Analysis of Sedimentation Equilibrium Distributions Reflecting Nonideal Macromolecular Associations

Peter R. Wills*, Michael P. Jacobsen†, and Donald J. Winzor†
*Department of Physics, University of Auckland, Auckland, New Zealand; and †Center for Protein Structure, Function and Engineering, Department of Biochemistry, University of Queensland, Brisbane, Queensland 4072, Australia

ABSTRACT A rigorous statistical–mechanical approach is adopted to derive general quantitative expressions that allow for the effects of thermodynamic nonideality in equilibrium measurements reflecting interaction between dissimilar macromolecular reactants. An analytical procedure based on these expressions is then formulated for obtaining global estimates of equilibrium constants and the corresponding reference thermodynamic activities of the free reactants in each of several sedimentation equilibrium experiments. The method is demonstrated by application to results from an ultracentrifugal study of an electrostatic interaction between ovalbumin and cytochrome *c* (Winzor, D. J., M. P. Jacobsen, and P. R. Wills. 1998. *Biochemistry.* 37:2226–2233). It is demonstrated that reliable estimates of relevant thermodynamic parameters are extracted from the data through statistical analysis by means of a simple nonlinear fitting procedure.

INTRODUCTION

As the culmination of studies designed to provide a rigorous theoretical basis for the characterization of macromolecular interactions by direct analysis of sedimentation equilibrium distributions (Wills et al., 1996; Jacobsen et al., 1996; Wills et al., 1997; Winzor et al., 1998, 1999), this investigation presents a comprehensive analysis of the effects of thermodynamic nonideality in multicomponent systems of arbitrary chemical complexity. The theory provides a foundation for the global analysis of sedimentation equilibrium data from systems involving interactions between dissimilar reactants. The validity of the theory is in no way dependent on the inevitable shortcomings that arise in its application to experimental data. Rather, it serves as a guide to the parameters which can be meaningfully extracted. We demonstrate how sedimentation equilibrium results for mixtures of ovalbumin and cytochrome *c* under conditions of neutral complex formation (Winzor et al., 1998) can be subjected to an appropriate method of global data analysis to provide an estimate of the thermodynamic equilibrium constant for the heterogeneous association reaction.

THEORY

In this section, we develop rigorous expressions for the relationships between thermodynamic activities and concentrations in multicomponent systems—a more extensive set of expressions than those presented recently (Winzor et al., 1999). Although the analysis of sedimentation equilibrium distributions is the direct objective of these developments, the quantitative expressions are pertinent to results obtained by using a wide range of physicochemical procedures. Previous studies have relied, at least implicitly, upon the calculation of activity coefficients of different chemical components as a means of allowing for effects of thermodynamic nonideality, but the present procedure obviates that requirement. Instead, we formulate the relationship between concentrations and thermodynamic activities in ways that are specifically applicable to the problems of evaluating equilibrium constants and the statistical–mechanical interpretation of thermodynamically defined osmotic virial coefficients (McMillan and Mayer, 1945). This approach overcomes the problem usually encountered in such studies: that the primary equations used in the analysis are transcendental functions in the main observational variables, the concentrations of various species.

General quantitative expressions

The starting point for analysis of the thermodynamics of a macromolecular solution amenable to statistical–mechanical interpretation is the standard virial expansion for osmotic pressure (Π) in terms of the molar concentrations (C) of the various solute components $i, j, k, \ldots,$

$$\frac{\Pi}{RT} = \sum_{\mathrm{i}} C_{\mathrm{i}} + \sum_{\{\mathrm{i\,j}\}} B_{\mathrm{ij}} C_{\mathrm{i}} C_{\mathrm{j}} + \sum_{\{\mathrm{i\,j\,k}\}} B_{\mathrm{ijk}} C_{\mathrm{i}} C_{\mathrm{j}} C_{\mathrm{k}} + \cdots, \qquad (1)$$

where the sets $\{i\,j\,k\ldots\}$ comprise all combinations, rather than permutations, of the relevant indices. Eq. 1 is equivalent to Eq. 89 of Hill and Chen (1973) with different notation for virial coefficients: in present terminology, each of the subscripts $i, j, k, \ldots$ ranges over the set of macromolecular solutes $\{A, B, C, D, \ldots\}$ present in the system. In terms of solute thermodynamic activities (z_{i}) defined under conditions of constant temperature (T) and chemical poten-

Received for publication 16 March 2000 and in final form 26 June 2000.
Address reprint requests to Donald J. Winzor, University of Queensland, Center for Protein Structure, Function and Engineering, Dept. of Biochemistry, Brisbane, QLD 4072, Australia. Tel.: +61-7-3365-2132; Fax: +61-7-3365-4699; E-mail: winzor@biosci.uq.edu.au.
This paper is dedicated to the memory of Mike Jacobsen, who died during the preparation of the manuscript.

0006-3495/00/10/2178/10 $2.00

tial of solvent (μ_s), namely,

$$\mu_i(T, \mu_s, \{C_j\}) = \mu_i^0(T, \mu_s) + RT \ln z_i(\{C_j\}), \quad (2)$$

the osmotic pressure is given by the alternative expression

$$\frac{\Pi}{RT} = \sum_i z_i + \sum_{\{i\,j\}} b_{ij} z_i z_j + \sum_{\{i\,j\,k\}} b_{ijk} z_i z_j z_k + \cdots, \quad (3)$$

where the sets $\{i\,j\,k\ldots\}$ once again span all combinations rather than permutations of individual indices.

By application of Eqs. 90 and 91 of Hill and Chen (1973), it can be seen that the two sets of coefficients in these equations for Π are related by the expressions

$$B_{ii} = -b_{ii} \quad (4a)$$

$$B_{ij} = -b_{ij}; \quad j \neq i \quad (4b)$$

$$B_{iii} = 4b_{ii}^2 - 2b_{iii} \quad (4c)$$

$$B_{iij} = 4b_{ij}b_{ii} + b_{ij}^2 - 2b_{iij}; \quad j \neq i \quad (4d)$$

$$B_{ijk} = 2[b_{ij}b_{jk} + b_{ik}b_{kj} + b_{ij}b_{ik} - b_{ijk}]; \quad j \neq i, \quad k \neq i, \quad k \neq j \quad (4e)$$

Note that B_{ij} and B_{ji} are equivalent designations of the same virial coefficient in Eq. 1, as are B_{iij}, B_{iji}, and B_{jii}, because the same combination of indices is merely being permuted in the different designations. Also required are the corresponding inverse relations for the coefficients, namely,

$$b_{ii} = -B_{ii} \quad (5a)$$

$$b_{ij} = -B_{ij}; \quad j \neq i \quad (5b)$$

$$2b_{iii} = 4B_{ii}^2 - B_{iii} \quad (5c)$$

$$2b_{iij} = 4B_{ij}B_{ii} + B_{ij}^2 - B_{iij}; \quad j \neq i \quad (5d)$$

$$2b_{ijk} = 2[B_{ij}B_{ik} + B_{ik}B_{jk} + B_{ij}B_{jk}] - B_{ijk}; \quad j \neq i, \quad k \neq i, \quad k \neq j \quad (5e)$$

These coefficients allow the specification of molar concentrations C_i in terms of thermodynamic activities z_i, or vice versa, in multicomponent systems. The former relations are obtained by applying the thermodynamic expression

$$C_i = z_i\left[\partial(\Pi/RT)/\partial z_i\right]_{T,\, z_{j\neq i}} = z_i + 2b_{ii}z_i^2 + \sum_{j\neq i} b_{ij}z_iz_j + 3b_{iii}z_i^3 + 2\sum_{j\neq i} b_{iij}z_i^2z_j + \sum_{j\neq i} b_{ijj}z_iz_j^2 + \sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} b_{ijk}z_iz_jz_k + \cdots. \quad (6)$$

