JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Theory and Simulation of Multicomponent Osmotic Systems

Sadish Karunaweera,[†] Moon Bae Gee,[†] Samantha Weerasinghe,[†,‡] and Paul E. Smith*[,†]

[†]Department of Chemistry, Kansas State University, Manhattan, Kansas 66506, United States

[‡]Department of Chemistry, University of Colombo, Colombo 00300, Sri Lanka

ABSTRACT: Most cellular processes occur in systems containing a variety of components, many of which are open to material exchange. However, computer simulations of biological systems are almost exclusively performed in systems closed to material exchange. In principle, the behavior of biomolecules in open and closed systems will be different. Here, we provide a rigorous framework for the analysis of experimental and simulation data concerning open and closed multicomponent systems using the Kirkwood−Buff (KB) theory of solutions. The results are illustrated using computer simulations for various concentrations of the solutes Gly, Gly_2, and Gly_3 in both open and closed systems, and in the absence or presence of NaCl as a cosolvent. In addition, KB theory is used to help rationalize the aggregation properties of the solutes. Here, one observes that the picture of solute association described by the KB integrals, which are directly related to the solution thermodynamics, and that provided by more physical clustering approaches are different. It is argued that the combination of KB theory and simulation data provides a simple and powerful tool for the analysis of complex multicomponent open and closed systems.

INTRODUCTION

Most biological processes occurring under cellular conditions involve systems that are open to some form of matter exchange. In contrast, most *in vitro* experiments study systems closed to matter exchange. It is therefore important to determine any differences in behavior expected under different thermodynamic constraints between otherwise similar systems. While the properties of closed systems have been studied in detail, the study of open systems is less common and yet can provide a wealth of thermodynamic information. Furthermore, the use of computer simulations to help understand biological systems is now common practice. However, simulations of open systems of biological interest remain quite rare. The main aim of the current work is to illustrate how simulation data can be combined with a rigorous theory of solutions (for both open and closed systems) to provide insights into the behavior of biologically relevant solutes and cosolvents.

The thermodynamics of open systems have been studied in detail.[1−6] The usual way to treat binary osmotic systems of a solute (2) in a primary solvent (1) employs a virial expansion for the osmotic pressure (Π) in terms of the solute number density (ρ_2), such that

$$\beta\Pi = \sum_{n\geq 1} \frac{1}{n} B_n \rho_2^n \tag{1}$$

where $\beta = 1/RT$, $B_1 = 1$, and several terms (2−5) are typically required in the sum. We note that the above osmotic virial coefficients (B_n) differ slightly from the usual values ($B_n' = B_n/n$) in an effort to simplify some of the results shown below. In the presence of an additional cosolvent (such as NaCl), equilibrium dialysis or isopiestic distillation techniques provide an alternative to the virial expansion approach.[7,8] The above equation can be directly applied to fit the experimental data using the B_n's as fitting constants. Experimental data concerning protein−protein interactions can be obtained from B_2; however, higher order osmotic virial coefficients are not normally required due to the low protein concentrations involved.[9−11] This is not the case for smaller and/or more soluble solutes.

Most statistical thermodynamic theories attempt to relate the virial coefficients to the underlying solute molecular distribution functions.[1,3,12] One of the more versatile approaches is provided by the Kirkwood−Buff (KB) theory of solutions.[12,13] KB theory provides thermodynamic expressions for various properties of both open and closed systems in terms of integrals over molecular distribution functions, commonly referred to as KB integrals (KBIs). In contrast to the traditional McMillan−Mayer (MM) approach, the resulting KB related expressions can easily be applied at any concentration in any multicomponent system. Furthermore, the combined use of KB theory and molecular simulation appears quite natural, as the KBIs can be obtained directly from the simulation data at the composition of interest.[14]

The application of KB theory to open systems has been widely recognized.[4,12,15,16] However, only recently have specific applications to evaluate either experimental or simulation data appeared. Kirkwood and Buff recognized the possible uses of their theory for osmotic systems in their original paper.[12] O'Connell and co-workers have since used KB theory to probe the exact relationships between osmotic virial coefficients and other thermodynamic properties of solution mixtures.[15] More recently, KB theory has been used to directly rationalize osmotic pressure data[16] and to reinterpret light scattering data which can also provide estimates of the second virial coefficient for proteins.[17] Just in the past decade, a considerable effort has focused on understanding equilibrium dialysis, and other closely related experimental data, in terms of cosolvent preferential binding.[14,18−22] Finally, KB theory has also been

Special Issue: Wilfred F. van Gunsteren Festschrift

Received: January 31, 2012
Published: May 28, 2012

applied to the study of reactive and association equilibria in a variety of ensembles.[23−26] Here, we extend these previous approaches to (i) provide a simple analysis of experimental osmotic pressure data, (ii) indicate how one can obtain valuable information concerning solute−solute distributions, (iii) compare and contrast similar properties in both open and closed systems, and (iv) illustrate how one can use KB theory to probe association equilibria describing the aggregation of solutes.

