_{i-1} + C \rightleftarrows C_i;\ K_{C_i} = \frac{[C_i]}{[C]^i}\\
&A + B \rightleftarrows AB;\ K_{AB} = \frac{[AB]}{[A][B]}\\
&AB + B \rightleftarrows AB_2;\ K_{AB_2} = \frac{[AB_2]}{[A][B]^2}\\
&AB + C \rightleftarrows ABC;\ K_{ABC} = \frac{[ABC]}{[A][B][C]}\\
&ABC + A \rightleftarrows A_2BC;\ K_{A_2BC} = \frac{[A_2BC]}{[A]^2[B][C]}\\
&ABC + D \rightleftarrows ABCD;\ K_{ABCD} = \frac{[ABCD]}{[A][B][C][D]}, \text{ etc.}
\end{aligned} \tag{1}$$

The treatment described here will apply to number-average molecular weights obtained from colligative data, such as osmotic pressure, vapor pressure, etc. It would be applicable to molecules of any size. The only restriction will be the assumption of ideal behavior for all species.

Complex associating systems are encountered more often in nature than in the literature. Some random examples might be the association of a purine base with two different pyrimidines; the interaction of two different protein monomer units, as influenced by the binding of a cofactor; and the hybridization of two related protein species consisting of multiple subunits.

Theory

Colligative measurements upon a heterogeneous system yield the number-average molecular weight, M_n, and hence the total molarity, m, of the system

$$M_n = \Sigma M_i m_i / m = c/m \tag{2}$$

where M_i and m_i are the molecular weight and molar concentration, respectively, of the ith species and c is the total weight concentration.

The total molar concentration is given by

$$m = \Sigma K_{A_iB_jC_kD_l} \ldots [A]^i[B]^j[C]^k[D]^l \ldots \tag{3}$$

The total number of moles per unit volume, $[A_t]$, of species A is given by

$$[A_t] = \Sigma i K_{A_iB_jC_kD_l} \ldots [A]^i[B]^j[C]^k[D]^l \ldots = [A]\frac{\partial m}{\partial [A]} \tag{4}$$

* From the Bureau of Medicine and Surgery, Navy Department, Research task MR005.06-0005A. *Received October 15, 1969.* The opinions in this paper are those of the author and do not necessarily reflect the views of the Navy Department or the naval service at large.

where [A] = molar concentration of monomeric A and for any species X

$$[X_t] = [X]\frac{\partial m}{\partial [X]}$$

Also, m_t, the total number of moles per unit volume of all monomer species, is given by

$$m_t = \Sigma[X]\frac{\partial m}{\partial [X]}$$

$$m_t = \Sigma[X]\frac{\partial m}{\partial [X]} = 0 \tag{5}$$

If the ratios of the total molar concentrations of all species are held constant

$$[B_t] = \beta_B[A_t] \tag{6}$$

where β_B is a constant

$$[C_t] = \beta_C[A_t]$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$[X_t] = \beta_X[A_t]$$

Also $[X]\partial m/\partial [X] = [X_t] = \beta_X[A]\partial m/\partial [A]$, and eq 5 may be rewritten

$$m_t - (\Sigma\beta_X)[A]\frac{\partial m}{\partial [A]} = 0 \tag{7}$$

where X = A, B, C, etc., and $\beta_A = 1$. Or

$$\alpha_n - (\Sigma\beta_X)\frac{\partial \ln m}{\partial \ln [A]} = 0 \tag{8}$$

where α_n = number-average degree of polymerization

$$= M_n/M_0 = m_t/m$$

where M_0 is the number-average molecular weight in the absence of association. Thus

$$d \ln [A] = (\Sigma\beta_X)\alpha_n^{-1}\frac{dm}{m} \tag{9}$$

$$\ln [A] = \Sigma\beta_X \int \alpha_n^{-1}\frac{dm}{m} + F([B], [C], \ldots) \tag{10}$$

where F is an arbitrary function of the concentrations of all free monomer units other than A.