The corresponding expressions of thermodynamic activities in terms of concentrations, obtained by series reversion and use of Eq. 4, are

$$z_i = C_i + 2B_{ii}C_i^2 + \sum_{j\neq i} B_{ij}C_iC_j + \frac{3}{2}[B_{iii} + 2B_{ii}^2]C_i^3 + \sum_{j\neq i}[B_{ij} + 2B_{ij}B_{ii}]C_i^2C_j + \frac{1}{2}\sum_{j\neq i}[B_{ij}^2 + B_{ijj}]C_iC_j^2 + \frac{1}{2}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}[2B_{ij}B_{ik} + B_{jk}(B_{ij} + B_{ik}) + B_{ijk}]C_iC_jC_k + \cdots. \quad (7)$$

These relationships lead to the following general expression for the activity coefficient, $\gamma_i = z_i/C_i$, of any component:

$$\ln \gamma_i = 2B_{ii}C_i + \sum_{j\neq i} B_{ij}C_j + \frac{3}{2}B_{iii}C_i^2 + \sum_{j\neq i} B_{iij}C_iC_j + \frac{1}{2}\sum_{j\neq i} B_{ijj}C_j^2 + \frac{1}{2}\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}} B_{ijk}C_jC_k + \cdots. \quad (8)$$

This expression is presented for completeness and to establish consistency with our previous treatment of first-order corrections for nonideality in the case of a single solute (Wills et al., 1996).

Allowance for chemical interaction

Eqs. 1–8 represent a complete thermodynamic description of a solution containing a specified number of separate macromolecular components (the range of any subscript, *i* or *j* or *k*, etc.). However, when there are very short-range contact forces between macromolecules that lead to the formation of relatively stable aggregates (homo- or hetero-dimers, homo- or hetero-trimers, etc.), it is customary to regard these aggregates as separately identifiable macromolecular species, in which case the number of nominal "components" is enlarged to incorporate these possibilities. It is more correct, strictly speaking, to regard dynamically reversible macromolecular aggregation as a special manifestation of thermodynamic nonideality (Hill and Chen, 1973; Wills et al., 1996)—an approach that allows elucidation of the association phenomenon from sedimentation equilibrium data in a simpler manner than that in which aggregates are regarded as additional species. However, aggregation is usually described in terms of the formation of molecular complexes: in a system composed of two dissimilar reactants A and B, we can, in the most general case, label the complexes C ≡ AB, D ≡ A_2, E ≡ B_2, F ≡ A_2B, G ≡ AB_2,

$H \equiv A_3$, $J \equiv B_3$, etc. The stoichiometric equilibrium constants (K_{AB}, K_{AA}, K_{BB}, K_{AAB}, K_{ABB}, K_{AAA}, K_{BBB}, etc.) are then expressed in terms of the activities z_A and z_B of the two thermodynamic components: $z_C = K_{AB}z_Az_B$, $z_D = K_{AA}z_A^2$, $z_E = K_{BB}z_B^2$, $z_F = K_{AAB}z_A^2\,z_B$, $z_G = K_{ABB}z_Az_B^2$, $z_H = K_{AAA}z_A^3$, $z_J = K_{BBB}z_B^3$, etc. The adaptation of Eq. 3 to the alternative Hill–Chen description of this system now follows.

The explicit expansion of Eq. 3 for a two-component system (A, B) involving potential aggregates of arbitrary composition ($C \equiv AB$, $D \equiv A_2$, $E \equiv B_2$, $F \equiv A_2B$, $G \equiv AB_2$, $H \equiv A_3$, $J \equiv B_3$, etc.) has the form

$$\begin{aligned}\Pi/RT = {} & z_A + z_B + z_C + z_D + z_E + b^*_{AB}z_Az_B + b^*_{AA}z_A^2 \\ & + b^*_{BB}z_B^2 + z_F + z_G + z_H + z_J + b^*_{AAB}z_A^2z_B \\ & + b^*_{ABB}z_Az_B^2 + b^*_{AAA}z_A^3 + b^*_{BBB}z_B^3 + b^*_{AC}z_Az_C \\ & + b^*_{BC}z_Bz_C + b^*_{AE}z_Az_E + b^*_{BD}z_Bz_D + b^*_{AD}z_Az_D \\ & + b^*_{BE}z_Bz_E + \cdots, \end{aligned} \tag{9}$$

where we have retained the order of terms in relation to aggregates (up to tri-molecular) of the components A and B; and where the asterisk on the coefficients indicates that certain components (the complexes, C, D, E, etc.) have been defined nonthermodynamically through the notional separation of the effects of associative and nonassociative intermolecular forces. It should be noted that this notional separation of forces can be determined quite arbitrarily from a statistical–mechanical point of view. However, in the case of strong, short-range associative forces that are usually encountered between macromolecules, the definition of an aggregate of any sort leaves little scope for ambiguity. Excluded volume forces and the relatively long-range forces due to the ionically screened electrostatic fields around molecules are generally regarded as nonassociative, whereas short-range dispersion forces and attractions between molecules due to specific surface group interactions are regarded as associative. One model for the quantitative analysis of this problem was developed by Wills and Georgalis (1981) in the context of the nonequilibrium thermodynamics of diffusion.

We must emphasize quite categorically that equilibrium constants for association reactions cannot be defined without the notional separation of associative and nonassociative forces between molecules because there is no thermodynamic procedure for distinguishing the configurations of a cluster of molecules that represent a complex of the molecules from those configurations in which they are regarded as being separate. This problem is implicit in all analyses of macromolecular associations, including self-association, but can generally be ignored because of the lack of ambiguity in making the notional separation of the prevalent molecular configurations into those that do or do not represent stable complexes.

Expression of the activities of all species in terms of z_A and z_B allows conversion of Eq. 9 to the form,

$$\begin{aligned}\Pi/RT = {} & z_A + z_B + (b^*_{AA} + K_{AA})z_A^2 + (b^*_{BB} + K_{BB})z_B^2 \\ & + (b^*_{AB} + K_{AB})z_Az_B \\ & + (b^*_{AAA} + b^*_{AD}K_{AA} + K_{AAA})z_A^3 \\ & + (b^*_{BBB} + b^*_{BE}K_{BB} + K_{BBB})z_B^3 + (b^*_{AAB} + b^*_{AC}K_{AB} \\ & + b^*_{BD}K_{AA} + K_{AAB})z_A^2z_B + (b^*_{ABB} + b^*_{BC}K_{AB} \\ & + b^*_{AE}K_{BB} + K_{ABB})z_Az_B^2 + \cdots. \end{aligned} \tag{10}$$

This relationship now needs to be compared with the corresponding expression for the two-component system with rigorously defined thermodynamic coefficients (b_{ij}, b_{ijk}, etc.). In present terminology, Eq. 3 becomes

$$\begin{aligned}\Pi/RT = {} & z_A + z_B + b_{AA}z_A^2 + b_{BB}z_B^2 + b_{AB}z_Az_B + b_{AAA}z_A^3 \\ & + b_{BBB}z_B^3 + b_{AAB}z_A^2z_B + b_{ABB}z_Az_B^2 + \cdots. \end{aligned} \tag{11}$$

