

DETECTION OF HETEROGENEITY IN SELF-ASSOCIATING SYSTEMS

DAVID A. YPHANTIS, JOHN J. CORREIA, MICHAEL L. JOHNSON* AND GAY-MAY WU**

Biochemistry and Biophysics Section, Biological Sciences Group and Institute of Materials Science, University of Connecticut, Storrs, Connecticut 06268

ABSTRACT

Heterogeneity in self-associating systems is discussed and extant methods of detecting such heterogeneity are reviewed. A proposal is made to detect heterogeneity by the variation in the apparent association constants, \hat{K} , observed in sedimentation equilibrium of different loading concentrations, c° . The sensitivity of this procedure is illustrated. It is shown that a correct model of the self-association examined is not needed, provided the ultracentrifuge data for all the loading concentrations is truncated so as to span the same observation concentration, $c(r)$, range for all the c° . Expressions are derived for the quantitative estimation of small extents of heterogeneity both from the diagnostic lack of overlap of curves of $M_k(r)$, the apparent molecular weights, vs $c(r)$ for the different c° and from the variation of \hat{K}_{app} for the different c° . These expressions are strictly valid only for systems where there is no hetero-association between the main self-associating component and the contaminating component. For other systems the expressions presented provide lower bounds for the extent of heterogeneity.

INTRODUCTION

The estimation of standard free energies of association (and derived parameters) from observed apparent association constants in self-associating systems can be influenced strongly by the presence of small amounts of contamination by molecules similar to the component of interest¹. Generally this fact would be considered to be a nuisance, especially when interest is focussed on the properties of the main component. However, this effect can provide a sensitive indication of heterogeneity in certain specific cases, allowing detection of small quantities of molecules closely related to the main component.

Heterogeneity in self-associating systems may have one or more of the following sources:

a) the presence of adventitious contaminating solutes with a molecular weight that is generally, different than that of the monomer of the component of interest.

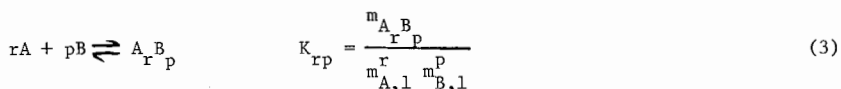
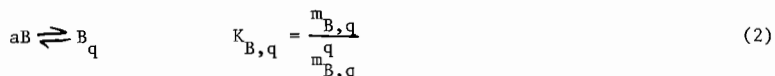
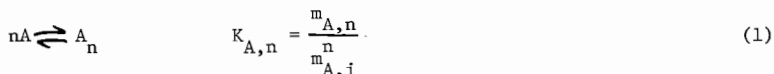
* Present address: Clinical Endocrinology Branch, National Institute of Arthritis, Metabolism and Digestive Diseases, National Institutes of Health, Bethesda, MD 20014

** Present address: Department of Biochemistry, Brandeis University, Waltham, MA 02154

- b) the presence of irreversible aggregates of the main component;
 c) the presence of "incompetent monomer" incapable of self-association;
 d) the presence of other chemical forms of the self-associating solute with, perhaps, slightly different sequence, differing location of amides, variations in carbohydrate substituents, etc., as in isozymes, with differing association constants but with effectively the same monomer molecular weights; and
 e) the presence of metastable conformers of the self-associating solute that are not in rapid equilibrium with each other over the time course of the experiment and that exhibit differing association behavior.

The first two types may be considered as "trivial heterogeneity" and generally can be removed from the sample by suitable preparative procedures. For example, irreversible aggregates in biochemical preparations frequently may be eliminated by simple gel exclusion chromatography immediately before the self-association experiment.

The association behavior of heterogeneous systems of the last three types, where the contaminating component, B, has the same monomer molecular weight as the component of primary interest, A, generally can be represented by equilibria of the form:



with various n , p , q and r , where $m_{i,j}$ is the molarity of the molecular species defined as the j -mer of the i -th component.

The problem treated here is the detection of small amounts of component B in the presence of large amounts of component A. To make this problem tractable we ignore the possible presence of mixed equilibria and set $K_{rp} = 0$ in eq. (3). [Since distinction between different components is based completely on their association constants, it is possible, with both K_{rp} and $K_{B,m}$ exactly equal to the corresponding $K_{A,n}$, to make the heterogeneity completely invisible to the observer. In general the presence of significant contributions from mixed equilibria will decrease the sensitivity of detection of heterogeneity.] For simplicity we treat in detail only systems where a single simple equilibrium, eq. (1), is effective; the presence of successive equilibria, with different values of n , can lead to interesting effects, but

these will be discussed elsewhere. It can be shown that contributions from the self-association of small enough amounts of component B can be completely ignored. Thus we consider in detail here only systems for which the assumptions

$$K_{B,m} = K_B = 0 \quad (4a)$$

$$K_{rp} = 0 \quad (4b)$$

are valid and for which there exists only a single value of n for the self-association of component A.

Unless otherwise specified, the equilibrium constants will be the molar association constants defined in equations (1) and (2) for the ideal self-association of the two components. The weight concentrations corresponding to the $m_{i,j}$ are the $c_{i,j}$ and are given in g/L. The heterogeneity is specified by the parameter θ ($=c_{B,T}/c_{A,T}$), the ratio of the total amount of component B to the total amount of component A in the solution being examined.

It is instructive to examine the error introduced into estimates of the standard free energy of association from the presence of a nonassociating component of the same size as the associating monomer. Figure 1 presents the differences between the values of $\Delta F^\circ/RT$ ($= -\ln \hat{K}$) calculated from \hat{K} , the apparent association constant that would be observed for such a heterogeneous

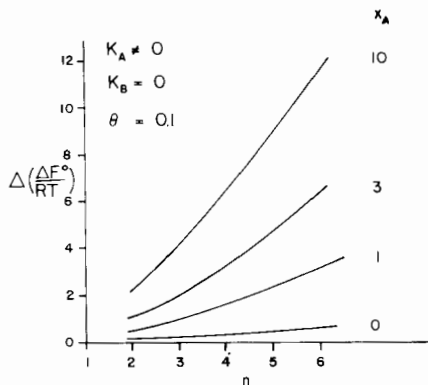


Fig. 1. Differences between true and apparent values of the standard free energy of association calculated from the corresponding association constants for systems with 10% of a non-associating component B, as a function of n , the degree of association, and for several extents of association, as measured by x_A , the molar ratio of n -mer to monomer for the self-associating component A. The values presented as $\Delta(\Delta F^\circ/RT)$ are equal to $-\ln(\hat{K}/K_A)$, where \hat{K} is the apparent association constant estimated for the total system and K_A is the true association constant of component A.

system, and the values of $\Delta F^\circ/RT (= -\ln K_A)$, calculated for K_A , the actual association constant of the main component. These differences are presented for associations ranging from monomer-dimer to monomer-hexamer, for varying extents of association as specified by $x_A (= m_{A,n}/m_{A,1})$, the ratio of polymer to monomer molarity for component A, and for heterogeneity specified by $\theta = 0.1$. It is obvious that a) rather large errors in ΔF° can be introduced by this modest heterogeneity, b) these errors in ΔF° increase with increasing degree of polymerization, n, and c) these errors become much larger with higher extents of association.

Clearly it is essential to have procedures available to detect and, if possible, quantitate such heterogeneity so as to be able to avoid potential pitfalls associated with this effect. We start by noting that the observed variation of apparent association constant with extent of reaction (Figure 1) may provide one such procedure to detect heterogeneity. Accordingly, we next derive expressions for the effects of heterogeneity on the apparent association constant.

EFFECTS OF HETEROGENEITY ON APPARENT ASSOCIATION CONSTANTS

The apparent association constant, \hat{K} , is defined as

$$\hat{K} = \frac{m_{T,n}}{m_{T,1}^n} = \frac{m_{A,n} + m_{B,n}}{(m_{A,1} + m_{B,1})^n} = \frac{K_A m_{A,1}^n}{(m_{A,1} + m_{B,1})^n} = \frac{m_{A,1} x_A}{(m_{A,1} + m_{B,1})^n} \quad (5)$$

where $m_{T,1}$ is the total molarity of all j-mers and where we have substituted relations (1) and (2) for the equilibria, made assumptions (4), and introduced the symbol x_i for the molar ratio of the n-mer to monomer for component i. We note that m_i^0 , the total molarity of the monomers of component i, is given by

$$m_A^0 = m_{A,1} + n m_{A,n} = m_{A,1} (1 + n x_A) \quad (6a)$$

$$m_B^0 = \theta m_A^0 = m_{B,1} \quad (6b)$$

where the last relation of eq.(6b) is valid only for $K_B = 0$. Substituting the values of the $m_{i,1}$ from eqs.(6) into eq.(5) and eq.(1) we obtain

$$\hat{K} = \frac{x_A m_{A,1}}{[m_A^0(1 + \theta) - n m_{A,1} x_A]^n} \quad (7a)$$

$$K_A = \frac{x_A m_{A,1}}{[m_A^0 - n m_{A,1} x_A]^n} \quad (7b)$$

Dividing eq.(7b) by eq.(7a), we obtain

$$\frac{\hat{K}_a}{K_a} = \frac{m_A^o - nm_{A,1}x_A + \theta m_A^o}{m_A^o - nm_{A,1}x_A} = \left[1 + \frac{\theta m_A^o}{m_A^o - nm_{A,1}x_A} \right]^n \quad (8)$$

Noting from eq.(6a) and the definition of x_A that

$$m_{A,1} = m_A^o / (1 + nx_A) \quad (9)$$

we then obtain the relation

$$\ln(\hat{K}/K_a) = -n \ln(1 + \theta < 1 + nx_A >) \quad (10a)$$

as generally applicable under assumptions (4). For θ sufficiently small we expand eq.(10) to obtain

$$\ln(\hat{K}/K_a) = -n\theta(1 + nx_A) \quad (10b)$$

Equations (10) predict the increasing differences between $\ln \hat{K}$ and $\ln K_a$ with increasing n and increasing extent of association illustrated in Figure 1.