If a function ψ is defined from eq 10

$$\psi \equiv \ln [A] - (\Sigma\beta_X)\int \alpha_n^{-1}\frac{dm}{m} - F = 0 \tag{11}$$

Then

$$\frac{\partial m}{\partial [A]} = -\frac{\partial \psi}{\partial [A]}\Big/\frac{\partial \psi}{\partial m} = \alpha_n m/[A]\Sigma\beta_X \tag{12}$$

and

$$\frac{\partial m}{\partial [B]} = -\frac{\partial \psi}{\partial [B]}\Big/\frac{\partial \psi}{\partial m} = -\frac{\partial F}{\partial [B]}\alpha_n m/\Sigma\beta_X \tag{13}$$

Equation 7 may be rewritten as

$$\alpha_n - \beta_A\frac{\partial \ln m}{\partial \ln [A]} - \frac{\partial \ln m}{\partial \ln [B]} - \left(\sum_{X\neq A,B}\beta_X\right)\frac{\partial \ln m}{\partial \ln [A]} = 0 \tag{14}$$

Substituting eq 12 and 13 into 14

$$\alpha_n - \beta_A\alpha_n/\Sigma\beta_X + \alpha_n\frac{\partial F}{\partial \ln [B]}\Big/\Sigma\beta_X - \left(\sum_{X\neq A,B}\beta_X\right)\times \alpha_n/\Sigma\beta_X = 0 \tag{15}$$

and

$$\beta_B = -\frac{\partial F}{\partial \ln [B]} \tag{16}$$

$$F = -\beta_B \ln [B] + G([C], [D], \ldots) \tag{17}$$

$$\ln [A] + \beta_B \ln [B] = (\Sigma\beta_X)\int \alpha_n^{-1}\frac{dm}{m} + G([C], [D], \ldots) \tag{18}$$

where G is an arbitrary function of the molar concentrations of all free monomer units other than A or B.

Repetition of the above process yields, finally

$$\Sigma\beta_X \ln [X] = \Sigma\beta_X \int \alpha_n^{-1}\frac{dm}{m} + C \tag{19}$$

where C is a constant of integration. Subtracting $\Sigma\beta_X \ln m$ $(= \Sigma\beta_X \int dm/m)$ from both sides, we obtain

$$\Sigma\beta_X \ln y_X = \Sigma\beta_X \int (\alpha_n^{-1} - 1)\frac{dm}{m} + C \tag{20}$$

where $y_X = [X]/m$ = mole fraction of free monomeric X.

In the absence of association, the first term on the right hand side vanishes, so that

$$C = \Sigma\beta_X \ln (\beta_X/\Sigma\beta_X) \tag{21}$$

Upon substituting, we have, finally

$$\Sigma\beta_X \ln y_X = \Sigma\beta_X \int (\alpha_n^{-1} - 1)\frac{dm}{m} + \Sigma\beta_X \ln (\beta_X/\Sigma\beta_X) \tag{22}$$

$$\equiv \phi_{m,\beta}$$

If only two different monomer units are present, eq 22 reduces to eq 24 of the previous paper (Steiner, 1968).

$$\ln y_A + \beta_B \ln y_B = (1 + \beta_B) \int (\alpha_n^{-1} - 1) \frac{dm}{m} + \ln [1/(1 + \beta_B)] + \beta_B \ln [\beta_B/(1 + \beta_B)] \quad (23)$$

If only one monomer is present, it reduces to the familiar form

$$\ln y_A = \int (\alpha_n^{-1} - 1) \frac{dm}{m} \quad (24)$$

In analyzing a real system, one would ordinarily obtain a series of experimental curves of α_n *vs.* m, in which all β's are held constant except one, β_Y. If the right hand side of eq 22 is designated as $\phi_{m,\beta}$, one may obtain $\phi_{m,\beta}$ as a function of β_Y at a constant value of m by interpolation of the above family of curves. At constant m

$$m([A], [B], \ldots, [Y], \ldots) = C \quad (25)$$

$$m - C = 0 \quad (26)$$

$$\left(\frac{\partial [Y]}{\partial [A]}\right)_m = -\frac{\partial m}{\partial [A]} \Big/ \frac{\partial m}{\partial [Y]} = \left(\frac{\partial y_Y}{\partial y_A}\right)_m = -\frac{\partial m}{\partial y_A} \Big/ \frac{\partial m}{\partial y_Y} \quad (27)$$

or

$$\left(\frac{\partial \ln y_Y}{\partial \ln y_A}\right)_m = -\frac{\partial m}{\partial \ln y_A} \Big/ \frac{\partial m}{\partial \ln y_Y} = -\frac{\partial m}{\partial \ln [A]} \Big/ \frac{\partial m}{\partial \ln [Y]} = -1/\beta_Y \quad (28)$$

$$d \ln y_A = -\beta_Y d \ln y_Y$$

$$\ln y_A = -\int \beta_Y \, d \ln y_Y = -\beta_Y \ln y_Y + \int \ln y_Y d\beta_Y + H(y_X) \quad (29)$$

where H is an arbitrary function of all y_X with $X \neq A, Y$.