By equating coefficients in Eqs. 10 and 11 with the aid of Eq. 5 we obtain

$$b_{AB} = b^*_{AB} + K_{AB} = K_{AB} - B^*_{AB} \tag{12a}$$

$$b_{AA} = b^*_{AA} + K_{AA} = K_{AA} - B^*_{AA} \tag{12b}$$

$$b_{BB} = b^*_{BB} + K_{BB} = K_{BB} - B^*_{BB} \tag{12c}$$

$$\begin{aligned} b_{AAB} = {} & b^*_{AAB} + b^*_{AC}K_{AB} + b^*_{BD}K_{AA} + K_{AAB} = 2B^*_{AB}B^*_{AA} \\ & + \frac{1}{2}(B^{*2}_{AB} - B^*_{AAB}) - B^*_{AC}K_{AB} - B^*_{BD}K_{AA} + K_{AAB} \end{aligned} \tag{12d}$$

$$\begin{aligned} b_{ABB} = {} & b^*_{ABB} + b^*_{BC}K_{AB} + b^*_{AE}K_{BB} + K_{ABB} \\ = {} & 2B^*_{AB}B^*_{BB} + \frac{1}{2}(B^{*2}_{AB} - B^*_{ABB}) - B^*_{BC}K_{AB} \\ & - B^*_{AE}K_{BB} + K_{ABB} \end{aligned} \tag{12e}$$

$$\begin{aligned} b_{AAA} = {} & b^*_{AAA} + b^*_{AD}K_{AA} + K_{AAA} \\ = {} & \frac{1}{2}(B^{*2}_{AA} - B^*_{AAA}) - B^*_{AD}K_{AA} + K_{AAA} \end{aligned} \tag{12f}$$

$$\begin{aligned} b_{BBB} = {} & b^*_{BBB} + b^*_{BE}K_{BB} + K_{BBB} \\ = {} & \frac{1}{2}(B^{*2}_{BB} - B^*_{BBB}) - B^*_{BE}K_{BB} + K_{BBB}. \end{aligned} \tag{12g}$$

Experimental estimation of these thermodynamic parameters (b_{ij}, b_{ijk}, etc.) thus has the potential to allow determination of the stoichiometric equilibrium constants K_{AB}, K_{AAB}, K_{ABB}, etc., for complex formation between A and B provided that magnitudes can be assigned to the notional

virial coefficients (b^*_{ij}, b^*_{ijk}, etc., or B^*_{ij}, B^*_{ijk}, etc.) describing the effects of nonassociative forces within clusters and the self-association constants (K_{ii}, K_{iii}, etc.). Should the latter equilibrium constants be nonzero, they can be determined in independent experiments involving only A or B.

It should be noted that the theory developed so far has quite general applicability and can be used in the analysis of experimental data gathered by using any equilibrium technique such as sedimentation equilibrium, static light scattering, osmotic pressure, isopiestic measurements, and so on. In what follows, we consider the special case of sedimentation equilibrium of systems in which the only significant reaction is binary heterogeneous association, A + B $\rightleftarrows$ C.

Adaptation to sedimentation equilibrium

At sedimentation equilibrium, the thermodynamic activity z_i of any component i, as defined in Eqs. 3 and 7, is given as a function of r, the distance from the center of rotation, by the relationship (Haschemeyer and Bowers, 1970; Milthorpe et al., 1975),

$$z_i(r) = z_i(r_F)\psi_i(r), \tag{13a}$$

where r_F is an arbitrary reference position, usually selected to be within the centrifuge cell, and

$$\psi_i(r) = \exp[M_i\phi_i(r^2 - r_F^2)] \tag{13b}$$

$$\phi_i = (1 - \bar{\nu}_i\rho_s)\omega^2/(2RT) \tag{13c}$$

M_i and $\bar{\nu}_i$ denote the solute molecular weight and partial specific volume respectively, ρ_s is the solvent density (Wills and Winzor, 1992; Wills et al., 1996), and ω is the angular velocity in an experiment conducted at absolute temperature T. The magnitude of $\psi_i(r)$ depends upon the selected reference radial position r_F, the solvent (through ρ_s), the experimental conditions (through T and ω), and the species to which it refers (through M_i and $\bar{\nu}_i$). For any given experiment, the ψ function for component j can be expressed in terms of that for component i by means of the relationship (Winzor et al., 1998),

$$\psi_j(r) = \psi_i(r)^{p_j}; \quad p_j = (M_j\phi_j)/(M_i\phi_i). \tag{14}$$

Consider a series of sedimentation equilibrium experiments in which the constitutive molar concentrations of different components, $\bar{C}_A$, $\bar{C}_B$, etc., have been determined as a function of radial distance. Combination of Eqs. 6 and 14 leads to the expressions

$$\bar{C}_i(r) = z_i(r_F)\psi_i(r) + 2b_{ii}z_i(r_F)^2\psi_i(r)^2 + \sum_{j\neq i} b_{ij}z_i(r_F)z_j(r_F)\psi_i(r)^{1+p_j} + \cdots \tag{15a}$$

$$\bar{C}_j(r) = z_j(r_F)\psi_i(r)^{p_j} + 2b_{jj}z_j(r_F)^2\psi_i(r)^{2p_j} + \sum_{k\neq j} b_{jk}z_j(r_F)z_k(r_F)\psi_i(r)^{p_j+p_k} + \cdots; \quad j \neq i. \tag{15b}$$

These equations represent a multicomponent generalization of the psi-function approach to the direct analysis of sedimentation equilibrium data for a single solute (Wills et al., 1996), that approach obviating the need for an iterative procedure to take account of thermodynamic nonideality and self-association based on transcendental equations in concentration variables (Johnson et al., 1981; Wills et al., 1981). With sufficiently precise data from a centrifuge run, it would be possible, in principle, to find $z_A(r_F)$, $z_B(r_F)$, etc., and any number of multinomial coefficients b_{AA}, b_{AB}, b_{BB}, etc., by nonlinear regression analysis of experimental [$\bar{C}_A(r)$, $\bar{C}_B(r)$, . . .] versus $\psi_A(r)$ or $\psi_B(r)$ records. However sedimentation equilibrium patterns are records of the total concentration, $\bar{C}_A(r) + \bar{C}_B(r) + \ldots$, or an absorbance representing some linear combination of component concentrations, as a function of radial distance. Such patterns are essentially a sum of weighted exponentials, deconvolution of which is a notoriously ill-posed problem.

There are no purely numerical procedures for obtaining best estimates of $z_A(r_F)$, $z_B(r_F)$, etc., that are independent of the model chosen for the thermodynamic interaction between the different components, because there will always be correlations between the best-fit values obtained for $z_A(r_F)$, $z_B(r_F)$, etc., and the various coefficients b_{AA}, b_{AB}, b_{BB}, etc., depending upon which terms in the expansion (Eq. 15) are used as the fitting function. Nevertheless, data from a number of different runs can be combined in global fitting to allow determination of the best-fit values of the thermodynamic parameters, b_{AA}, b_{AB}, b_{BB}, etc., across a wide range of concentrations of all solute species, together with values of the reference activities, $z_A(r_F)$, $z_B(r_F)$, etc., specific to each centrifuge run. The capability of modern analytical centrifuges to deconvolve by direct measurement the radial distribution patterns of some components from the total solute concentration profile gives the experimenter a significant advantage in this regard.

A practical strategy for simultaneous nonlinear regression analysis of multiple experiments is to hold the thermodynamic interaction coefficients, b_{ij}, b_{ijk}, etc., constant while best-fit values of the reference activities, $z_A(r_F)$, $z_B(r_F)$, etc., for the various components are found by minimizing the sum-of-squares residual for each experiment individually, and then to hold the reference activity parameters constant while the best-fit values of the thermodynamic coefficients are found by minimizing the "global" sum-of-squares residual by regression of the entire [$\bar{C}_A(r)$, $\bar{C}_B(r)$, . . .] versus $\psi(r)$ data from all of the experiments together. The iterative application of this two-step procedure will yield global best-fit values of all parameters for a chosen model of

macromolecular interactions (choice of terms in the expansion given by Eq. 15) relevant to the complete data set comprised of many centrifuge runs conducted with different speeds and different loading concentrations of the various solute components.