The application of KB theory to open systems can be further illustrated using computer simulation data. The simulation of open systems by Monte Carlo methods is quite straightforward.[27,28] Molecular dynamics simulations of open systems are more problematic due to technical issues surrounding particle creation and annihilation.[29] The simplest methods involve the application of semipermeable physical boundaries (virtual membranes) between various regions of the system which directly mimic the experimental situation.[30−32] A similar approach is adopted here for the study of small Gly_n (n = 1−3) solutes with and without NaCl as a cosolvent.

THEORY

General Background. In the following sections, we will consider solutions containing a principle solvent (1), a solute (2), and in some cases an additional cosolvent (3). The equilibrium concentration of each species is expressed in terms of the number densities (molarities), $\rho_i = N_i/V$, or dimensionless molalities, $m_i = \rho_i/\rho_1$, and each species has an associated chemical potential, μ_i. Temperature will be assumed to be constant throughout. The osmotic system(s) of interest involve a central fixed volume of interest (V), which is separated from a large bulk solvent region by a barrier permeable (open) to the solvent, and in some applications the cosolvent, but not to the solute. The bulk solvent is held at a constant chemical potential (μ_1) defined by the solvent at a particular temperature and a fixed outside pressure (P_O). Here, we use pure water at a temperature T = 298.15 K and a pressure P_O = 1 bar throughout. The pressure generated inside the central fixed volume region (P_I) in the presence of the solute then provides the osmotic pressure via $\Pi = P_I - P_O$. The osmotic pressure, the virial coefficients, and the integrals defined below are then a function of T, $\mu_1(P_O)$, and ρ_2. In the presence of a cosolvent, the dependence extends to include ρ_3, when the barrier is impermeable (closed) to the cosolvent, or μ_3, when the barrier is permeable (open) to the cosolvent. However, in the following sections, we have not included all of these dependencies in an effort to simplify the notation used.

Kirkwood−Buff Theory of Binary Osmotic Systems. In this section, we outline how KB theory can be used to understand osmotic systems. One of the advantages of using KB theory is that the solution thermodynamics can be formulated in terms of integrals which have a well-defined physical significance. This is also true of the MM theory of solutions, but there one is restricted to an interpretation in terms of distributions at infinite dilution in the primary solvent.[13] This restriction is not required by KB theory, although MM theory is obtained, as expected, under infinite dilution conditions. The following integrals are required:[12]

$$G_{\alpha\beta} = \frac{1}{V}\iint [g^{(2)}_{\alpha\beta}(r_1, r_2) - 1]\, dr_1\, dr_2$$

$$G_{\alpha\beta\gamma} = \frac{1}{V}\iiint [g^{(3)}_{\alpha\beta\gamma}(r_1, r_2, r_3) - g^{(2)}_{\alpha\beta}(r_1, r_2) - g^{(2)}_{\alpha\gamma}(r_1, r_3) - g^{(2)}_{\beta\gamma}(r_2, r_3) + 2]\, dr_1\, dr_2\, dr_3 \quad (2)$$

and correspond to integrals over the orientationally averaged two body $g^{(2)}$ and three body $g^{(3)}$ distribution functions between the centers of mass of species α, β, and γ, defined in the Grand Canonical ensemble, and integrated over all relative center of mass positions r_1 of particle 1 of species α, etc. They clearly resemble the integrals appearing in the treatment of imperfect gases or the MM theory of solutions.[33] The key difference is that the solute integrals G_{22} and G_{222} are composition dependent in KB theory. Hence, the distributions (g_{22} etc.) are for pairs of solute molecules after averaging over all other solute and solvent degrees of freedom at the composition of interest. The physical interpretation of the G_{22} integral in open systems is quite simple. A positive value indicates a tendency for the solute to self-associate, while a negative value indicates a preference for solute solvation. We will see that G_{222} provides a measure of triplet solute correlations and determines how G_{22} changes with composition. Alternatively, one can express the above integrals in terms of particle−particle number fluctuations,