The variations in apparent association constant with extent of association could, in principle, be used to infer association heterogeneity in a preparation. One simple method would be to compare the shapes of the curves for an apparent molecular weight average, say the apparent weight average molecular weight, of the sample as a function of concentration. However, such inferences would only be valid if there were no other source for variation of the apparent association constants (or, equivalently, of departure of the curve of the apparent molecular weight vs concentration from the curve for a homogeneous simple system) than heterogeneity. Unfortunately, one rarely has such assurances. Most systems exhibit the effects of some nonideality, which, at least qualitatively, can generate somewhat similar variations in \hat{K} (or in shapes of curves) as generated by heterogeneity. It is fairly common to find sequential self-association reactions, with increasing values of n . The variation of the apparent association constants observed for such complex systems readily may mask the contrary effects of modest amounts of heterogeneity. Even in the absence of such common complications, the sensitivity of detection of heterogeneity from variations in shape of the molecular weight curves (or from variations in $\ln \hat{K}$ with association) may be fairly low. For example, the curves for $M_w(c)$ vs c for a monomer dimer system with $\theta = 0.1$ may be fit, for $K_A c < 1$, rather well (with maximum deviations in $M_w(c)$ less than about 1%), assuming a

single constant value of the association constant ($K \approx 0.768$). Since experimental error is generally larger than these deviations, the heterogeneity would be virtually invisible to this approach. A more general procedure is needed, preferably one that does not depend on the assumption of the complete correct model for the association of the main component of the system.

DETECTION OF HETEROGENEITY BY EQUILIBRIUM SEDIMENTATION

Such a procedure was suggested by Squire and Li² for sedimentation equilibrium experiments. It is based on the phase rule; for a single-phase system with two chemical components, the solvent and the self-associating solute, there are three degrees of freedom. At constant temperature and pressure the behavior of the system is completely determined by the solute concentration and all the properties of the system must be a function of concentration only. In practice most systems fulfill the requirement of two components at constant pressure by approximation: Although the solvent component itself usually consists of more than one component - for biochemical systems typically this would be water and one or more buffer components - the solvent may be considered as a single component in the thermodynamic sense if its composition is effectively invariant over the conditions of interest; similarly if the volume change on associations is small, as it usually is, then the requirement of constant pressure may be relaxed. (If the volume change is known not to be negligible, then the effects of pressure may be compensated for by calculation).

Since all properties of the system depend solely on the solute concentration, the various point average molecular weight moments, such as $M_w(r)$ and $M_z(r)$, the weight and z-average molecular weights at radius r , must be functions only of $c(r)$, the concentration at the point r . Accordingly for solutes that consist of a single thermodynamic component the curves of $M_k(r)$ vs. $c(r)$ for different loading (initial) concentrations, c^0 , must be exactly superposable. Conversely, divergence of these curves implies that the solute consists of more than one component and thus must be heterogeneous. By similar reasoning one may compare the curves of $M_w(r)$ vs. $c(r)$ for the same loading concentration but observed for sedimentation equilibrium at differing angular velocities, again with lack of coincidence indicating the presence of more than one thermodynamic component. Generally the procedure of examining different loading concentrations in the same rotor and at the same time would be preferred since it tends to minimize extraneous variations in experimental conditions, such as differences in speed, in temperature, and in history between different experiments.

A calculated example of this procedure to detect heterogeneity in associating systems is furnished by Figure 2. The several curves of apparent weight average

molecular weight shown have been calculated for the indicated loading concentrations, assuming a heterogeneous associating system with two solute components: main component A, which dimerizes reversibly with $K_A = 1$, and component B, present at the level of 1/10 that of component A, and which is incapable of any association. The moderate heterogeneity of this system is obvious.

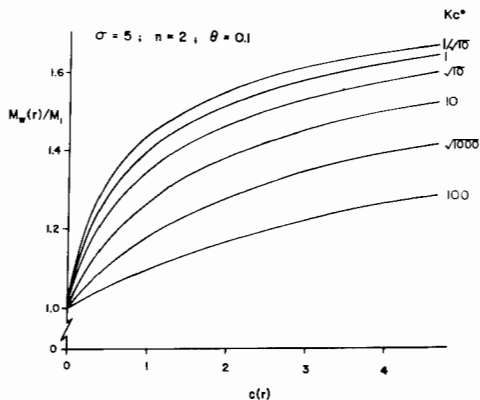


Fig. 2. Apparent weight-average molecular weights, at each point, $M_w(r)$, presented as a function of the concentration at that point, $c(r)$, calculated for several loading concentrations, c^0 , for a monomer-dimer system with $K_d = 1$, $\sigma_1 = 5 \text{ cm}^{-2}$, $\Delta\xi = 1.965 \text{ cm}^2$ and with $\theta = 0.1$, corresponding to the presence of about 9% of "incompetent monomer". The lack of overlap of these curves is evidence that the solute consists of more than one thermodynamic component, provided that there is no pressure dependence of the association and provided that the solvent can be represented as a single thermodynamic component.

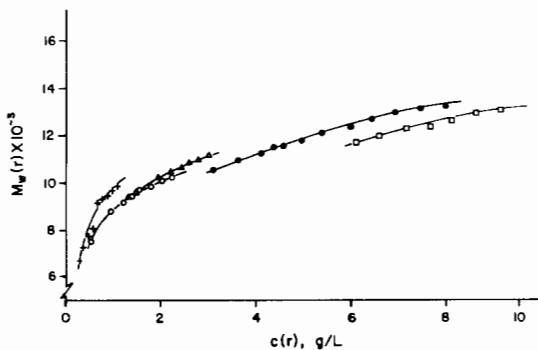


Fig. 3. Apparent weight average molecular weights, $M_w(r)$, as a function of $c(r)$ observed by Jeffrey and Coates³ for insulin at pH 2.6, 0.1 ionic strength at 25°C. The divergence of the curves for different loading concentrations indicates heterogeneity in association behavior.

An illustration of the application of this procedure to data from the literature is given in Figure 3. Here we present data from Jeffrey and Coates³ for the self-association of insulin at pH 2.6, replotted to distinguish data corresponding to the different loading concentrations and the different equilibrium speeds. The lack of coincidence of the several curves indicates significant association heterogeneity.

The detection of association heterogeneity in β -lactoglobulin A by this procedure and the subsequent partial resolution of the protein into nearly homogeneous fractions with differing self-association constants has been described by Roark and Yphantis⁴. Numerous other applications of this procedure can be found in the literature.

A useful variant of the procedure of comparing curves of an apparent molecular weight moment, $M_k(r)$, vs the observation concentrations, $c(r)$, for different loading concentrations is the replacement of the observation concentrations by their corresponding concentration gradients, $\frac{dc}{d\xi}(r)$. These concentration gradients are given, for two component systems, by

$$\frac{dc}{d\xi}(r) = \frac{1}{r} \frac{dc}{dr} = \frac{c(r) M_w(r) (1-\bar{v}\rho) \omega^2}{RT \left\langle 1 + c \frac{\partial \ln y}{\partial \ln c} \right\rangle_{P,T}} = c(r) \sigma_w(r) \quad (11)$$

where $\xi = r^2/2$, \bar{v} is the partial specific volume of the solute, ρ is the solvent density, ω the angular velocity, R the gas constant, T the absolute temperature, y the activity of the solute on the c -scale, P the pressure and $\sigma_w(r)$ is the effective reduced weight average molecular weight of the solute. It is obvious from eq.(11) that for homogeneous systems $\frac{dc}{d\xi}(r)$ is a function of $c(r)$ alone, provided only that variations in $\bar{v}\rho$ may be neglected, since both $M_w(r)$ and y are functions of concentration alone. This variant is useful when the concentrations $c(r)$ are not explicitly and absolutely known, for example when the assumption of mass conservation throughout the ultracentrifuge cell may not be reliable because of solute adsorption, precipitation, etc. An example of the application of this variant may be found in studies of the self-association of tyrocidine-B⁵.

The several curves of Figure 2 correspond to differing values of the apparent dimerization constants, K_2 , ranging from 0.664 at the lowest loading concentration to 0.060 at the highest c° presented. This variation of apparent association constant with loading concentration appears to be useful as a sensitive indicator of heterogeneity in self-associating systems. The degree of sensitivity depends on two factors: the uncertainty in the estimated values of the association constant, \hat{K} , obtained from the sedimentation equilibrium experiments and the magnitude of the change in the apparent equilibrium constants with variations in loading concentrations.

The intrinsic variation in \hat{K} with loading concentration depends on the properties of the associating system under study and on the product $\sigma_1 \Delta \xi = (\xi_b - \xi_a) \sigma_1$, where ξ_a and ξ_b are the values of $r^2/2$ at the meniscus and at the base of the solution column, and where σ_1 is the effective reduced molecular weight ($= \omega^2 s / D = \omega^2 M(1 - \bar{v} \rho) / RT$)—where s and D are the sedimentation and diffusion coefficient of the ideal monomer. In addition to these factors, the uncertainties in the \hat{K} critically depend on the details of the experimental procedures used to estimate the \hat{K} . The parameters and conditions used in our estimates of these uncertainties naturally mirror our experimental techniques:

The ultracentrifuge cells⁶ used in our experiments provide six channels for the observation of three solution-solvent pairs. This facilitates the simultaneous examination of different loading concentrations in the same rotor at the same time, speed and temperature. These cells may be cleaned without disassembly and can provide reproducible patterns even for speeds above 50,000 rpm. The high pressure mercury-arc lamp usually used as interferometer light source has been replaced by a pulsed argon-ion laser^{7,8,9} fired in phase with the ultracentrifuge rotor. A minicomputer system controls synchronization of the laser with the rotor so as to illuminate the appropriate cell or counter-weight at the designated time intervals with the selected number of flashes, and advances the motorized film transport. The resultant Rayleigh interferograms have been measured by one of three automatic image measurement systems: The first system, patterned after that of DeRosier, Munk and Cox¹⁰, scans the images mechanically in a raster pattern and records relative light intensities as a function of position; fringe shifts are then calculated by Fourier transforms from these recorded intensities. Unfortunately our mechanical raster scanning system is time consuming, requiring almost 90 minutes for the scan of an interferogram at 300 radial points within one solution channel. Accordingly, this system has been replaced by a television based system¹¹ that utilizes the T.V. raster scan to provide the "y-scan" across the interference fringes. This speeded up the data acquisition by nearly two orders of magnitude. Unfortunately this T.V. system exhibited only limited resolution at moderately high gradients and has recently been replaced, in turn, by our currently operational system. This third system utilized a 1024-element self-scanned photodiode array to perform the fringe scan¹². It has been set up using a modified "real time" Walsh transform to obtain the phase of the fringes. This system measures, calculates and records data from two radial points per second with a reproducibility of better than 10^{-3} fringe on appropriate photographic exposures.