In relation to the global fitting step, the use of the $\psi(r)$ function serves the purpose of establishing an experimental scale for the estimation of thermodynamic activities, common to all experiments, which can be used effectively to fit concentration profiles to Eq. 6, in which the activities of different components are independent thermodynamic parameters. This may be understood as follows. We first establish a common point of reference for all runs, for heuristic reasons choosing the center of rotation, $r = 0$, and defining the standard psi-function as

$$\psi_i^0(r) = \exp[M_i\phi_i r^2] \tag{16a}$$

$$= \psi_i(r)\exp[M_i\phi_i r_F^2] \tag{16b}$$

We now define an experimental quasi-activity scale ζ_i for each component (cf. Eq. 13a). The value of ζ_i for any data point from any run is given by

$$\zeta_i = \zeta_i(0)\psi_i^0(r), \tag{17}$$

and $\zeta_i(0)$ represents the best available estimate of $z_i(0)$ for the data set from that experiment. $z_i(0)$ represents the notional thermodynamic activity that species i would exhibit at the center of rotation if the solution column were to extend all the way inward. Global fitting of the dependences of $\bar{C}_A(r)$, $\bar{C}_B(r)$, . . . upon $\psi_A^0(r)$ or $\psi_B^0(r)$ then amounts to the use of Eq. 6 with ζ_i replacing z_i to obtain best-fit values of the thermodynamic parameters b_{AA}, b_{AB}, b_{BB}, etc., which can be used for the improvement of the estimates $\zeta_A(0)$, $\zeta_B(0)$, etc. of the reference activities, $z_A(0)$, $z_B(0)$, etc., internal to each run and the subsequent recalculation of the ζ_i values for each data point in a further iteration of the procedure.

Eq. 6 is the basic thermodynamic relation that allows equilibrium constants to be obtained by analysis of sedimentation equilibrium patterns. In contrast, no equilibrium constant can be determined independently of the establishment of the scale for the thermodynamic activities of the components of the system. The simultaneous estimation of the reference activities, $z_A(0)$, $z_B(0)$, etc., for different experiments serves to normalize the experimental scales, $\zeta_A(r)$, $\zeta_B(r)$, etc., which represent the thermodynamic activities of the various species.

Finally, we consider a series of sedimentation equilibrium experiments in which neither reactant self associates and in which the constitutive molar concentrations, $\bar{C}_A$ and $\bar{C}_B$, of associating components A and B may both be determined as a function of radial distance. For the specific case in which the only associative clusters are 1:1 complexes between A and B, we now write extended forms of Eq. 15, a and b, with i = A and j = B as

$$\begin{aligned}\bar{C}_A(r) = {} & z_A(r_F)\psi_A(r) + 2b_{AA}z_A(r_F)^2\psi_A(r)^2 \\ & + b_{AB}z_A(r_F)z_B(r_F)\psi_A(r)^{1+p_B} + 3b_{AAA}z_A(r_F)^3\psi_A(r)^3 \\ & + 2b_{AAB}z_A(r_F)^2z_B(r_F)\psi_A(r)^{2+p_B} \\ & + b_{ABB}z_A(r_F)z_B(r_F)^2\psi_A(r)^{1+2p_B} + \ \ldots, \end{aligned} \tag{18a}$$

$$\begin{aligned}\bar{C}_B(r) = {} & z_B(r_F)\psi_A(r)^{p_B} + 2b_{BB}z_B(r_F)^2\psi_A(r)^{2p_B} \\ & + b_{AB}z_A(r_F)z_B(r_F)\psi_A(r)^{1+p_B} \\ & + 3b_{BBB}z_B(r_F)^3\psi_A(r)^{3p_B} \\ & + b_{AAB}z_A(r_F)^2z_B(r_F)\psi_A(r)^{2+p_B} \\ & + 2b_{ABB}z_A(r_F)z_B(r_F)^2\psi_A(r)^{1+2p_B} + \ \ldots, \end{aligned} \tag{18b}$$

in which expressions for the coefficients, b_{ij}, b_{ijk}, etc., are provided by Eq. 12. These equations are easily adapted to experimental variables that are linear combinations of $\bar{C}_A(r)$ and $\bar{C}_B(r)$. Practically speaking, elucidation of sedimentation equilibrium patterns for two-component systems requires the measurement of at least two experimental variables, A_I and A_{II}, (absorbances at specific wavelengths, λ_I and λ_{II}, or number of interference fringes) that depend on the cell path length, l = 1.2 cm, and the concentration of the solutes as a function of radial distance:

$$A_I(r) = [\varepsilon_{I,A}C_A(r) + \varepsilon_{I,B}C_B(r)]l, \tag{19a}$$

$$A_{II}(r) = [\varepsilon_{II,A}C_A(r) + \varepsilon_{II,B}C_B(r)]l. \tag{19b}$$

The quantities ε_I and ε_{II} are molar extinction coefficients in the case of experimental variables, $A_I(r)$ and $A_{II}(r)$, that are absorbances (at wavelengths, λ_I and λ_{II}), and they are molar refractive index increments when the variable is the number of fringes recorded in interference measurements of concentration.

RESULTS

We have used Eqs. 18 and 19 to analyze sedimentation equilibrium distributions recorded as absorbances at 280 and 410 nm in a recent experimental study of the 1:1 interaction between ovalbumin and cytochrome *c* (Winzor et al., 1998). The constituent concentrations of the two proteins were sufficiently small to decrease the effects of thermodynamic nonideality to such a level that only terms of second order in constituent activities needed to be retained in the use of Eq. 18 (as in Eq. 15). Osmotic second virial coefficients for nonassociative interactions between the species, ovalbumin (A), cytochrome *c* (B), were assigned magnitudes on the basis of spherical geometry. The

expressions used were (Wills and Winzor, 1992; Winzor and Wills, 1995)

$$2B^*_{ii} = \frac{32\pi N_A R_i^3}{3} + \left[\frac{Z_i^2}{2I}\right]\left[\frac{1 + 2\kappa R_i}{(1 + \kappa R_i)^2}\right], \qquad (20a)$$

$$B^*_{ij} = \frac{4\pi N_A[(R_i + R_j)^3]}{3} + \left[\frac{Z_i Z_j}{2I}\right] \cdot \left[\frac{1 + \kappa R_i + \kappa R_j}{(1 + \kappa R_i)(1 + \kappa R_j)}\right]; \quad j \neq i, \qquad (20b)$$

in which a molar ionic strength (I) of 0.03 was used to calculate the inverse screening length (κ) as $3.27 \times 10^7 \sqrt{I}$ (in cm^{-1} at 20°C). The radii (R_i) and net charges (Z_i) of species were as follows: $R_A = 2.92$ nm, $Z_A = -12$; $R_B = 1.90$ nm, $Z_B = +12$; $R_C = 3.20$ nm, $Z_C = 0$. N_A is Avogadro's number. Such calculations led to the following magnitudes for the various second virial coefficients: $2B^*_{AA} = 1968$ L/mol, $2B^*_{BB} = 1895$ L/mol, $B^*_{AB} = -1345$ L/mol.

Results of joint regression analysis, using the Levenberg–Marquardt method to minimize the total sum-of-squares residuals of the two absorbance distributions from a single sedimentation equilibrium run at 15,000 rpm and 20°C, followed by solution of Eq. 19 to yield concentration distributions, are presented in Fig. 1. The buoyant molecular weights, $M_i(1 - \bar{\nu}_i\rho_s)$, of reactants ovalbumin (A) and cytochrome c (B) are 11,340 mol^{-1} and 3,645 mol^{-1}, respectively, and the extinction coefficients used were

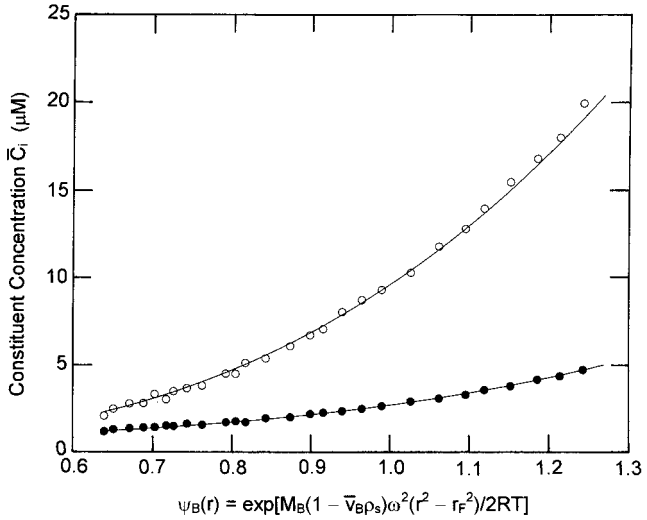

FIGURE 1 Illustration of the joint regression analysis of both constituent concentration distributions to evaluate the binding constant for the 1:1 interaction between ovalbumin (A) and cytochrome c (B) from a single sedimentation equilibrium experiment (No. 9 in Table 2): ●, $\bar{C}_A(r)$; ○, $\bar{C}_B(r)$. Values of $\psi_B(r)$ are based on a reference radial position (r_F) of 7.050 cm. Solid lines denote the best-fit descriptions of the combined data sets to Eq. 18, a and b, in terms of three fitted parameters: $z_A(r_F)$, $z_B(r_F)$, and K_{AB}. For purposes of clarity, only every fourth data point is shown.