$$\rho_2(1 + \rho_2 G_{22}) = \frac{\langle \delta N_2 \delta N_2 \rangle}{V} = F_{22} \quad (3)$$

and

$$\rho_2(1 + 3\rho_2 G_{22} + \rho_2^2 G_{222}) = \frac{\langle \delta N_2 \delta N_2 \delta N_2 \rangle}{V} = F_{222} \quad (4)$$

where $\delta N_2 = N_2 - \langle N_2 \rangle$ and the angular brackets denote an ensemble average for a local region within the solution mixture. Here, N_2 is the instantaneous number of solute molecules observed in a small local fixed volume of the solution open to all species. KB theory relates the properties (particle number fluctuations) of systems open to all species to the properties of semiopen (osmotic) or closed (isothermal isobaric) systems under the same average thermodynamic conditions. We note that one does not have to use the commonly employed superposition approximation for the triplet distributions, or invoke additive potentials, when using KB theory. The evaluation of G_{22} and other G_{ij} values for various solutes represents the major focus of this work.

The application of KB theory to binary osmotic systems provides expressions for derivatives of the osmotic pressure in terms of the above integrals and the solute number density. The first derivative is given by[12]

$$\beta\left(\frac{\partial \Pi}{\partial \rho_2}\right)_{\mu_1} = \frac{1}{1 + \rho_2 G_{22}} \quad (5)$$

Clearly, ideal osmotic behavior requires either a small solute concentration or G_{22} = 0 for all compositions. A tendency for solute self-association ($G_{22} > 0$) would result in a lower than ideal ($\beta\Pi^{id} = \rho_2$) osmotic pressure as the solute concentration is increased and *vice versa*. An expression for the second derivative has also been provided and can be written[26]

$$\beta\left(\frac{\partial^2\Pi}{\partial\rho_2^2}\right)_{\mu_1} = -\frac{G_{22} + \rho_2(G_{222} - G_{22}^2)}{(1 + \rho_2 G_{22})^3} \tag{6}$$

Both derivative expressions apply at any solute concentration. Taking derivatives of the right-hand side of eq 5 and equating with the right-hand side of eq 6 provides an expression for the derivative of G_{22} with respect to solute concentration at constant T and solvent chemical potential (all such derivatives will be indicated with a prime)

$$G_{22}' = \frac{G_{222} - 2G_{22}^2}{1 + \rho_2 G_{22}} \tag{7}$$

Hence, if $G_{222} = 2G_{22}^2$ for all compositions, the value of G_{22} will be independent of composition, whereas one requires $G_{222} = 0$ for ideal systems. However, when $G_{222} > 2G_{22}^2$, then G_{22} will tend to increase with composition and *vice versa*. When G_{22} is independent of composition, one finds that $\beta\Pi = G_{22}^{-1} \ln(1 + \rho_2 G_{22})$.

Given a set of osmotic virial coefficients, one can directly express the composition dependence of G_{22} (G_{22}') and G_{222} according to

$$G_{22} = -\frac{Y_n}{\rho_2(1 + Y_n)}$$

$$G_{222} = \frac{Y_n(1 + Y_n)(1 + 2Y_n) - \rho_2 Y_n'}{\rho_2^2(1 + Y_n)^3}$$

$$Y_n = \sum_{n\geq 2} B_n \rho_2^{n-1} \tag{8}$$

The above expressions describe the composition dependence of the experimental or simulated solute self-association and represent the principle quantities of interest in this study. Expansion of the above expressions in a power series in the solute number density leads to

$$G_{22} \approx -B_2 - [B_3 - B_2^2]\rho_2 - [B_4 - 2B_2B_3 + B_2^3]\rho_2^2 - \ldots$$

$$G_{222} \approx -[B_3 - 3B_2^2] - [2B_4 - 9B_2B_3 + 3B_2^3]\rho_2 - \ldots \tag{9}$$

and provide the limiting values of G_{22} and G_{222} for an infinitely dilute solute

$$G_{22}^\infty = -B_2,\ G_{222}^\infty = 3B_2^2 - B_3 \tag{10}$$

together with the derivative of G_{22}

$$G_{22}^{\infty\prime} = G_{222}^\infty - 2(G_{22}^\infty)^2 = B_2^2 - B_3 \tag{11}$$

The above expressions are necessarily equivalent to those of MM theory, except for the fact that we have not inferred the superposition approximation for the triplet potential of mean force to simplify and evaluate B_3. The above relationships lead to the following osmotic pressure expansion:

$$\beta\Pi = \rho_2 - \frac{1}{2}G_{22}^\infty\rho_2^2 - \frac{1}{3}[G_{222}^\infty - 3(G_{22}^\infty)^2]\rho_2^3 + \ldots \tag{12}$$

which provides the B_2 and B_3 coefficients in terms of KB integrals and is in agreement with previous results.[12] Hence, MM theory is obtained from KB theory when the required derivatives are obtained at infinitely dilute solute concentrations.