The overall reproducibility of our measurement system, for the net fringe displacements of the solutions corrected for blank distortions, ranges from 1 to 3 μm fringe displacement ($0.3 - 1 \times 10^{-2}$ fringe). For our discussions

we assume the value of 3 μm , since such reproducibility can be obtained by careful manual measurements on interferograms obtained with the usual AH-6 mercury arc light source^{6,10,13}.

These measured fringe displacements are then used to characterize the system of interest through one or more of our data analysis programs: BIOSPIN, ORTHO and NONLIN. All these programs accept data in the form of fringe displacements as a function of comparator x-coordinates, for both the run itself and its corresponding blank, obtained by running water vs water in the same cell, without disassembly, before and/or after the run itself. The blank is subtracted, after appropriate interpolation, and the blank corrected data further processed. BIOSPIN² calculates several point average apparent molecular weight averages, such as the number, weight, z and (z+1)-averages. ORTHO¹⁴ requires the user to supply a value of σ_1 for the monomer, and a set of m integers, n_i , to specify the various n-mers thought to be present in the system being studied. This program then fits the observed fringe displacements, y_i , to the form

$$y_j = A_0 + \sum_{i=1}^m A_i \exp(n_i \sigma_1 [\xi_j - \xi_1]) \quad (12)$$

where A_0 is the program's estimate of the fringe offset at the first data point and the A_i are the concentrations in terms of fringe displacement at the first data point (where $\xi = \xi_1$) that are estimated for the various n_i -mers. The various K_i are then obtained from these A_i by applying the appropriate formulation for the equilibrium constant desired. As the name of the program implies, the actual fitting is by means of orthogonal functions generated from the set of the $\exp(n_i \sigma_1 [\xi_j - \xi_1])$ over all the data points ξ_j by a Gram-Schmidt orthonormalization¹⁵. It should be noted that nowhere is the questionable assumption of mass conservation made. The program NONLIN^{16,17} fits observed fringe displacements, y_i , to the more general form

$$y_j = A_0 + \sum_{i=1}^m A_i \exp(\lambda_i [\sigma_1 \langle \xi_j - \xi_1 \rangle - \sum_{k=1}^4 \frac{k+1}{k} B_k c^k(\xi_j)]) \quad (13)$$

where the λ_i , the analogs of the n_i , need not be integers. The value of σ_1 , the effective reduced molecular weight of the monomer, is not required here, but can be estimated, along with the λ_i , by the program. In addition the program can be used to estimate values of the several virial coefficients, B_k . These estimations may be carried out for single channels of data or for several channels of data analyzed jointly. Applications of this program to nonideal and/or associating systems have been reported^{13,18,19,20}.

We now wish to estimate the uncertainty in the association constants. With all measurement systems the maximum practical gradient is limited to concentration gradients, $(dc/d\xi)$, below about 15 mm fringe displacement/cm², in 12 mm thick centerpieces, by the effects of Wiener skewing²¹. This limitation to a definite maximum gradient that can be measured reliably in an equilibrium experiment must be taken into account in any examination of the reproducibility of determining the values of the K.

The introduction of this restriction proceeds most straightforwardly using the formulation associated with ORTHO to propagate errors and to keep track of analyses in terms of a linear system of equations. For such analyses one usually knows (or can determine) the value of σ_1 . One then has the minimum number of parameters to determine; one more than the number of molecular species analyzed for. For a linear system it can be shown that the root mean square (rms) uncertainty in the $\ln \hat{K}$, which will be denoted by $\langle \ln K \rangle$, is proportional to $\langle y \rangle$, the rms differences between the observed fringe displacements and their calculated values--or standard deviation of the y_1 from the fitting. The program ORTHO then readily provides the value of $\langle \ln K \rangle / \langle y \rangle$, the ratio of the effective standard deviation in $\ln \hat{K}$ to that of y , for various specific experimental arrangements, including the number and disposition of data points, the number and type of molecular species for which analysis is to be carried out, and the value of $\sigma_1 \Delta \xi$ --from the relation

$$\frac{\langle \ln K_n \rangle}{\langle y \rangle} = \left(\frac{\beta_n}{K_n m_1} \right)^2 + \left(\frac{\beta_1}{m_1} \right)^2 \quad (14)$$

where the $\beta_n = \partial \langle A_1 \rangle / \partial \langle y \rangle$ are coefficients available from the program and the value of m_1 is the monomer molarity at the first data point, ξ_1 . Values of m_1 and K_n to minimize the value of $\langle \ln K \rangle / \langle y \rangle$ subject to the imposition of a maximum gradient $(dc/d\xi)_{\max}$ at the base of the solution column may be chosen using undetermined LaGrangian multipliers. The resulting values of m_1 and K_n then specify conditions for the best possible estimate of K_n that can be obtained for the particular experimental arrangement considered and determine the magnitude of $\langle \ln K \rangle$ for any $\langle y \rangle$. Figure 4 presents such values of the uncertainty in $\ln \hat{K}_2$ for a typical experimental arrangement: 101 equispaced data points in a 3 mm high solution column and an rms fitting error, $\langle y \rangle$, of 3 μm . The errors in $\ln \hat{K}_2$ are presented on a logarithmic scale, as a function of σ_1 over the experimentally useful range from 0.1 to 10 cm^{-2} , for the estimation of a monomer-dimer equilibrium alone and, using formulations similar to eq.(14), for the simultaneous estimation of dimerization and trimerization equilibria, etc., up to the simultaneous determination of equilibria among monomer, dimer, trimer, tetramer and hexamer. It can be seen that the values of $\ln \hat{K}_2$ can be obtained

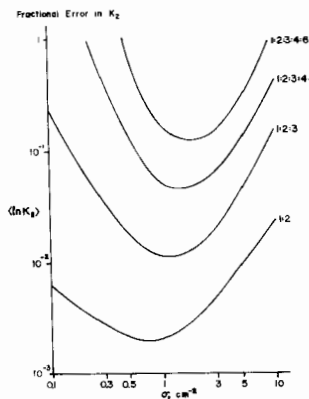


Fig. 4. Estimated minimal fractional uncertainty in apparent dimerization constants originating from an rms fitting deviation of $3 \mu\text{m}$ in fringe displacement, presented as a function of σ_1 , the effective reduced molecular weight of the monomer. The calculations assumed that the fringe displacements used as data were obtained at 101 (=N) equispaced radial points ranging from 6.4 to 6.7 cm and that these data were subject to the restraint that the maximum concentration gradient, $dc/d\xi = (dc/dr)/r = 15 \text{ mm fringe displacement/cm}^2$. The several curves correspond to the estimated uncertainties when the system was analyzed as containing the indicated species, e.g. the top curve corresponds to a system analyzed for the simultaneous presence of monomer, dimer, trimer, tetramer and hexamer. These estimated uncertainties are minimal values, corresponding to optimal values of c^0 and of the several values of the association constants, K_i . They represent the random errors associated with the best possible estimates of K_2 under the specified conditions for σ_1 , N, $\Delta\xi$, and $(dc/d\xi)_{\text{max}}$.

with an uncertainty of $\sim 3 \times 10^{-3}$ for $\sigma_1 \sim 1$, increasing gradually with either increasing or decreasing σ_1 , and strongly increasing, by a factor of about 4, for each additional component being simultaneously determined. The behavior for a monomer-tetramer equilibrium is quite similar. The main differences are in the locations of the minima which are shifted to lower values of σ_1 , with relatively minor quantitative differences between the actual values of $\langle \ln K_4 \rangle$ and of $\langle \ln K_2 \rangle$.

If the concentration gradient at the base of the solution column is made smaller than the limiting value of $15 \text{ mm fringe displacement/cm}^2$ that was used here, then the estimates of K_2 become less well-defined, with larger values of $\langle \ln K_2 \rangle$. Similarly if the values of K_2 differ from the optimal values for which Figure 4 was drawn, then the values $\langle \ln K_2 \rangle$ also increase. However, the rate of increase of $\langle \ln K_2 \rangle$ with changes in K_2 is relatively slow, as shown in Figure 5 which presents values of $\langle \ln K_2 \rangle$ for typical values of $\sigma_1 = 1, 3, \text{ and } 5 \text{ cm}^{-2}$, in the most useful range of σ_1 , as a function of $\log_{10} K_2$. It can be seen that the minima are relatively broad and that there is a moderate latitude in the values of K_2 since $\langle \ln K_2 \rangle$ is no more than twice its minimum value for about a 100-fold range in K_2 . A similar graph for the equivalent monomer-tetramer system shows

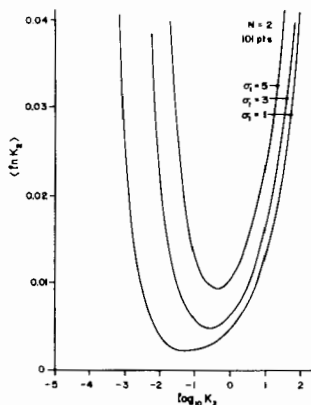


Fig. 5. Estimated minimal fractional uncertainties in the apparent dimerization constants for systems analyzed as consisting only of monomer and dimer presented as a function of $\ln K_2$. Procedure and restrictions are the same as in Figure 4, except for values of K_2 .

slightly greater latitude in K_4 , with about a 200-fold range K_4 for a doubling of the minimum uncertainty in K_4 .