$\varepsilon_{280,A} = 2.97 \times 10^4\ M^{-1}\ cm^{-1}$, $\varepsilon_{280,B} = 2.32 \times 10^4\ M^{-1}\ cm^{-1}$, $\varepsilon_{410,A} = 0$ and $\varepsilon_{410,B} = 1.06 \times 10^5\ M^{-1}\ cm^{-1}$. Figure 1 shows the extent of conformity between the experimental data for ovalbumin, $\bar{C}_A(r)$, and cytochrome c, $\bar{C}_B(r)$, and the best-fit description, namely, $z_A(r_F) = 8.625 \pm 0.016$ μM, $z_B(r_F) = 1.622 \pm 0.005$ μM, $K_{AB} = (7.69 - 7.84) \times 10^4\ M^{-1}$. The errors cited in $z_A(r_F)$ and $z_B(r_F)$ represent 95% confidence intervals based on the statistical standard errors in these parameters (derived from the diagonal elements of the correlation matrix), and the asymmetric range of values for K_{AB} about the best-fit value of $7.75 \times 10^4\ M^{-1}$ arises from the fact that the standard free energy of association between A and B, $\Delta G^0_{AB} = -RT \ln K_{AB}$, was used as the fitting parameter. This procedure is recommended by Johnson et al. (1981), who also note that use of the correlation matrix can lead to errors in fitted parameters being underestimated, depending on the data and the form of the fitted function. Significant nondiagonal elements in the correlation matrix (Table 1) show that the estimates of the parameters are not independent—a factor adding to concerns that the confidence intervals derived from the diagonal elements of the matrix may be misleadingly narrow.

To assess the adequacy of the simple estimates of the errors based on the correlation matrix (see Appendix), we used Monte Carlo simulation to generate 500 data sets with points chosen according to the error distributions of the experimental data. The sets of estimates of the parameters, $z_A(r_F)$, $z_B(r_F)$, and ΔG^0_{AB}, obtained by nonlinear fitting to these 500 data sets, are shown in Fig. 2. For purposes of visualization, points in the 3-dimensional parameter space (one point for each data set) have been projected onto planes formed by the axes representing two of the parameters. Also shown are bars representing 95% confidence intervals for the single parameters derived from the diagonal elements of the trivariate correlation matrix (Table 1) and ellipses representing 95% confidence intervals derived, using the standard value χ^2 (95%, 2) = 5.99, from the eigenvectors and eigenvalues of the inverse correlation matrix for the corresponding bivariate parameter distributions (see Appendix for details). It is striking that the confidence intervals derived from the correlation matrix give a good representation, both qualitatively and quantitatively, of statistical uncertainties and correlations in the estimates of the thermodynamic parameters obtained by nonlinear fitting to sedimentation equilibrium data using Eq. 15. This demonstrates that in-

TABLE 1 Correlation matrix for a three-parameter fit to data from a sedimentation equilibrium experiment ($r_F = 7.050$ cm)

Parameter	$z_A(r_F)$	$z_B(r_F)$	ΔG^0_{AB}
$z_A(r_F)(M)$	0.0709	0.0078	0.0612
$z_B(r_F)(M)$	0.0078	0.0065	0.0273
ΔG^0_{AB} (erg/mol)	0.0612	0.0273	0.1380

Entries scaled by a factor of 10^{-15} after ΔG^0_{AB} scaled by 10^{16}.

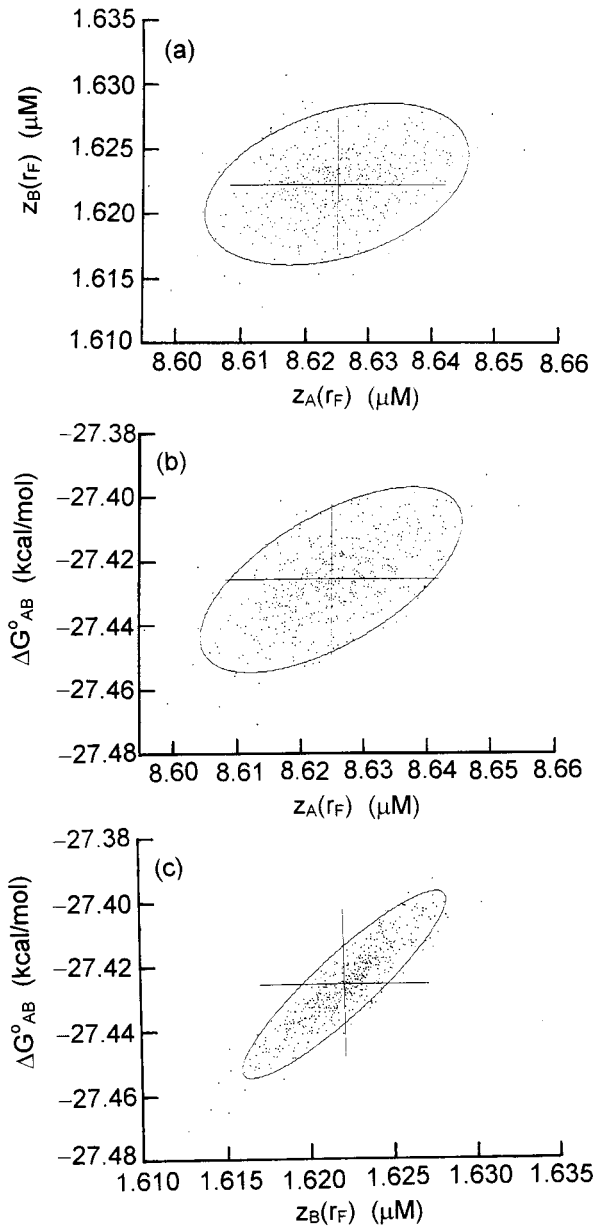

FIGURE 2 Results of statistical analyses of Monte Carlo simulated data representing a sedimentation equilibrium experiment (archetype: No. 9 in Table 2). The results [$z_A(r_F)$, $z_B(r_F)$, ΔG^0_{AB}] derived from each of 500 data sets have been projected into the planes formed by the axes of two of the three parameters. The expected 95% confidence intervals for individual parameters and bivariate parameter distributions are shown as bars and ellipses, respectively.

version of the curvature matrix used in nonlinear fitting does not give grossly misleading indications of statistical reliability in the estimates of the thermodynamic parameters that appear in Eq. 15, which is what we require to know. After all, uncertainties in parameters are estimated so that undue confidence is not placed in the values obtained by using the chosen fitting procedure—not for the sake of accurately measuring the statistical errors themselves. It would appear that the choice of terms from the equation, corresponding to the molecular interactions taken into account, and the justification for that choice, are of more relevance than statistical considerations to the assessment of uncertainties in the estimates of equilibrium constants.