Repetition of this process for a third species Z in turn yields

$$\ln y_A = -\beta_Y \ln y_Y + \int \ln y_Y d\beta_Y - \beta_Z \ln y_Z + \int \ln y_Z d\beta_Z + H'(y_X) \quad (30)$$

where $H'(y_X)$ is an arbitrary function of the remaining y_X, with $X \neq A, Y, Z$.

By repetition of this process one obtains finally

$$\ln y_A = -\sum_{X \neq A} \beta_X \ln y_X + \sum_{X \neq A} \int \ln y_X d\beta_X \quad (31)$$

Substitution in eq 22 yields

$$\sum_{X \neq A} \int \ln y_X d\beta_X = \phi_{m,\beta} \quad (32)$$

We then have

$$\ln y_Y = \left(\frac{\partial \phi}{\partial \beta_Y}\right)_{m,\beta_{X \neq Y}} \quad (33)$$

Equation 33 is valid, irrespective of the complexity of the system, and provides a general method for obtaining the monomeric concentration of any component from colligative data.

By repetition of the above measurements for each component in turn, one may obtain the values of each y_X and hence each value of [X] as a function of m for various values of the β_X's.

For a system containing three monomers, A, B, and C, one would obtain curves of α_n^{-1} as a function of m for varying values of β_B, holding B_C equal to β_C^*. In this way $\phi_{m,\beta}$ would be obtained as a function of β_B for various values of m, using eq 22. From eq 33 y_B and [B] would be computed as a function of β_B for the chosen set of values of m. Selecting a particular value β_B^*, a second set of curves of α_n^{-1} as a function of m would be obtained for a series of values of β_C, holding β_B equal to β_B^*. In this way y_C and [C] would be obtained for $\beta_C = \beta_C^*$ and the same values of m as before. At this stage y_B and y_C are known as a function of m for $\beta_B = \beta_B^*$ and $\beta_C = \beta_C^*$. Next y_B and [A] may be computed from eq 22. The process may be repeated, if desired, for other values of β_B^* and β_C^*.

In practice, the slopes corresponding to eq 33 could best be evaluated by fitting ϕ to a polynomial in β_X, using a suitable computer program. A number of simple programs are available, such as the BASIC program POLFIT, available in the General Electric library. The slope can be evaluated at any value of β by differentiating the polynomial.

The integral of eq 22 may be evaluated either graphically or by trapezoidal integration, using a suitable computer program.

In this way m may be obtained as a function of the molarities of free monomer of each of the species present in the system, corresponding to eq 3. This may be done for a wide range of values of [A], [B], and [C]. In particular, by appropriate extrapolation, m may be obtained as a function of the concentration of any one monomer species, for constant values of the other two.[1]

For the case of three different species, A, B, and C, which both self-associate and associate with each other in all possible combinations, it is desirable to make separate measurements upon solutions containing only one species.

Defining S_A, S_B, and S_C as the values of m for solutions

[1] This could be done as follows for the three component system. (1) For each of the selected values of m, curves of ϕ *vs.* β_B are obtained for 5–6 values of β_C. (2) From these, curves of [B] *vs.* β_B are constructed for each value of β_C and the corresponding values of m. (3) In a similar manner curves of [C] *vs.* β_C are constructed for a set of values of β_B. (4) Selecting the values [B*] and [C*], curves of β_B *vs.* β_C consistent with these values are constructed for [B*] and for [C*] at each value of m. The intersection of each pair of curves yields, for each value of m, the values of β_B^* and β_C^* consistent with the (constant) values [B*] and [C*]. (5) From the curves of ϕ *vs.* β_B for various values of β_C, a curve of ϕ *vs.* β_C for $\beta_B = \beta_B^*$ is constructed for each value of m. (6) The values of ϕ^* corresponding to β_B^* and β_C^* at the corresponding value of m is interpolated and the values of [A] for each value of m, with [C] = [C*] and [B] = [B*], are obtained.