The uncertainties in the estimated association constants show almost the expected proportionality to the reciprocal of the square root of N , the number of points: over the range from 31 to 301 equi-spaced data points the values of $\langle \ln K_2 \rangle$ are proportional to $N^{-0.48}$.

These simple estimations appear to be usefully applicable to estimates of the uncertainties of the association constants from NONLIN. Table I presents comparisons of the average estimated values of K_2 obtained from NONLIN, keeping the value of σ_1 fixed at the appropriate value, for various σ_1 . The data to be

TABLE I

Comparison of observed variations in $\ln \hat{K}_2$ for different noise perturbations with the predicted variations. All calculations are for monomer-dimer systems with $K_2 = 1$.

σ_1 cm ⁻²	$\frac{dc}{d\xi}$ max	\hat{K}/K_A	$\langle \ln K_2 \rangle_{\text{obs}}$	$\langle \ln K_2 \rangle_{\text{pred}}$
0.5	1.38	0.996	0.041	0.031
1	4.7	1.005	0.016	0.010
1.5	15	1.008 ₈	0.003 ₆	0.004 ₁
3	10	0.999 ₁	0.007 ₂	0.008 ₄
3	2.5	0.996	0.037	0.030
5	8.3	0.991	0.015	0.017
5	15	1.005	0.012	0.011
10	12	1.002	0.049	0.036

fitted was calculated for the indicated ideal dimerization. These "perfect" data were then modulated by adding to the data a set of Gaussian random numbers, with mean of zero and standard deviation of 3 μm , so as to simulate typical experimental error. This "noisy" simulated data was then fitted by NONLIN, and the procedure repeated five to ten times for each case shown. The mean of the values of the estimated values of $\ln K_2$ is compared to the true value of $\ln K_2$ ($=\ln K_A$) in the third column of the table and the standard deviation of these values from their mean is presented in the fourth column. The corresponding estimations of $\langle \ln K_2 \rangle$ were then carried out for the number of data points and for the maximum gradient present and measurable in the simulated solution column (any points with $\frac{dc}{d\xi} > 15$ mm fringe displacement/cm² were discarded and not used in the fitting procedures.) These "predicted" estimates in column 5 compare reasonably with the observed uncertainties in $\ln K_2$. Similar comparisons for other monomer-oligomer systems were likewise successful in predicting the observed uncertainties. Accordingly we use these simple estimations of the uncertainties in $\langle \ln K_A \rangle$ for examining sensitivity to association heterogeneity.

The expressions already developed for the effect of heterogeneity on apparent association constants, eqs.(10), are only valid for the ultracentrifuge under conditions where there is no sensible redistribution of solute, i.e. for small values of $\sigma_1 \Delta \xi$, as for equilibrium in short columns at low speeds or for experiments at very low speeds. Other expressions are required for more general conditions. These may be obtained as follows: starting with the formal statement of mass conservation for each component within a sectoral solution column, and invoking the definition of $\sigma_n(r)$, the apparent number average molecular weight of that component we write

$$c_k \circ \Delta \xi = \int_a^b c_j \circ d\xi = \int_a^b c_j(\xi) d\xi = \frac{c_j(\xi)}{\sigma_{n,i}(\xi)} \quad b \quad a \quad (15a)$$

substituting for the c_j and $\sigma_{n,i}$ we obtain

$$c_j \circ \Delta \xi = M_1 \left[\frac{m_{j,i} + nM_{j,n}}{\sigma_1 \left(\frac{m_{j,1} + nM_{j,n}}{m_{j,1} + m_{j,n}} \right)} - \frac{\lambda m_{j,1} + n\lambda^n m_{j,n}}{\sigma_1 \left(\frac{\lambda m_{j,1} + n\lambda^n m_{j,n}}{\lambda m_{j,1} + \lambda^n m_{j,n}} \right)} \right] \quad (15b)$$

$$= M_1 / \sigma_1 \{ m_{j,1} (1 + x_{j,n}) - \lambda m_{j,1} (1 + x_{j,n} \lambda^{n-1}) \}$$

where M_1 is the molecular weight of the monomer, $m_{j,i}$ is the molarity of the i -mer of the j -th component at the base of the solution channel, λ is $\exp(-\sigma_1 \Delta \xi)$, the ratio of the monomer concentration at the meniscus to the monomer concentration

at the base of the solution column, and $x_{j,n}$ is $m_{j,n}/m_{j,1}$, the ratio of the n-mer to the monomer molarity at the base for the j-th component. Substituting the parameter β for $(1-\lambda^n)/(1-\lambda)$ this becomes

$$c_j^{\circ} \Delta \xi = \frac{m_{j,1} M_1 (1-\nu)}{\sigma_1} (1 + \beta x_{j,n}) \quad (15c)$$

With the substitution $\gamma_j = \sigma_1 \Delta \xi c_j^{\circ} / [M_1 (1-\lambda)]$ we can express the monomer molarity of component j as

$$m_{j,1} = \gamma_j / (1 + \beta x_{j,n}) \quad (16)$$

and this expression, in turn, can be used to obtain the value of the apparent association constant for a heterogeneous system with components A and B as

$$\hat{K} = \frac{m_{T,n}}{m_{T,1}} = \frac{m_{A,n} + m_{B,n}}{(m_{A,1} + m_{B,1})^n} = \frac{K_A m_{A,1}^n + K_B m_{B,1}^n}{(m_{A,1} + m_{B,1})^n} = \frac{K_A + K_B (m_{B,1}/m_{A,1})^n}{(1 + m_{B,1}/m_{A,1})^n} \quad (17)$$

Dividing by K_A and introducing eq.(16) for the $m_{j,1}$ we obtain

$$\hat{K}/K_A = \frac{1 + (K_B/K_A)(\gamma_B/\gamma_A)^n [(1 + \beta x_{A,n})/(1 + \beta x_{B,n})]^n}{[1 + (\gamma_B/\gamma_A)(1 + \beta x_{A,n})/(1 + \beta x_{B,n})]^n} \quad (18a)$$

When $\theta = \gamma_B/\gamma_A = c_B^{\circ}/c_A^{\circ}$ is sufficiently small, this expression can be approximated by

$$\hat{K}/K_A = 1/(1 + \theta < 1 + \beta x_{A,n} >)^n \quad (18b)$$

or

$$\ln (\hat{K}/K_A) = -n \ln (1 + \theta < 1 + \beta x_{A,n} >) \approx -n \theta (1 + \beta x_{A,n}) \quad (19a)$$

with the relation on the right valid in the limit as $\theta \rightarrow 0$. When $n\sigma_1 \Delta \xi \ll 1$, i.e. for sufficiently low speed and/or short enough solution columns, β approaches n and $x_{A,n}$ approaches $x_{A,n}^{\circ}$ the molar ratio of n-mer to monomer in the original solution and

$$\ln (\hat{K}/K_A) = -n \ln (1 + \theta < 1 + n x_{A,n}^{\circ} >) \approx -n \theta (1 + n x_{A,n}^{\circ}), \quad (19b)$$

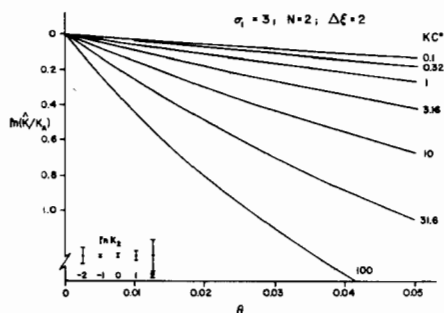
expressions identical to eqs.(10). On the other hand, for high speed equilibrium

conditions with $\sigma_1 \Delta\xi \gg 1$, the value of β approaches unity and eq.(19a) becomes

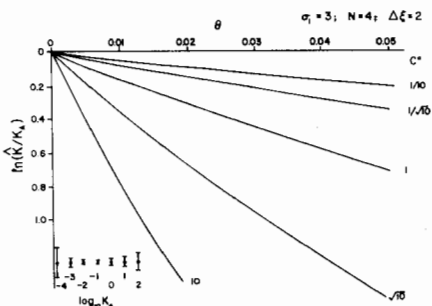
$$\ln \left(\widehat{K}/K_A \right) = -n \ln (1 + \theta < 1 + x_{A,n} >) \approx -n \theta (1 + x_{A,n}) \quad (19c)$$

Expressions (19) provide simple direct estimates of $\ln \widehat{K}/K_A$ for heterogeneous systems with sufficiently small amounts of heterogeneity. More generally eq.(17) may always be used if the values of $m_{A,1}$ and $m_{B,1}$ are known. The required values of $m_{A,1}$ and $m_{B,1}$ may be estimated by successive approximations for any given system. Figure 6a presents values of $\ln \left(\widehat{K}/K_A \right)$ for monomer-dimer systems with $K_B=0$ as a function of θ , the relative amount of the contaminating non-associating component B. In this particular figure the value of σ_1 was taken as 3 cm^{-2} , a value corresponding to intermediate speeds, as recommended by Teller et al¹ for studies of self-associating systems, and the value of $\Delta\xi = 2 \text{ cm}^2$, approximately the value corresponding to a 3 mm high solution column near the center of the ultracentrifuge cell. The several curves correspond to different values of $K_A c^0$ and therefore various loading concentrations. The ranges indicated on the lower left hand corner of the figure correspond to ranges of four times the expected standard deviation (± 2 standard deviations) in $\ln \widehat{K}$. The different ranges correspond to different values of K_2 (over a 10^4 range) for the associating component, as indicated by the label $\log_{10} K_2$. From this figure it appears possible, at least in favorable cases, to detect considerably less than 1% heterogeneity using experiments with a 3- or 10-fold range of initial concentrations.

a



b

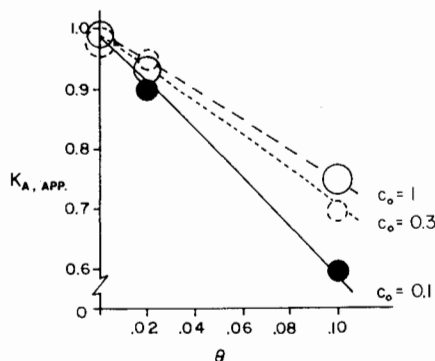


Figs. 6. Values of apparent dimerization (panel a) and apparent tetramerization (panel b) constant \widehat{K} , presented in the form $\ln \left(\widehat{K}/K_A \right)$, as a function of θ for a system with $K_A = 1$, $K_B = 0$, $\sigma = 3 \text{ cm}^{-2}$ and $\Delta\xi = 2 \text{ cm}^2$. The several curves shown correspond to different initial (loading) concentrations. The ranges indicated on the lower left hand corner are 4 times the predicted optimal error in $\ln \widehat{K}$ (± 2 standard deviations) for data sets with 101 equispaced points and $3 \mu\text{m}$ rms noise.