A feature evident from Fig. 1 is the slight systematic departure of both experimental distributions from what would be their individual best-fit counterparts. However, if either experimental data set [absorbance at 410 nm as a measure of $\bar{C}_B(r)$ or absorbance at 280 nm as a scaled measure of the total protein concentration] is analyzed separately using the appropriately truncated form of Eq. 18 as a model of the thermodynamic interaction between ovalbumin and cytochrome *c*, then the only parameter estimates that make sense are the reference activities [$z_B(r_F) = 1.63$ (cf. 1.62) μM from analysis of the 410 nm data and $z_A(r_F) = 0.77$ (cf. 8.63) μM from analysis of the 280 nm data]; and the fitting cannot be relied on to have given any very useful information. This state of affairs demonstrates that a reasonable body of data, representative of the scale of interaction between two components over a broad range of measured concentrations of both components, is needed to allow a trustworthy estimate of the association constant K_{AB} and assessment of specific interactions between A and B.

We have conducted global analysis of data from 22 different centrifuge runs recorded at both 280 and 410 nm. We devised a fitting procedure that was robust vis-à-vis the initial estimates of parameters because of the difficulty in obtaining initial convergence in the fitting of all 45 parameters (the equilibrium constant and both reference activities from each of 22 experiments) to the experimental data. Our procedure was to fix a value of K_{AB} while individual best-fit values of $z_A(r_F)$ and $z_B(r_F)$ were found peculiar to each experiment; and then, with these reference activities fixed, to allow a new value of K_{AB} to be determined by minimization of the total sum-of-squares residuals for the entire data set from all experiments. These two steps were used iteratively in tandem until reasonable convergence was achieved. Finally, all 45 parameters were allowed to vary simultaneously to achieve the best global fit of the data. Inspection of the 45 × 45 correlation matrix revealed that the best fit was based on correlations between estimates of reference activities from different experiments that were as significant as those within individual experiments.

Representative data comprising the two constituent concentration distributions from seven experiments are illustrated in Fig. 3, *A* and *B*. Each constituent concentration is shown as a function of ζ_A and ζ_B, the experimentally derived thermodynamic activity scales spanning all of the data sets, according to the relationships defined in Eq. 6 via Eqs. 15–17. Fig. 4, *A* and *B* display the theoretical surfaces $(\bar{C}_A, \bar{C}_B) = f(z_A, z_B)$ to which the global data were fitted: the curves defined by parameter values describing global fitting are also shown for each of the seven representative runs.

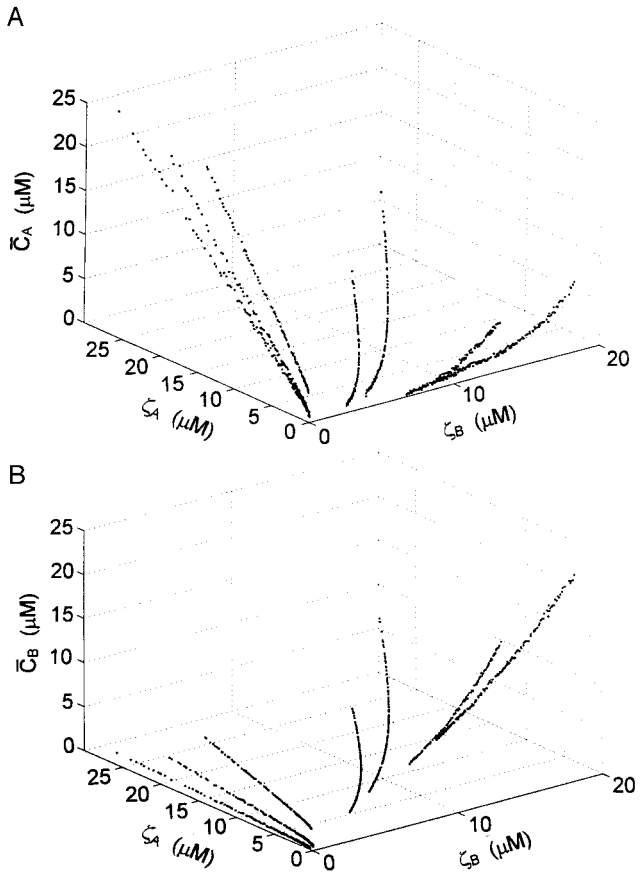

FIGURE 3 Representative data defining (*A*) $\bar{C}_A(\zeta_A, \zeta_B)$ and (*B*) $\bar{C}_B(\zeta_A, \zeta_B)$ surfaces from 7 of 22 sedimentation equilibrium experiments on mixtures of ovalbumin and cytochrome *c* (Experiments 1, 9, 10, 14, 15, 19, and 21 of Table 2) by global analysis of the two constituent concentrations (Eqs. 6, 15, and 17). Axes are labeled ζ_A and ζ_B because the positioning of the data points along these axes has been derived through the numerical fitting procedure (see text).

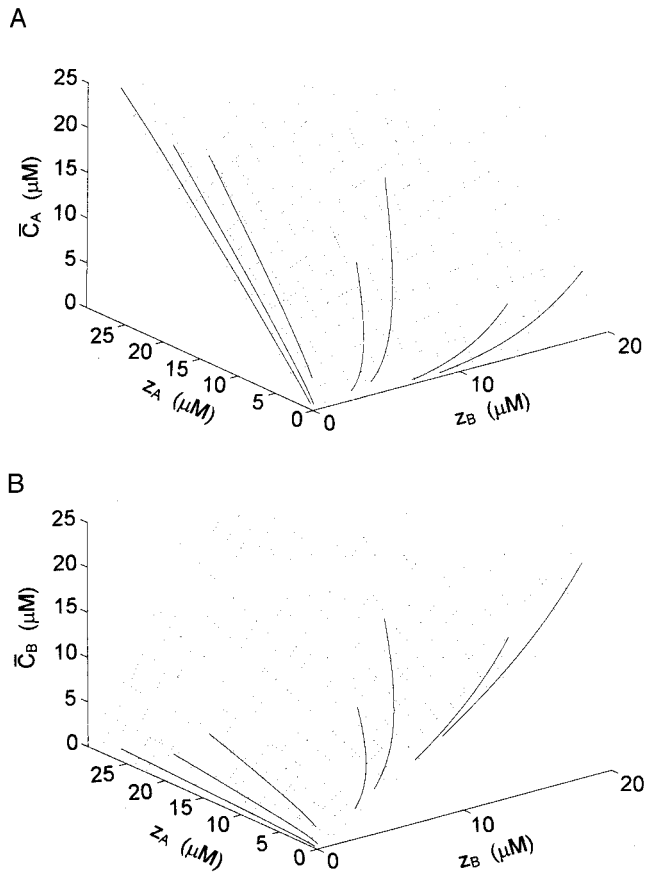

FIGURE 4 Representative best-fit descriptions of data from the seven distributions shown in Fig. 2 that have been obtained by global analysis of all 22 experiments listed in Table 2 in terms of Eq. 15 for a model with A, B, and C (the 1:1 complex AB) the only species present. The curves representing (*A*) $\bar{C}_A(r)$ and (*B*) $\bar{C}_B(r)$ for the seven experiments are calculated from a thermodynamic model with all parameters specified (surface represented as a grid) and the axes specify variation in the actual thermodynamic activities z_A and z_B.

Experimental details and the best-fit estimates of $z_A(r_F)$ and $z_B(r_F)$ for each of 22 sedimentation equilibrium runs are summarized in Table 2. The deviation of the parameters obtained for experiment No. 9 under global fitting conditions from the values obtained previously for this experiment are small, but larger than would be expected purely on the basis of statistical considerations, thereby testifying to effects such as the accommodation of differing systematic errors across separate experiments in the global fitting procedure.

A best-fit value of $8.39 \times 10^4\ M^{-1}$ was obtained for the equilibrium constant, K_{AB}, pertaining to the 1:1 interaction between ovalbumin and cytochrome *c* under the chosen experimental conditions (pH 6.3, I 0.03), with a narrow 95% confidence interval defined by the range $(8.36–8.42) \times 10^4\ M^{-1}$ as determined by the statistical standard error in the corresponding fitting parameter, ΔG^0_{AB}. However, there are other important sources of uncertainty in K_{AB} that render more precise determination of the standard error rather pointless: potential systematic errors in individual data sets, and possible inadequacy of the chosen model of macromolecular interaction, $A + B \rightleftarrows C$.