containing only A, B, or C, respectively, we then have

$$S_A = [A] + K_{A_2}[A]^2 + K_{A_3}[A]^3 + \ldots = \Sigma K_{A_i}[A]^i$$

$$S_B = [B] + K_{B_2}[B]^2 + K_{B_3}[B]^3 + \ldots = \Sigma K_{B_i}[B]^i \quad (34)$$

$$S_C = [C] + K_{C_2}[C]^2 + K_{C_3}[C]^3 + \ldots = \Sigma K_{C_i}[C]^i$$

where $K_{A_1} = K_{B_1} = K_{C_1} = 1$. If S_A, S_B, and S_C are known from independent measurements, they may be subtracted from m to yield

$$m - S_A - S_B - S_C = K_{AB}[A][B] + K_{AC}[A][C] + K_{BC}[B][C] + K_{ABC}[A][B][C] + K_{AB_2}[A][B]^2 + \ldots \equiv T_{ABC} \quad (35)$$

Also, when β_B and [B] are zero, S_B is zero and

$$T_{AC} \equiv m - S_A - S_C = K_{AC}[A][C] + K_{AC_2}[A][C]^2 + K_{A_2C}[A]^2[C] + \ldots \quad (36)$$

and

$$\left[\frac{d(T_{AC}/[A])}{d[C]}\right]_{C=0} = K_{AC} \text{ (for [A] constant); or}$$

$$\lim_{[A],[C]\to 0} \{T_{AC}/[A][C]\} = K_{AC}$$

$$\left[\frac{d}{d[C]}\{(T_{AC} - K_{AC}[A][C])/[A][C]\}\right]_{C=0} = K_{AC_2} \text{ (for [A] constant)}$$

Similarly, when C or A is absent, one may define

$$T_{AB} = m - S_A - S_B = K_{AB}[A][B] + K_{AB_2}[A][B]^2 + \ldots$$

$$T_{BC} = m - S_B - S_C = K_{BC}[B][C] + K_{BC_2}[B][C]^2 + \quad (37)$$

and

$$K_{AB} = \left[\frac{d(T_{AB}/[A])}{d[B]}\right]_{B=0}, \text{ etc. (for [A] constant)}$$

$$K_{BC} = \left[\frac{d(T_{BC}/[B])}{d[C]}\right]_{C=0}, \text{ etc.}$$

In this way, using measurements made in the absence of one component, K_{AB}, K_{AC}, and K_{BC} may be evaluated. Also, since T_{AB}, T_{AC}, and T_{BC} are known as functions of [A], [B], and [C], they may be subtracted from T_{ABC} to give

$$T_{ABC} - T_{AB} - T_{BC} - T_{AC} \equiv U = K_{ABC}[A][B][C] + K_{A_2BC}[A]^2[B][C] + K_{AB_2C}[A][B]^2[C] + K_{ABC_2}[A][B][C]^2 + \ldots \quad (38)$$

We also have

$$U/[A][B][C] = K_{ABC} + K_{A_2BC}[A] + K_{AB_2C}[B] + K_{ABC_2}[C] + \ldots \quad (39)$$

$$\lim_{\substack{[A]\to 0\\[B]\to 0\\[C]\to 0}} \{U/[A][B][C]\} = K_{ABC} \quad (40)$$

and, for constant β's

$$\lim_{[A],[B],[C]\to 0} \{(U/[A][B][C] - K_{ABC})/[A]\} = K_{A_2BC} + \beta_B K_{AB_2C} + \beta_C K_{ABC_2}, \text{ etc.} \quad (41)$$

In principle, the association constants characteristic of the mixed species might be determined from eq 40 and 41, using data for different values of β_B and β_C. Thus, if three values of the right hand side of eq 41 are obtained for three different values of β_B or β_C, a set of three simultaneous linear equations is obtained, which may be solved for K_{A_2BC}, K_{AB_2C}, and K_{ABC_2}.