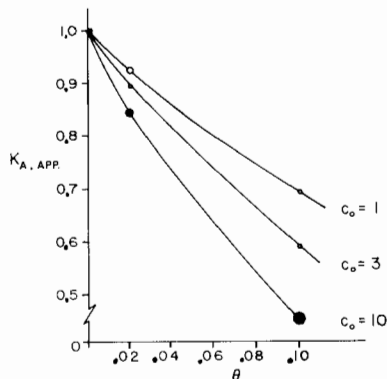
Figure 6b presents the same type of information for the equivalent monomer-tetramer system. The sensitivity to heterogeneity appears to be significantly greater than for the monomer-dimer case, over a considerably larger range in K_4 (a range of about 10^6).

The sensitivity to heterogeneity has been tested using calculated concentration distributions for several heterogeneous systems and the program NONLIN to analyze these distributions after they had been modulated by the addition of sets of Gaussian random numbers of known variance so as to simulate typical random errors. Five to ten sets of such noisy simulated data were analyzed for each distribution. The uncertainties of the derived parameters were taken as the standard deviations of the derived parameters from the mean of their values. Two such examples are given in Figures 7 for monomer-dimer equilibria, systems for which it is most difficult to detect association heterogeneity. Values of the K are presented as a function of θ for three loading concentrations and for high speed conditions, $\sigma_1 = 5 \text{ cm}^{-2}$, in Fig (7a) and for moderate speeds, $\sigma_1 = 1.5 \text{ cm}^{-2}$, in Fig. (7b). The radii of the circles correspond to the standard deviation of the values of \hat{K} . From Figure (7a) we can infer that a 2% contamination by "incompetent monomer" would barely be detectable under these high speed conditions. Comparison with Figure (7b) shows that considerably more sensitivity in detecting heterogeneity would be expected at moderate speeds, largely because of the increased precision in obtaining values of \hat{K} .

a.



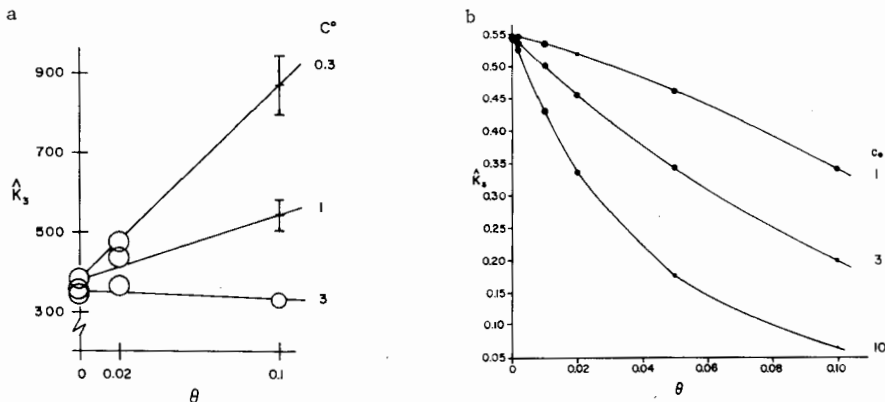
b.



Figs. 7. Illustration of the detection of heterogeneity with simulated data. Concentration distributions were calculated at 100 equispaced points from $r = 6.4$ to 6.7 cm for $K_2 = 1$ and the indicated values of c_o . All data points corresponding to concentration gradients, $dc/d\xi$, greater than $15 \text{ mm fringe displacement/cm}^2$ were discarded and the remaining data modulated by random numbers with a variance of $3 \mu\text{m}$ to simulate random experimental error. Five to ten of these "noisy" simulated data sets were individually fitted by NONLIN to the appropriate model. The mean and standard deviation from the mean of the estimated values of K_2 are presented as the points and their indicated ranges as a function of θ . Panel (a) corresponds to $\sigma_1 = 5 \text{ cm}^{-2}$ and panel (b) to $\sigma_1 = 1.5 \text{ cm}^{-2}$.

The procedure of comparing association constants for equilibrium experiments at different initial loading concentrations as outlined so far depends on the model used to obtain an apparent association constant from the experimental observations. For example, if a homogeneous monomer-dimer system were analyzed as a monomer-trimer system, then the apparent association constant that would be obtained with such an incorrect model would depend strongly on the region of the data analyzed, with the apparent trimerization constant decreasing with higher observation concentrations. Thus even a homogeneous system would give rise to different apparent constants with different loading concentrations unless the concentration ranges of the data used to estimate the apparent association constants for the different c° were maintained exactly the same for all estimations. Such truncation of the data sets to be used to a common concentration range insures that the errors from the choice of an incorrect model affect all the estimations of K for the different loading concentrations equally. (Strictly speaking, the distribution of data points for all the experiments at the different loading concentrations should be identical, but this requirement may be relaxed if the point density is sufficiently high.)

This procedure of using a common concentration span for determining the K is strongly urged since one rarely can be absolutely sure that the model of the associating system that is being used to analyze the data is completely correct. The effectiveness of this procedure for analyzing a system for heterogeneity even with an incorrect model, provided only that the data used span a common observation concentration range, is illustrated in Figures 8. The first of these



Figs. 8. Analyses of simulated self-associating systems for heterogeneity using incorrect models: a) monomer-dimer system with $K_2 = 1$ and $\sigma_1 = 2 \text{ cm}^{-2}$ analyzed as a monomer-trimer system; and b) monomer-dimer-tetramer system with $K_2 = K_4 = 1$ and $\sigma_1 = 2 \text{ cm}^{-2}$ analyzed as a monomer-trimer system. In both figures five to ten "noise" perturbed (rms = 3 μm fringe displacement) data sets with about 100 equispaced data points for which $dc/d\xi < 15 \text{ mm fringe displacement/cm}^2$ were analyzed as described in the text.

(Fig. 8a) presents the results of an analysis of a heterogeneous monomer-dimer system with $\sigma_1 = 2 \text{ cm}^{-2}$ in terms of a trimerization reaction -- clearly an incorrect choice of model. The procedure used for "mis-analyzing" this system was to first truncate the data for all three initial concentrations so that this data spanned a common concentration range. Then this truncated data was fit by NONLIN, using the data from all three channels jointly, to determine the "best value" of σ_1 for a monomer-trimer fit of the complete trimmed data set. Each channel of data was then analyzed separately using NONLIN, as a monomer-trimer system with this specific communal "best value" of σ_1 and the resultant K_3 noted. This procedure was repeated for all the initial concentrations, in each case using five different sets of Gaussian random numbers to mimic the effect of random noise ($3 \mu\text{m}$) on the calculated distributions. Figure 8a presents the values of the apparent trimerization constants thus obtained, with their standard deviation from the mean indicated by the dimensions of the points, as a function of θ , the extent of heterogeneity. Clearly one can distinguish the effects of heterogeneity for $\theta \geq \sim 0.02$, although the sensitivity appeared distinctly less than the nearly comparable case illustrated in Figure 7b for a correctly analyzed monomer-dimer system.

Figure 8b presents a similar analysis for a monomer-dimer-tetramer system with $\sigma_1 = 2 \text{ cm}^{-2}$ again analyzed, as a monomer-trimer system. In this case one can clearly detect the presence of less than 2×10^{-3} of a non-associating monomer even though an incorrect model was used for the analysis. Part of the reason for the increased sensitivity in this case is the much better fitting of the monomer-dimer-tetramer data by a trimerization as compared to the monomer-dimer data of Figure 8a: in both cases the perturbation by random numbers to simulate experimental error corresponded to an rms of $3 \mu\text{ms}$; for the monomer-dimer data the rms of the fit was about $7 \mu\text{m}$, while for the monomer-dimer-tetramer data the rms of the fit was about $3 \mu\text{m}$, indicating a nearly perfect description of the data by the assumed trimerization.