One obvious source of systematic error is the determination of the baseline for the absorbance, and hence concentration of solute species, in each experiment. It is our preferred practice to reach the best estimate of the baseline from all of the available information concerning the experimental set-up, rather than to make the baseline a fitting parameter of the analytical model. However, we acknowledge that baseline variation can be used better to gauge the uncertainties in estimates of the thermodynamic parameter of interest. In the end, the choice of the analytical model of the macromolecular interaction must be justifiable on grounds that take into account information beyond the data gathered in sedimentation equilibrium experiments, and the data-fitting procedure should not unduly influence the model.

A best estimate of K_{AB}, obtained by using Eq. 18 truncated at terms corresponding to bimolecular interactions,

TABLE 2 Global analysis of sedimentation equilibrium data for the interaction of ovalbumin (A) with cytochrome *c* (B), pH 6.3, I 0.03

Run	Speed (rpm)	$\bar{C}_A(r)$ (μM)*	$\bar{C}_B(r)$ (μM)*	$z_A(r_F)$ (μM)†	$z_B(r_F)$ (μM)†
1	15,000	1.1–17.9	4.6–18.2	2.862 (±0.007)	6.657 (±0.009)
2	15,000	1.1–14.9	4.7–15.7	2.428 (±0.006)	6.370 (±0.006)
3	15,000	0.5–15.3	4.4–16.2	2.535 (±0.004)	6.260 (±0.006)
4	15,000	1.8–5.7	8.7–17.8	1.887 (±0.008)	13.02 (±0.016)
5	15,000	0.0–6.8	8.1–18.5	1.734 (±0.006)	12.83 (±0.013)
6	15,000	1.6–11.3	8.6–20.7	1.992 (±0.006)	13.03 (±0.013)
7	15,000	2.3–18.1	1.2–4.3	8.891 (±0.013)	1.567 (±0.002)
8	15,000	2.2–13.1	1.3–3.4	9.076 (±0.010)	1.631 (±0.003)
9	15,000	2.1–21.2	1.2–5.1	8.592 (±0.013)	1.608 (±0.002)
10	20,000	0.4–22.1	0.2–2.4	6.581 (±0.012)	0.613 (±0.002)
11	20,000	0.1–16.4	0.3–2.0	7.135 (±0.007)	0.676 (±0.002)
12	20,000	0.5–14.0	0.3–1.7	6.668 (±0.013)	0.689 (±0.003)
13	20,000	0.2–19.6	0.1–0.9	7.322 (±0.012)	0.285 (±0.002)
14	20,000	1.5–25.2	0.1–1.0	9.672 (±0.012)	0.237 (±0.002)
15	15,000	0.5–11.5	2.9–10.5	1.967 (±0.002)	4.454 (±0.003)
16	15,000	0.1–9.2	2.9–9.0	1.773 (±0.006)	4.034 (±0.004)
17	15,000	0.5–9.4	3.1–9.5	1.714 (±0.006)	4.363 (±0.004)
18	20,000	0.0–5.0	4.5–17.3	0.569 (±0.002)	11.01 (±0.003)
19	15,000	0.1–4.1	8.8–20.8	1.082 (±0.006)	15.11 (±0.009)
20	15,000	0.8–6.5	8.8–21.8	0.473 (±0.005)	15.27 (±0.014)
21	15,000	0.1–4.2	6.8–16.6	1.087 (±0.006)	11.30 (±0.012)
22	15,000	0.0–11.9	7.0–14.9	0.953 (±0.008)	11.66 (±0.011)

From Winzor et al., 1998.

*Concentration ranges calculated from the absorbance (280 and 410 nm) data by solving Eq. 19.

†Thermodynamic activity of reactant at the reference radial position, taken as 7.050 cm for all experiments: numbers in parentheses denote twice the standard errors in the returned estimates.

could be biased due to the effects of trimolecular interactions that are accommodated statistically but whose significance cannot be determined from the data alone. On those grounds, it is to be expected that the uncertainty in our estimate of K_{AB} is likely to be somewhat larger than the value we have cited on the basis of statistical analysis. In that regard, it is notable that inclusion of a trimolecular term in the statistical analysis of our data gave rise variously to lack of convergence in the statistical fitting or physically meaningless results such as negative equilibrium constants. The main lesson to be taken is that, although Eq. 15 covers every possible case, generality is no substitute for an informed approach to the choice of model for statistical analysis.

DISCUSSION

An important theoretical outcome of this investigation is its provision of completely general quantitative expressions that include rigorous allowance for the effect of thermodynamic nonideality in the analysis of equilibrium measurements reflecting interactions between dissimilar reactants. In the practical context of a system that undergoes a reaction $A + B \rightleftarrows C$ our expressions have formed the basis of an analysis for obtaining a global estimate of the equilibrium constant and the corresponding reference thermodynamic activities of the two reactants pertaining to each of several sedimentation equilibrium experiments. Although this analysis is similar to existing procedures for the study of heterogeneous association by sedimentation equilibrium (Laue et al., 1993; Lewis et al., 1993; Kim et al., 1994; Bailey et al., 1996), it differs because of its ability to take separate account of effects arising from thermodynamic nonideality (nonassociative molecular forces) and "chemical reaction" (specific molecular complex formation). Our analysis highlights the fact that the results of thermodynamic measurements, such as sedimentation equilibrium measurements, make no distinction between molecules that may be said to have formed a complex and those just being close together—a situation necessitating independent consideration of the effects of the nonassociative forces.

There remains the problem of more generally investigating the practical limits of our approach to the determination of equilibrium constants in heterogeneous systems. These limits will inevitably be encountered in a case-by-case manner and will depend upon the precision, quantity, and range of data that can be obtained from sedimentation equilibrium experiments. The investigation of weak associations (that may well be of considerable biological importance) will require experiments to be conducted at relatively high reactant concentrations, in which case reliable information about nonassociative forces between species will need to be gathered so that the effects of those forces can be taken into account through use of Eq. 12. At these higher concentrations, the increasingly complex interactions between clusters of three or more molecules will become significant and therefore need to be considered (Hill, 1955). Fortunately, use of the present analysis couched in terms of the molar-based thermodynamic activity (Eq. 2) yields a buoyancy term that contains the solvent density (Eq. 13) and is therefore independent of solute concentration (Wills and Winzor, 1992). Further discussion of its relationship to the traditional analysis couched in terms of the molal-based activity and a concentration-dependent solution density in the buoyancy term has been presented recently (Wills et al., 2000).

Finally, we have demonstrated that the simplest statistical methods available for the evaluation of equilibrium constants for a simple $A + B \rightleftarrows C$ reaction can yield estimates with credible reliability when the appropriate thermodynamic relations are used to describe the interaction model being fitted to experimental data. An informed approach to the choice of model is thus the key element to success in evaluation of the binding constant and the statistical uncertainty inherent in the estimate of its magnitude.

APPENDIX: ERROR ELLIPSES FOR BIVARIATE GAUSSIAN DISTRIBUTIONS

In assessing confidence intervals or regions for parameters evaluated by using techniques based on least square fitting, it is often adequate to

assume that variations in alternative possible data sets that would arise as a result of random experimental errors would, in turn, give rise to estimates of the n fitted parameters $\mathbf{a} = (a_1, a_2, a_3, \ldots, a_n)$ that are distributed in a simple Gaussian fashion:

$$p(\mathbf{a}) = \frac{1}{(2\pi)^{1/2}\sqrt{\det \alpha^{-1}}} \exp\left(-\frac{1}{2}\mathbf{a}^t\boldsymbol{\alpha}\mathbf{a}\right), \qquad \text{(A1)}$$

where $\mathbf{a}^t$ is the transpose of the vector $\mathbf{a}$, and $\boldsymbol{\alpha}$ is the curvature matrix whose klth element is defined by

$$\alpha_{kl} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left[\frac{\partial f(x_i, \mathbf{a})}{\partial \mathbf{a}_k}\frac{\partial f(x_i, \mathbf{a})}{\partial \mathbf{a}_l}\right], \qquad \text{(A2)}$$

for a set of N data points (x_i, y_i) with individual standard errors σ_i and a fitting function $f(\mathbf{x}, \mathbf{a})$. The inverse of the curvature matrix, α^{-1}, is the correlation matrix, $\mathbf{C}$, and det α^{-1} represents the determinant of $\mathbf{C}$.