In practice, it would generally be preferable to use multiple regression analysis. Equation 39 may be regarded as a series expansion of $U/[A][B][C]$, for which the constant coefficients are to be determined. This is of the type

$$y = \alpha_0 + \sum_i \alpha_i X_i \quad (42)$$

where the α's are constant and the X_i are variables and correspond to [A], [B], [C], [A][B], [A][C], etc. If, as in the present case, y is known for a range of values of the X_i's then the problem becomes a standard one in multiple regression analysis. A number of suitable computer programs are available for obtaining the coefficients, such as the ALGOL program MULRG in the General Electric library. The paper of Magar (1969) or the book of Draper and Smith (1966) may be consulted for details.

A third, somewhat more laborious approach may be used if, by suitable interpolation, U is obtained as a function of a single free monomer concentration, holding the other two constant. Regression analysis may also be used in this case (Magar, 1969). Thus, if particular constant values of [A] and [B] are chosen, eq 39 may be regarded as a polynomial in [C], or if [A] and [C] are held constant, as a polynomial in [B], etc. Application of the statistical methods described by Magar, whose paper should be consulted for details, permits evaluation of the coefficient of each power of [C]. Thus, for [A] and [B] held constant at [A*] and [B*], eq 38 may be written

$$U = \alpha_1[C] + \alpha_2[C]^2 + \ldots \quad (43)$$

where

$$\alpha_1 = K_{ABC}[A^*][B^*] + K_{A_2BC}[A^*]^2[B^*] + K_{AB_2C}[A^*][B^*]^2 = \ldots \quad (44)$$

Repetition for different sets of values of [A*] and [B*] yields a set of simultaneous linear equations, which may be solved for the association constants. This approach has the advantage of utilizing all the data and avoiding excessive dependence upon the points at low concentrations.

A Simple Example. If additional information is available,

this procedure can often be greatly simplified. Consider the system

$$\begin{aligned} A + B &\rightleftarrows AB \\ B + C &\rightleftarrows BC \\ AB + C &\rightleftarrows ABC \\ BC + A &\rightleftarrows ABC \end{aligned} \tag{45}$$

Here A, B, and C do not self-associate, A does not combine with C, and AB, BC, and ABC are the only mixed species. This would correspond, for example, to the simultaneous association of the "double-headed" Bowman–Birk soy bean inhibitor with trypsin and α-chymotrypsin (Steiner and Frattali, 1969).

Here, we have

$$m - [B_t] = [A] + [C] \tag{46}$$

This provides a relationship between [A] and [C], in terms of the known quantities m and $[B_t]$. One needs thus to apply eq 33 only to a single set of curves of α_n^{-1} *vs.* m for varying values of β_B to obtain [B]. Use of this in conjunction with eq 22 and 46 provides all the needed information.

Discussion

The theory outlined in the preceding sections can be applied to interactions in solution of either biopolymers or small molecules, as well as mixtures of the two. Perhaps the most generally useful application would be the combined use of eq 22 and 33 to determine the concentration of the free monomeric form, and hence the mole fraction combined, of any component of a complex interacting system. This may be of particular utility for mixtures of proteins, for which alternative methods for determining free monomer concentrations may not be readily applicable. For example, in a complicated mixture in which species A, B, and C are present one might use this approach to determine the influence of A upon the combination of B with C, etc.

The problem of determining all the association constants for a system containing three or more components is admittedly laborious. Assuming that a minimum of 5–6 values of β_X must be examined to permit application of eq 33, it can be estimated that a total of about 15 curves of α_n *vs.* m would have to be obtained in order to characterize a three component system for which there are no restrictions upon the modes of association. However, the current availability of rapid membrane and vapor pressure osmometry devices reduces the time consumed by such an operation to a reasonable figure. Of course, in most cases it may be known from other information that several possible types of association do not occur, thereby simplifying the problem considerably. For example, for the three component case considered above, if the monomers did not self-associate and if no association occurred between one of the three possible monomer pairs, the number of α_n^{-1} *vs.* m curves required would be reduced to 5 to 6.

A further simplification becomes possible if an independent method is available for estimating the free monomer concentrations of one or more components. For example, if one component is a small molecule, it may be possible to combine equilibrium dialysis with colligative measurements to obtain its concentration directly.

References

Draper, N. R., and Smith, H. (1966), Applied Regression Analysis, New York, N. Y., Wiley.

Magar, M. E. (1969), *J. Theor. Biol. 22*, 291.

Steiner, R. F. (1968), *Biochemistry 7*, 2201.

Steiner, R. F., and Frattali, V. (1969), *J. Agr. Food Chem. 17*, 513.