An application of this procedure to experimental data is illustrated by some of our work with bovine insulin. The sample of insulin used had been chromatographically purified^{22,23}, first on SP-sephadex in 10M acetic acid and then on QAE-Sephadex at pH 9 in 50% dimethylformamide. This insulin preparation appeared to have less than one part per thousand present of any contaminant, including desamido-insulins, as determined by disc-gel electrophoresis and chromatography. Figure 9 presents the apparent z-average molecular weights, $M_z(r)$, observed for two loading concentrations of this insulin at two angular velocities as a function of the normalized concentration gradient $\frac{dc}{d\xi}(r)/\omega^2$. The solvent was 0.1 M NaCl, 0.1 M TRIS and 0.001 M EDTA pH 7 at 25°C, conditions under which insulin shows extensive association²⁴. This preparation clearly appears to be heterogeneous in association behavior, as indicated by the divergence

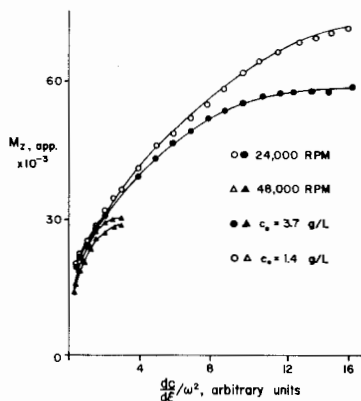


Fig. 9. Apparent z-average molecular weight presented as a function of the normalized concentration gradient for insulin at pH 7 in 0.1 M NaCl, 0.1 M TRIS and 0.001 M EDTA at 25°C.

of the several curves. However, at very low concentrations, where the equilibria appear to involve primarily monomer, dimer and tetramer²³, the apparent molecular weights are sensibly independent of loading concentration or angular velocity. Table II presents values of the apparent dimerization constants determined for insulin from two replicate interferograms under these conditions. These values were obtained by truncating the experimental observations to fringe displacements less than 1 mm in a 30 mm thick centerpiece corresponding to concentrations less than 6.9×10^{-5} M (in monomer units). This truncation spans most of the range where the apparent molecular weights appear to be independent of both concentration

TABLE II

Test of homogeneity of initial association stages of insulin. Experimental conditions: pH 7.0, 0.1 M TRIS, 0.1 M NaCl, 10^{-3} M EDTA, 25°C, 48,000 rpm in 30 mm thick centerpiece.

c° g L ⁻¹	Plate G305		Plate G308	
	RMS of fit μm	\hat{K}_2^* Lg ⁻¹	RMS of fit μm	\hat{K}_2^* Lg ⁻¹
0.34	3.0	310	2.8	306
1.01	2.5	277	2.7	291
2.44	4.2	313	4.6	306

* Fit as a monomer-dimer-tetramer system for \hat{K}_2 , $\xi(c_a)$, $c(a)$, keeping $\hat{K}_4 = 6.36 \times 10^3$ L³g⁻³, over the concentration range corresponding to 0-1 mm fringe displacement (0- 6.9×10^{-5} M in monomer equivalent molarity).

and speed and also avoids regions where there is a significant participation by oligomers greater than the tetramer. The model used in fitting the experimental observations was that of a monomer-dimer-tetramer equilibrium with the value of the tetramerization constant fixed at a value consistent with our analyses of this system²³. The lack of significant variation in \hat{K}_2 with loading concentration that is shown in Table II suggests that this insulin sample is indeed homogeneous, within experimental error, in its association properties for this initial region where only the smallest oligomers participate. The heterogeneity at higher concentrations is thus ascribable to the participation of higher degree of polymerization.

It would be useful to obtain quantitative estimates of the heterogeneity actually present in such experiments. As shown in Figure 10, even as little as 1% heterogeneity can generate significant divergence between curves of the apparent molecular weight averages of different loading concentrations if $n = 4$. This divergence between such curves increases strongly for larger n . Clearly it would be useful to have procedures for making quantitative estimates of the extent of heterogeneity from curves of apparent molecular weight averages vs concentration, or from values of \hat{K} , obtained with different c° . Accordingly we present here, as a first step towards such analyses, some considerations applicable to single ideal systems that can be described by a single value of n .

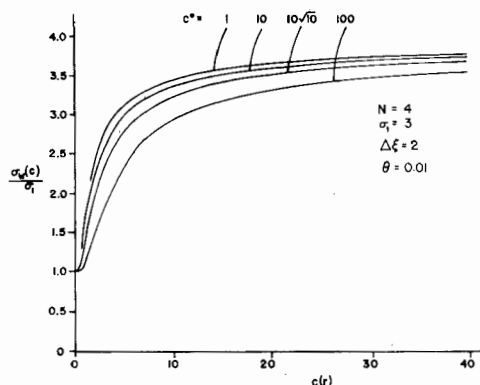


Fig. 10. Apparent reduced weight average molecular weight as a function of observation concentration for several loading concentrations, as indicated. These curves were calculated for a monomer-tetramer system. The effects of the small extent of heterogeneity ($\theta = 0.01$) are clearly discernible, especially at the higher loading concentrations.

First we examine the sensitivity to heterogeneity of the procedure of examining curves of weight average molecular weight, $M_w(r)$, vs $c(r)$ for different loading concentrations. Indications of heterogeneity are provided by the differences between different curves at the same $c(r)$ but for different values of c° . Accordingly we wish to evaluate the derivative of the weight average

molecular weight with respect to initial concentration, c° , keeping association constant, heterogeneity and observation concentration constant. It is convenient to express the derivative in logarithmic form and to split the desired derivative into two parts according to the chain rule

$$\begin{aligned} \left(\frac{\partial \ln M_w(r)}{\partial \ln c^\circ} \right)_{K_A, \theta, c(r)} &= \left(\frac{\partial \ln M_w(r)}{\partial \ln \hat{K}} \right)_{K_A, \theta, c(r)} \left(\frac{\partial \ln \hat{K}}{\partial \ln c^\circ} \right)_{K_A, \theta, c(r)} \\ &= \left(\frac{\partial \ln M_w}{\partial \ln \hat{K}} \right)_c \left(\frac{\partial \ln \hat{K}}{\partial \ln c^\circ} \right)_{K_A, \theta} \end{aligned} \quad (20)$$

The partial derivative of $\ln M_w(r)$ with respect to $\ln \hat{K}$, at constant K_A , θ , and $c(r)$, is identical to the derivative

$$\left(\frac{\partial \ln M_w}{\partial \ln \hat{K}} \right)_c = \left(\frac{\partial \ln M_w}{\partial \ln \hat{K}} \right)_{m_1} + \left(\frac{\partial \ln M_w}{\partial \ln m_1} \right)_{\hat{K}} \left(\frac{\partial \ln m_1}{\partial \ln \hat{K}} \right)_c \quad (21)$$

where for any ideal association equilibrium the total concentration, c , is given in terms of the monomer concentration, m_1 , by

$$c = M_1 (m_1 + n \hat{K} m_1^n) \quad (22)$$

and where the weight average molecular weight, M_w , is given by

$$M_w = M_1 (m_1 + n^2 \hat{K} m_1^n) / (m_1 + n \hat{K} m_1^n) \quad (23)$$

Here we consider m_n and m_1 to be the total molar concentration of n -mer and monomer present. The derivatives appearing in eq.(21) may be evaluated from relations (22) and (23) to obtain

$$\begin{aligned} \left(\frac{\partial \ln M_w}{\partial \ln \hat{K}} \right)_c &= \frac{n^2 \hat{K} m_1^{n-1}}{1 + n^2 \hat{K} m_1^{n-1}} - \frac{\hat{K} m_1^{n-1}}{1 + \hat{K} m_1^{n-1}} - \left(\frac{n^2 (n-1) \hat{K} m_1^{n-1}}{1 + \hat{K} m_1^{n-1}} - \frac{n(n-1) \hat{K} m_1^{n-1}}{1 + \hat{K} m_1^{n-1}} \right) \frac{\hat{K} m_1^{n-1}}{1 + n^2 \hat{K} m_1^{n-1}} \\ &= \frac{n(n-1) \hat{x}(r)}{\langle 1 + n^2 \hat{x}(r) \rangle^2} \end{aligned} \quad (24)$$

where, as before, we have substituted $\hat{x}(r) = K_{T,1}^{n-1} = m_{T,n}/m_{T,1}$ for the apparent molar ratio of n -mer to total monomer at the point r . The derivative represented by eq. (24) has the maximum value of $(n-1)/4n$ at $\hat{x}(r) = 1/n^2$. The other

derivative needed in eq. (20) may be obtained by noting that the apparent association constant, \hat{K} , is a function of K_A , θ and c° , and that the total initial loading concentration c° can be considered to be a function of K_A , θ and $m_{A,1}$, the molarity of the monomer at the base of the solution column. Thus we may write

$$\left(\frac{\partial \ln \hat{K}}{\partial \ln c^\circ} \right)_{K_A, \theta} = \left(\frac{\partial \ln \hat{K}}{\partial \ln m_{A,1}} \right)_{K_A, \theta} \left(\frac{\partial \ln m_{A,1}}{\partial \ln c^\circ} \right)_{K_A, \theta} \quad (25)$$

The derivatives on the right may be evaluated by noting that $x_{A,n}$, the molar ratio of n-mer to monomer for component A at the base of the solution column, is equal to $K_A m_{A,1}^{n-1}$. Combining this relation with eq. (15c) we obtain

$$\left(\frac{\partial \ln m_{A,1}}{\partial \ln c^\circ} \right)_{K_A, \theta} = \frac{m_{A,1} + \beta K_A m_{A,1}^n}{(m_{A,1} + n\beta K_A m_{A,1}^n)} = \frac{\sigma_1}{\sigma_n(b)} \quad (26)$$

where we have substituted the number average reduced molecular weight at the base of the cell for its explicit definition. Similarly we obtain from eq. (19a) the other derivative needed for eq. (25) as

$$\left(\frac{\partial \ln \hat{K}}{\partial \ln m_{A,1}} \right)_{K_A, \theta} = - \frac{\theta \beta n (n-1) x(b)}{1 + \theta <1 + \beta x(b)>} \quad (27)$$

where we write $x(b)$ for $x_{A,n}$, the molar ratio of n-mer to monomer for component A at the base of the solution column, changing notation slightly so as to distinguish this ratio from the x of eq. (24) which refers to the molar ratio of n-mer to monomer corresponding to the observation concentration, $c(r)$. Combining eqs. (20) and (24-27) we obtain the desired result:

$$\left(\frac{\partial \ln M_w(r)}{\partial \ln c^\circ} \right)_{K_A, \theta, c(r)} = \frac{n^2 (n-1)^2 \theta \beta \hat{x}(r) x(b) \sigma_1}{<1+n \hat{x}(r)>^2 (1+\theta <1+\beta x(b)>) \sigma_n(b)} \quad (28a)$$