The correlation matrix characterizes the goodness of fit of the chosen fitting function $f(\mathbf{x}, \mathbf{a})$ to the experimental data set. The diagonal elements of $\mathbf{C}$ are measures of the standard errors in the estimates of the parameters and the off-diagonal elements measure correlations in the estimates of separate parameters. The correlation matrix $\mathbf{C}_m$ for any subset m of the n parameters is created simply by striking out of $\mathbf{C}$ the rows and columns corresponding to the $(n - m)$ parameters not of interest. The joint distribution of the reduced parameter set $\mathbf{a}_m$ is then given by

$$p_m(\mathbf{a}_m) = \frac{1}{(2\pi)^{m/2}\sqrt{\det \mathbf{C}_m}} \exp\left(-\frac{1}{2}\mathbf{a}_m^t\mathbf{Q}_m\mathbf{a}_m\right), \qquad \text{(A3)}$$

where $\mathbf{Q}_m = \mathbf{C}_m^{-1}$ is the inverse of $\mathbf{C}_m$. For such a multivariate Gaussian distribution, confidence regions can be represented as hyper-ellipsoids in the parameter space of interest.

The ellipse defining a confidence region for the bivariate Gaussian distribution of a subset composed of just two parameters $\mathbf{a}_2 = (a_k, a_l)$ is the locus of points defined by the parametric equation

$$\begin{pmatrix} a_k \\ a_l \end{pmatrix} = \lambda_1\mathbf{e}_1\cos\theta + \lambda_2\mathbf{e}_2\sin\theta, \qquad \text{(A4)}$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the eigenvectors of the matrix $\mathbf{Q}_2$. The lengths of the semi-axes of the ellipse are given in terms of the corresponding eigenvalues, μ_1 and μ_2, of $\mathbf{Q}_2$ as

$$\lambda_1 = \sqrt{\frac{c}{\mu_1}}; \quad \lambda_2 = \sqrt{\frac{c}{\mu_2}}, \qquad \text{(A5)}$$

where $c = \mathbf{a}_m^t\mathbf{Q}_m\mathbf{a}_m$ is a constant that defines the chosen level of confidence: $c = \chi^2$ (95%, $m = 2$) for a 95% level of confidence. The ellipse defines a locus of points around which $p_2(\mathbf{a}_2)$ is uniform. The integral of p_2 over the region inside the ellipse is the chosen level of confidence.

It should be noted that the use of error ellipses as we have described depends upon the validity of the assumption that the multivariate parameter distribution is Gaussian in respect to the real error distribution of the data set. This assumption can be tested by, for example, Monte Carlo simulation—the procedure adopted in the present study.

The support of this investigation by the Australian Research Council is gratefully acknowledged, as is the assistance and mathematical expertise of Sze M. Tan in undertaking the statistical analyses.

REFERENCES

Bailey, M. F., B. E. Davidson, A. P. Minton, W. H. Sawyer, and G. J. Howlett. 1996. The effect of self-association on the interaction of the *Escherichia coli* regulatory protein TyrR with DNA. *J. Mol. Biol.* 263:671–684.

Haschemeyer, R. H., and W. F. Bowers. 1970. Exponential analysis of concentration or concentration difference data for discrete molecular weight distributions in sedimentation equilibrium. *Biochemistry.* 9:435–445.

Hill, R. L. 1955. Molecular clusters in an imperfect gas. *J. Chem. Phys.* 23:617–622.

Hill, R. L., and Y. D. Chen. 1973. Theory of aggregation in solution. I. General equations and application to the stacking of bases, nucleosides, etc. *Biopolymers.* 12:1285–1312.

Jacobsen, M. P., P. R. Wills, and D. J. Winzor. 1996. Thermodynamic analysis of the effects of small inert cosolutes in the ultracentrifugation of noninteracting proteins. *Biochemistry.* 35:13173–13179.

Johnson, M. L., J. J. Correia, D. A. Yphantis, and H. H. Halvorson. 1981. Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* 36:575–588.

Kim, S.-J., T. Tsukiyama, M. S. Lewis, and C. Wu. 1994. Interaction of the DNA-binding domain of *Drosophila* heat shock factor with its cognate DNA site: a thermodynamic analysis using analytical ultracentrifugation. *Protein Sci.* 3:1040–1051.

Laue, T. M., D. F. Senear, S. Eaton, and J. B. A. Ross. 1993. 5-Hydroxytryptophan as a new intrinsic probe for investigating protein–DNA interactions by analytical ultracentrifugation: study of the effect of DNA on self-assembly of the bacteriophage λ *c*I repressor. *Biochemistry.* 32:2469–2472.

Lewis, M. S., R. I. Shrager, and S.-J. Kim. 1993. Analysis of protein–nucleic acid and protein–protein interactions using multi-wavelength scans from the XL-A analytical ultracentrifuge. *In* Modern Analytical Ultracentrifugation. T. M. Schuster and T. M. Laue, editors. Birkhäuser, Boston. 94–115.

McMillan, W. G., and J. E. Mayer. 1945. The statistical thermodynamics of multicomponent systems. *J. Chem. Phys.* 13:276–305.

Milthorpe, B. K., P. D. Jeffrey, and L. W. Nichol. 1975. Direct analysis of sedimentation equilibrium results obtained with polymerizing systems. *Biophys. Chem.* 3:169–176.

Wills, P. R., and Y. Georgalis. 1981. Concentration dependence of the diffusion coefficient of a dimerizing protein: bovine pancreatic trypsin inhibitor. *J. Phys. Chem.* 85:3978–3984.

Wills, P. R., and D. J. Winzor. 1992. Thermodynamic nonideality and sedimentation equilibrium. *In* Analytical Ultracentrifugation in Biochemistry and Polymer Science. S. E. Harding, A. J. Rowe, and J. C. Horton, editors. Royal Society of Chemistry, Cambridge, U.K. 311–330.

Wills, P. R., L. W. Nichol, and R. J. Siezen. 1981. The indefinite self-association of lysozyme: consideration of composition-dependent activity coefficients. *Biophys. Chem.* 11:71–82.

Wills, P. R., M. P. Jacobsen, and D. J. Winzor. 1996. Direct analysis of solute self-association by sedimentation equilibrium. *Biopolymers.* 38:119–130.

Wills, P. R., M. P. Jacobsen, and D. J. Winzor. 1997. Direct analysis of sedimentation equilibrium distributions reflecting macromolecular interactions. *Progr. Colloid Polym. Sci.* 107:1–10.

Wills, P. R., D. R. Hall, and D. J. Winzor. 2000. Interpretation of thermodynamic non-ideality in sedimentation equilibrium experiments on proteins. *Biophys. Chem.* 84:217–225.

Winzor, D. J., and P. R. Wills. 1995. Thermodynamic nonideality and protein solvation. *In* Protein–solvent interactions. R. B. Gregory, editor. Marcel Dekker, New York, 483–520.

Winzor, D. J., M. P. Jacobsen, and P. R. Wills. 1998. Direct analysis of sedimentation equilibrium distributions reflecting complex formation between dissimilar reactants. *Biochemistry.* 37:2226–2233.

Winzor, D. J., M. P. Jacobsen, and P. R. Wills. 1999. Allowance for thermodynamic nonideality in the analysis of sedimentation equilibrium distributions reflecting complex formation between dissimilar reactants. *Progr. Colloid Polym. Sci.* 113:69–75.