The above expressions can be used to analyze experimental or simulated osmotic pressure data for any type of solute. It should be noted, however, that the G_{22} integral diverges ($\propto \rho_2^{-1/2}$) for low concentration salt solutions.[34] KB theory can still be applied to study salt solutions, but with less interpretive power as provided for nonionic systems. For both ionic solutes and cosolvents, we then distinguish between the traditional salt concentration (ρ_s) and the total ion concentration (ρ_2 or ρ_3).[35] Before leaving this section, we note that KB theory can be used to provide an expansion in terms of solute molality,[25,36,37] but the expressions then involve the G_{21} integrals and become somewhat more complicated to interpret.

Kirkwood–Buff Theory of Ternary Osmotic Systems. Ternary osmotic systems are more complicated and yet just as important. In particular, the effects of osmolytes (or molecular crowding) on protein folding and association under cellular (open) conditions requires a detailed knowledge of osmotic systems and their behavior.[38−40] Here, we provide expressions to illustrate the effects of a cosolvent (3) on the osmotic pressure displayed by a solute (2) in a primary solvent (1), which depend on whether the system is open or closed with respect to cosolvent. The following expressions then hold:[25]

$$RTd\ln\rho_1 = (1 + N_{11})d\mu_1 + N_{12}d\mu_2 + N_{13}d\mu_3$$

$$RTd\ln\rho_2 = N_{21}d\mu_1 + (1 + N_{22})d\mu_2 + N_{23}d\mu_3$$

$$RTd\ln\rho_3 = N_{31}d\mu_1 + N_{32}d\mu_2 + (1 + N_{33})d\mu_3$$

$$dP = \rho_1 d\mu_1 + \rho_2 d\mu_2 + \rho_3 d\mu_3 \tag{13}$$

where we have written $N_{ij} = \rho_j G_{ij}$, and the last expression corresponds to the Gibbs–Duhem equation at constant T. These differentials can be applied toward the analysis of systems in any ensemble where T is held constant. Several different cases will be considered.

If the system is open to both the solvent and the cosolvent, then one has $d\mu_1 = d\mu_3 = 0$ and $dP = d\Pi$, which on insertion into the above expressions provide

$$\beta\left(\frac{\partial\Pi}{\partial\rho_2}\right)_{\mu_1,\mu_3} = \frac{1}{1 + N_{22}} \tag{14}$$

In this situation, there is no explicit dependence of the osmotic pressure on the KB integrals involving either the solvent or cosolvent. However, the value of G_{22} will depend implicitly on the cosolvent concentration. The difference between the G_{22} values in the presence and absence of the cosolvent can be obtained from

$$\beta\left(\frac{\partial\Pi}{\partial\rho_2}\right)_{\mu_1,\mu_3} - \beta\left(\frac{\partial\Pi}{\partial\rho_2}\right)_{\mu_1} \approx -[N_{22}(\rho_3) - N_{22}(0)] \tag{15}$$

which is valid for low solute concentrations. If $G_{22}(\rho_2) > G_{22}(0)$, then the presence of the cosolvent tends to increase the self-association of the solute and is characterized by a lower solute osmotic pressure in the presence of the cosolvent compared to that in pure solvent (for the same solute concentration). The above conditions are the same as found in equilibrium dialysis experiments. Here, one can quantify the relative binding of the cosolvent (G_{23}) and solvent (G_{21}) to the solute via the preferential binding parameter[36,37]

$$\Gamma_{23} = \left(\frac{\partial m_3}{\partial m_2}\right)_{\mu_1,\mu_3} = \frac{N_{23} - m_3 N_{21}}{1 + N_{22} - N_{12}} \tag{16}$$

where $m_i = \rho_i/\rho_1$ is the (dimensionless) molality of i. This property is particularly useful when describing the effects of cosolvents on molecular association as demonstrated below.