Generally we are interested in the detection and quantitation of small extents of heterogeneity. Accordingly we consider the limit

$$\lim_{\theta \rightarrow 0} \left(\frac{\partial \ln M_w(r)}{\partial \ln c^\circ} \right)_{c(r)} = - \frac{\theta n^2 (n-1)^2 \beta \hat{x}(r) x(b) \sigma_1}{<1+n \hat{x}(r)>^2 \sigma_n(b)} \quad (28b)$$

This expression for small θ may be further simplified for high speed equilibrium experiments when $\sigma_n(b)/\sigma_1 \approx n$ to

$$\left(\frac{\partial \ln M_w(r)}{\partial \ln c^\circ} \right)_{c(r)} \approx - \frac{\theta n(n-1) \hat{x}(r) x(b)}{\langle 1+n \hat{x}(r) \rangle^2} \quad (28c)$$

At the point where $\hat{x}(r) = 1/n^2$ the value of $M_n(r)$ is $M_1(n+1)/2$ and eq. (24) has a maximum; thus eqs. (28a,b,c) will also have maxima at this point. For this point of maximum sensitivity we then obtain the simple relation

$$\left(\frac{\partial \ln M_w(r)}{\partial \ln c^\circ} \right)_{c(r)} \bigg|_{M_n = M_1(n+1)/2} = -(n-1)^2 \theta x(b) / 4n \quad (28d)$$

valid for high speed equilibrium ($\sigma_1 \Delta \xi \gg 1$) with $x(b) \gg 1$, and for small extents of heterogeneity.

Combining eqs. (25-27) we obtain the variation of the apparent association constant for a simple heterogeneous system with loading concentration as

$$\left(\frac{\partial \ln \hat{K}}{\partial \ln c^\circ} \right)_{\theta, K_A} = - \frac{\theta \beta n(n-1) x(b) \sigma_1}{(1+\theta \langle 1+\beta x(b) \rangle) \sigma_n(b)} \quad (29a)$$

As before, this complete expression may be simplified for specific cases of interest, e.g. for sufficiently small extents of heterogeneity as:

$$\left(\frac{\partial \ln \hat{K}}{\partial \ln c^\circ} \right)_{\theta, K_n} = -\theta \beta n(n-1) x(b) \sigma_1 / \sigma_n(b) \quad (29b)$$

and for high speed equilibrium we obtain the simple relation

$$\left(\frac{\partial \ln \hat{K}}{\partial \ln c^\circ} \right)_{\theta, K_a} = -(n-1) \theta x(b) \quad (29c)$$

valid when $x(b)$, $\sigma_1 \Delta \xi \gg 1$ and $\theta \ll 1$.

Expressions (28) and (29) all contain the term $x(b)$. A useful approximation for $x(b)$ in terms of the initial concentration, c° , may be obtained when $\theta \ll 1$ so that $x(b) \approx \hat{x}(b)$, and when $x(b)$ is large, as is often the case of high speed equilibrium experiments. Under these conditions $c(a) \ll c(b)$, $c_1(b) \ll c_n(b)$, $\sigma_n(b) \approx n\sigma_1$, and eq. (15), as written for the main component, becomes applicable to all of the solution:

$$c^\circ \Delta \xi \frac{c(b)}{\sigma_n(b)} \approx \frac{c_1(b) + c_n(b)}{n\sigma_1} \approx \frac{c_n(b)}{n\sigma_1} = \frac{\hat{K}_c c_1^n(b)}{n\sigma_1} \quad (30)$$

where \hat{K}_c ($=\hat{K}/M_1^{n-1}$) is the concentration (gL^{-1}) scale based apparent association constant, and c , c° , $c_n(b)$ and $c_1(b)$ refer to all components jointly. Equation (30) is then solved for $c_1(b)$, which, in turn, is used to obtain the desired approximation as

$$\hat{x}(b) = \frac{m_{T,n}(b)}{m_{T,1}(b)} \approx \frac{\hat{K}_c c_1^n(b)}{nc_1(b)} = \left(\frac{\hat{K}_c}{n}\right)^{\frac{1}{n}} (c^\circ \sigma_1 \Delta \xi)^{\frac{n-1}{n}} \quad (31)$$

which is valid provided that both $x(b)$ and $\sigma_1 \Delta \xi \gg 1$. This approximation may be inserted into the appropriate eqs. (28) and (29) to obtain simple limiting relations in terms of the readily accessible variables n , σ_1 , $\Delta \xi$, K_c and c°

$$\left(\frac{\partial \ln M_w(r)}{\partial \ln c^\circ} \right)_{c(r)} \bigg|_{M_n = M_1(n+1)/2} = -\theta \langle n-1 \rangle^2 / 4n \left(\hat{K}_c c^\circ{}^{n-1} / n \right)^{\frac{1}{n}} (\sigma_1 \Delta \xi)^{\frac{n-1}{n}} \quad (28D)$$

and

$$\left(\frac{\partial \ln K}{\partial \ln c^\circ} \right) = -\theta(n-1) \left(\hat{K}_c c^\circ{}^{n-1} / n \right)^{\frac{1}{n}} (\sigma_1 \Delta \xi)^{\frac{n-1}{n}} \quad (29C)$$

that are useful for estimation of small values of θ from high speed sedimentation equilibrium experiments. Eqs. (28a) and (29a) are valid for large values of θ , say $\theta > 0.1$, however, these expressions may be difficult to use since values of $x(b) = x_{A,n}$ can differ significantly from the apparent $\hat{x}(b)$ estimated from the \hat{K} when θ cannot be assumed to be small. In such cases we note that for $K_B \approx 0$

$$\hat{x}(b) = \frac{x_{T,n}}{x_{T,1}} = \frac{x_{A,n} m_{A,1}}{m_{A,1} + m_{B,1}} = \frac{x_{A,n}}{1 + \theta(1 + \beta x_{A,n})} \quad (32)$$

Solving eq. (32) for $x_{A,n}$ and inserting the result in eqs. (28a) and (29a) we obtain

$$\left(\frac{\partial \ln M_w(r)}{\partial \ln c^\circ} \right)_{c(r)} = - \frac{n^2 (n-1)^2 \theta \beta \hat{x}(r) \hat{x}(b)}{\langle 1+n \hat{x}(r) \rangle^2} \quad (28A)$$

and

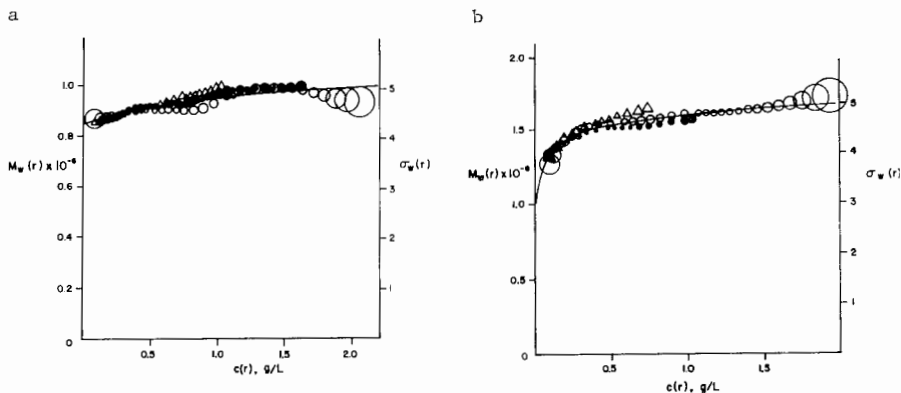
$$\left(\frac{\partial \ln K}{\partial \ln c^\circ} \right) = -\theta \beta n(n-1) \hat{x}(b) \frac{\sigma_1}{\sigma_n(b)} \quad (29A)$$

which is exactly the same expression as eq. (29b) except that $x(b)$ has been replaced by its apparent value, $\hat{x}(b)$, the apparent molar ratio of n -mer to total monomer at the base of the solution column. The approximation of eq.(31) is still valid for high speed equilibrium and for $\hat{x}(b) \gg 1$, and may be used here also to obtain the needed estimates of $x(b)$.

The data of Table II may be used in conjunction with eqs.(29) to obtain quantitative estimates of the heterogeneity of the initial association of our insulin preparation. This data leads to a least squares estimate of the derivative $(\partial \ln \hat{K} / \partial \ln c^\circ) = (+0.1 \pm 3.9) \times 10^{-2}$. In experiments at 48,000 rpm the value of $\sigma_1 \Delta \xi = 3.05$, making the use of eq.(29c) marginal. Accordingly we calculated points for a graph of $\hat{K}(b)$ vs c° by assuming several values of c_1 and then combining eqs.(30) and (31). From this curve we estimated the value of $\hat{K}(b)$ for the geometrical mean $c^\circ (=0.94 \text{ g L}^{-1})$ and the geometrical mean $\hat{K} = 300 \text{ L g}^{-1}$ as $\hat{K}(b) = 20.5$; the corresponding value of $\sigma_1 / \sigma_n(b) = 0.512$. Eq.(29b) then leads to the estimate $\theta = 0.000 \pm 0.002$. [Equation (29c), which is strictly applicable only for both high speed equilibrium and extensive association, yields the same estimate.]

Unfortunately the heterogeneity for insulin evident in figure 9 is more difficult to evaluate properly since the higher stages of associating depicted there are as yet poorly characterized and since theory for evaluating heterogeneity in systems with several sequential equilibria is not available. However, the maximum value of the derivative $(\partial \ln M_w / \partial \ln c^\circ)_{dc/d\xi}$, analogous to the derivatives of eqs.(28) is 0.2-0.3. When n is large eq.(28d) - or its analog that may be derived for the z-average molecular weight - suggests that this value might arise from heterogeneity of a few percent or so, provided $x(b)$ were sufficiently large. Clearly more work is needed before any definite conclusions can be reached about the heterogeneity apparent with these larger oligomers.

We illustrate the estimation of heterogeneity with two further examples, taken from our work with Limulus polyphemus hemocyanin²⁰. Figure 11a presents the apparent weight average molecular weights observed as a function of $c(r)$ for three loading concentrations at pH 5, in acetate buffer, 0.05 ionic strength with 0.45 M KCl added at 25.8°C and at 6800 rpm. Under these conditions, Limulus hemocyanin exhibits a slightly nonideal reversible equilibrium between the 24S dodecamer and a 36S 24-mer. These curves are systematically displaced towards lower values with increasing loading concentration indicating heterogeneity. Separate fits of the initial regions, for fringe displacements up to 1 mm, to a monomer-dimer equilibrium gave the following apparent association constants: $\hat{K}_2 = 0.213 \text{ L g}^{-1}$ for $c^\circ = 0.2 \text{ g L}^{-1}$; 0.189 L g^{-1} for 0.4 g L^{-1} ; and 0.119 L g^{-1} for 2 g L^{-1} . The typical rms error for these fits corresponded to about 3 μm fringe displacement. From these data we estimate the derivative $(\partial \ln \hat{K} / \partial \ln c^\circ)$ as $-0.3 (+0.1, -0.2)$. The value of σ_1 is 4.32 cm^{-2} and $\Delta \xi = 1.965 \text{ cm}^2$. We take the geometric mean (corresponding to the arithmetic average of the ΔF°) of the values of \hat{K}_2 as 0.168 L g^{-1} and of the loading concentration as $\bar{c}^\circ = 0.68_3$, giving an effective value of $\hat{K}c^\circ = 0.115$. Equation (31) gives poor estimates



Figs. 11. Apparent weight average molecular weights, $M_w(r)$, (and values of $\sigma_w(r)$) as a function of $c(r)$ observed for Limulus polyphemus hemocyanin by Johnson and Yphantis²⁰. Panel a) presents the data obtained at pH 5, 0.5 M total ionic strength at 25.8°C and 6800 rpm. The data can be represented by a slightly nonideal dimerization of the 24S dodecamer to the 36S molecular species. Panel b) displays the observations at pH 6.1, 26°C and 5200 rpm. Here the hemocyanin is more strongly associated with both dimers (36S) and tetramers (60S) apparently participating in the equilibria. In both cases the triangles correspond to initial concentrations, c° of 0.2 g L⁻¹, the filled circles to $c^\circ = 0.8$ g L⁻¹ and the open circles to $c^\circ = 2$ g L⁻¹.

of $\hat{x}(b)$ since the extent of association at the base is relatively small. Accordingly we again estimated values of $\hat{x}(b)$ (=0.265) and of $\sigma_N(b)/\sigma_1$ (=1.21) by assuming several values of c_1 and thus generating curves of $\hat{x}(b)$ and of $\sigma_N(b)$ vs c° , Equation (29A) then yields the estimate of θ as 0.70 (from 0.47 to 1.16).

Clearly this estimated value of θ is not small and thus eq.(29b) is not strictly applicable. The appropriate equation for large θ is eq.(29A) which provides the even larger estimate of $\theta = 2.3$ (from 0.9 to no defined upper limit).

Figure 11b presents $M_w(r)$ as a function of $c(r)$ for the same three loading concentrations as before (0.2, 0.8 and 2 g L⁻¹), but at pH 6.1, with 0.05 ionic strength phosphate buffer and 0.45 M KCl added. The temperature was 26°C and the ultracentrifuge speed 5200 rpm. Under these conditions hemocyanin is associated somewhat more strongly than at pH 5 with the apparent participation of both a dimer and a tetramer of the 24S dodecamer. The curves of Figure 11b diverge more distinctly than do those of Figures 11a, clearly demonstrating heterogeneity. Evaluation of the apparent dimerization constants from fringe displacements truncated to 0.7 mm for the three channels separately gave values of \hat{K}_2 of 92.6 L g⁻¹ at $c^\circ = 0.2$ g L⁻¹; 42.8 L g⁻¹; at 0.8 g L⁻¹, and 23.8 L g⁻¹ at 2 g L⁻¹. The geometrical average value of \hat{K}_2 was 45.5 L g⁻¹, the corresponding c° was again 0.683 g L⁻¹, $\sigma_1 = 2.5$ cm⁻² and $\Delta\xi = 1.965$ cm². Here the association

is sufficiently extensive and the value of $\sigma_1 \Delta \xi$ (=4.93) sufficiently large that eq.(29c) appears to be quite valid. The value of $(\partial \ln \hat{K} / \partial \ln c^\circ) = -0.59$ (-0.49 to -0.76) leads to the estimate $\theta = 0.068$ (0.056 to 0.087). At first glance it is surprising that the much more evident divergence of the curves in Figure 11b compared to Figure 11a, and the significantly larger spread of the values of \hat{K}_2 at pH 6 compared to pH 5, correspond to significantly smaller estimates of heterogeneity. Examination of eqs.(28) and (29) quickly convinces one of the crucial role played by the extent of association in the superficial appearances of heterogeneity, with the divergences both in the point average molecular weight curves and in the values of \hat{K}_2 being essentially proportional to $x(b)$.

DISCUSSION

It should be pointed out that in all three experiments examined here the data were severely truncated to minimize the effects of higher association stages or nonideality. In general, considerably wider ranges of data are available for determining the K and thus the precision and sensitivity of estimating heterogeneity should be greater than in the cases presented. Furthermore, the cases analyzed have all been effectively monomer-dimer systems which are least sensitive to the effects of association heterogeneity.

Conditions conducive to maximum sensitivity include high centrifugation speed and long solution columns, to maximize the extent of association; choice of systems with large n ; and the use of a large range of c° for increased precision in determining the derivatives of eqs.(28) and (29). The sensitivity to heterogeneity can be multiplied by utilizing separation procedures based on association behavior, such as gel exclusion chromatography under associating conditions with comparisons of early and retarded fractions. Conversely the effects of heterogeneity may be minimized by utilizing small values of $\sigma_1 \Delta \xi$, thus effectively approaching determinations with no sensible differential redistributions of solutes, and the choice of conditions so as to keep Kc_0^{n-1} small. Again separation procedures can be used to isolate relatively homogeneous fractions on the basis of their association behavior.

In closing it must be emphasized that the relations derived depend completely on the assumption that the hetero-association of different components to form mixed oligomers can be neglected. Generally that is not so, and the estimates provided here must therefore be considered as minimal estimates of association heterogeneity. Similarly, heterogeneity can be underestimated in cases where there exists a considerable amount of a strongly self-associating minor component. Indeed it may be shown that if θ and K are both large enough the derivatives of eqs.(28) and (29) go through zero and become positive. Occasionally such cases have been seen experimentally.

SUMMARY

Some types of heterogeneity in self-associating systems can be detected with high sensitivity. Expressions to quantitate the effects of such heterogeneity show that the usual subjective impressions obtained from the lack of coincidence of the $M_K(r)$ vs c (c) for different c^0 can be highly misleading since sensitivities depend greatly on both the extent and degree, n , of association present.

ACKNOWLEDGEMENTS

This work has been supported by grants from the National Institutes of Health (AM-18001) and from the National Science Foundation (PCM 76-21847). We thank Raymond Kikas for his technical assistance.

REFERENCES

1. Teller, D. C., Horbett, T. A., Richards, E. G. and Schachman, H. K. (1969) *Ann. N.Y. Acad. Sciences*, 164, 66-101.
2. Squire, P. G. and Li, C. H. (1961) *J. Am. Chem. Soc.*, 83, 3521-3528.
3. Jeffrey, P. D. and Coates, J. H. (1966) *Biochemistry*, 5, 489-498.
4. Roark, D. E. and Yphantis, D. A. (1969) *Ann. N.Y. Acad. Sciences*, 164, 245-278.
5. Williams, R. C., Jr., Yphantis, D. A. and Craig, L. C. (1972) *Biochemistry*, 11, 70-77.
6. Ansevin, A. T., Roark, D. E. and Yphantis, D. A. (1970) *Analytical Biochem.*, 34, 237-261.
7. Paul, G. H. and Yphantis, D. A. (1972) *Analytical Biochem.*, 48, 588-604.
8. Paul, C. H. and Yphantis, D. A. (1972) *Analytical Biochem.*, 48, 605-612.
9. Laue, T. M., Domanik, R. A., Rhodes, D. and Yphantis, D. A. (1977) *Biophys. J.*, 17, 101a (Abstract), and in preparation.
10. DeRosier, D. J., Munk, P. and Cox, D. J. (1972) *Analytical Biochem.*, 50, 139-153.
11. Domanik, R. A., Yphantis, D. A. and Lue, T. M. (1977) *Biophys. J.*, 17, 101a (Abstract) and in preparation.
12. Yphantis, D. A., Domanik, R. A. and Laue, T. M. (1978) In preparation.
13. Szuchet, S. and Yphantis, D. A. (1976) *Arch. Biochem. Biophys.*, 173, 495-516.
14. Correia, J. J. and Yphantis, D. A. (1978) In preparation.
15. Davis, P. and Rabinowitz, P. (1954) *J. Assoc. Computing Machinery*, 1, 183-191.
16. Johnson, M. L. (1973) Ph.D. Thesis, University of Connecticut.
17. Johnson, M. L. and Yphantis, D. A. (1978) In preparation.
18. Szuchet, S. and Yphantis, D. A. (1973) *Biochemistry*, 12, 5115-5127.
19. Szuchet, S. (1976) *Arch. Biochem. Biophys.*, 177, 437-460.
20. Johnson, M. L. and Yphantis, D. A. (1978) *Biochemistry*, 17, 1448-1455.
21. Yphantis, D. A. (1964) *Biochemistry* 3, 297-317.
22. Wu, G.-M. (1974) Ph.D. Thesis, State University of New York at Buffalo.
23. Wu, G.-M. and Yphantis, D. A. (1978) In preparation.
24. Pekar, A. H. and Frank, B. H. (1972) *Biochemistry*, 11, 4013-4016.