If the system is only open to the solvent, then one has $d\mu_1 = d\rho_3 = 0$ and $dP = d\Pi$, which on insertion into the above expressions provide

$$\beta\left(\frac{\partial \Pi}{\partial \rho_2}\right)_{\mu_1,\rho_3} = \frac{1 + N_{33} - N_{23}}{(1 + N_{22})(1 + N_{33}) - N_{23}N_{32}} \tag{17}$$

where N_{23} can be considered a measure of the solute−cosolvent affinity. When species 2 and 3 are both proteins, this provides insight into mixed protein−protein interactions that can be extracted from experimental osmotic data. In either case, the above expression reduces to eq 5 when $G_{23} = 0$. Equation 17 is much more complicated in comparison to eq 5 or 14. However, one can extract information on the cosolvent and solute association via

$$\beta\left(\frac{\partial \Pi}{\partial \rho_2}\right)_{\mu_1,\rho_3} - \beta\left(\frac{\partial \Pi}{\partial \rho_2}\right)_{\mu_1,\mu_3} \approx -N_{23} \tag{18}$$

which is valid for low solute and cosolvent concentrations.

Solute Association Equilibria in Osmotic and Closed Systems. The previous analysis indicates how one can obtain information concerning G_{22} for solutes. It should be noted that this is the most relevant property describing solute−solute association that relates to the thermodynamics of the solution. It involves both the direct binding between solute molecules, together with more subtle and/or long-range changes in the solute−solute distribution with respect to a random bulk distribution (see eq 2). Hence, solute−solute association could increase without inferring the formation of well-defined dimers etc. However, a much more physical picture of solute−solute association is provided by spectroscopic studies, where information may be provided concerning the concentration of specific tightly bound dimers. KB theory can also be used to study these types of association equilibria.[23−26] The results for binary and ternary systems are presented here and compared to equivalent results for closed systems.

If we consider a solute which can exist as a monomer (M) and an aggregate (A) consisting of n monomers, then one can define an equilibrium constant for the association reaction nM → A such that $K = \rho_A/\rho_M^n$ under the equilibrium conditions $\mu_A = n\mu_M$. We note that the equilibrium constant defined here is not dimensionless. One could include a standard concentration in the definition of the equilibrium to make K dimensionless. However, we will only be concerned with changes in the equilibrium constant (KB theory is mute on the value of K itself), and hence this factor will disappear. Previous studies indicate that[25]

$$RTd\ln K = (N_{A1} - nN_{M1})d\mu_1 + (N_{A2} - nN_{M2})d\mu_2 + (N_{A3} - nN_{M3})d\mu_3 \tag{19}$$

for a ternary system. The above differential complements the expressions in eq 13 and involves KB integrals describing the correlation between each solute form and the primary components of the solution. The relationships between the solute KB integrals (independent of solute form) and the integrals for solute specific forms are given by[25]

$$\begin{aligned} \delta_{2j} + N_{2j} &= f_A N_{Aj} + f_M N_{Mj} \\ N_{M2} &= 1 + N_{MM} + nN_{MA} \\ N_{A2} &= n + nN_{AA} + N_{AM} \end{aligned} \tag{20}$$

where δ_{ij} is the Kronecker delta function and the monomer and aggregate fractions are given by $f_M = \rho_M/\rho_2$ and $f_A = n\rho_A/\rho_2$, respectively. More details can be found in the original literature.[25,26]

Using eqs 13 and 19 for binary systems ($\rho_3 = 0$), one finds the following expressions for the effect of increasing solute concentration on the solute association equilibrium in open

$$\begin{aligned} RT\left(\frac{\partial \ln K}{\partial \Pi}\right)_{\mu_1} &= G_{A2} - nG_{M2} \\ \left(\frac{\partial \ln K}{\partial \rho_2}\right)_{\mu_1} &= \frac{G_{A2} - nG_{M2}}{1 + N_{22}} \end{aligned} \tag{21}$$

and closed

$$\left(\frac{\partial \ln K}{\partial \rho_2}\right)_{P} = \frac{(G_{A2} - nG_{M2}) - (G_{A1} - nG_{M1})}{1 + N_{22} - N_{12}} \tag{22}$$

systems. After taking derivatives of eq 19 with respect to pressure, one can then eliminate the G_{i1} terms to provide

$$\begin{aligned} \left(\frac{\partial \ln K}{\partial \rho_2}\right)_{P} &= \frac{G_{A2} - nG_{M2} + \Delta V^*}{1 + N_{22} - \rho_2 RT\kappa_T} \\ &\approx \frac{G_{A2} - nG_{M2} + \Delta V^*}{1 + N_{22}} \end{aligned} \tag{23}$$

where ΔV^* is the change in volume for the process, which can be expressed in terms of KBIs but is simpler to interpret in this form. The above expressions demonstrate that the change in the association equilibrium differs in open and closed systems (possessing the same average thermodynamic properties) by terms in both the numerator and denominator. The compressibility term in the denominator will typically be small (10^{-3}) and can be neglected, and the difference between the ensembles is related to the magnitude of ΔV^*. In open systems, an increase in solute concentration increases the equilibrium constant if association of the solute, in any form, is larger to the aggregate than n times the monomer. In closed systems, the effect of water association is also directly present and can be represented in terms of the volume change associated with the aggregation process. Hence, open systems will resist (compared to closed systems) any processes which result in an increase in volume by a term related to $\Pi\Delta V^*$.

Ternary systems are more complicated and involve additional KB integrals. Furthermore, component 3 may be held at constant chemical potential or fixed concentration, and one can follow the equilibrium by varying either the solute or cosolvent concentration. If the cosolvent concentration is held fixed and the solute concentration varied, one finds (in addition to eq 17) that

$$RT\left(\frac{\partial \ln K}{\partial \Pi}\right)_{\mu_1,\rho_3} = \frac{(G_{A2} - nG_{M2})(1 + N_{33}) - (N_{A3} - nN_{M3})G_{23}}{1 + N_{33} - N_{23}}$$

$$\left(\frac{\partial \ln K}{\partial \rho_2}\right)_{\mu_1,\rho_3} = \frac{(G_{A2} - nG_{M2})(1 + N_{33}) - (N_{A3} - nN_{M3})G_{23}}{(1 + N_{22})(1 + N_{33}) - N_{23}N_{32}} \quad (24)$$

These expressions also describe the effect of varying the cosolvent concentration for a fixed solute concentration after a simple index change ($2 \leftrightarrow 3$).

The previous expressions are greatly simplified if we restrict ourselves to situations in which the solute concentration is negligible (a common biological situation) and the cosolvent concentration is varied. Then we find for open systems

$$RT\left(\frac{\partial \ln K}{\partial \Pi}\right)^{\infty}_{\mu_1,\rho_2} = G_{A3} - nG_{M3}$$

$$\left(\frac{\partial \ln K}{\partial \rho_3}\right)^{\infty}_{\mu_1,\rho_2} = \frac{G_{A3} - nG_{M3}}{1 + N_{33}} \quad (25)$$

while for closed systems we have

$$\left(\frac{\partial \ln K}{\partial \rho_3}\right)^{\infty}_{P,m_2} = \frac{(G_{A3} - nG_{M3}) - (G_{A1} - nG_{M1})}{1 + N_{33} - N_{13}} = \frac{\rho_3^{-1}(\Gamma^{\infty}_{A3} - n\Gamma^{\infty}_{M3})}{1 + N_{33} - N_{13}} \quad (26)$$

and provides the KB expression for the m value of protein denaturation when A → D, M → N, and $n = 1$. Performing the same manipulation as for binary systems, one finds

$$\left(\frac{\partial \ln K}{\partial \rho_3}\right)^{\infty}_{P,m_2} = \frac{G_{A3} - nG_{M3} + \Delta V^*}{1 + N_{33} - \rho_3 RT\kappa_T} \approx \frac{G_{A3} - nG_{M3} + \Delta V^*}{1 + N_{33}} \quad (27)$$

which takes a similar form as before.[41] Hence, the change in the equilibrium constant for association will be larger (more positive) in closed versus open systems when the volume change for association is positive and *vice versa*. The ease with which the above manipulations can be performed for multicomponent systems in any ensemble represents a particular advantage of the KB approach.

In summary, we have provided a series of expressions which can be applied to understand the behavior of open and closed systems. In particular, expressions describing variations in both the osmotic pressure and association equilibria in terms of a series of KB integrals have been provided. The main difference between equilibria in open and closed systems relates to the volume change accompanying the process. Hence, a significant difference between ensembles would only be expected for large volume changes and/or osmotic pressures. Finally, we want to be clear concerning the exact interpretation of the KBIs. The KBIs are defined in a Grand Canonical ensemble open to all species. Hence, $G_{ij} = G_{ij}(T, V, \mu_1, \mu_2)$ for binary systems. The KBIs obtained from an analysis of the osmotic data correspond to changes in μ_2, ρ_2, or Π, whichever is more convenient to use. The KBIs will not be the same as those obtained from isothermal isobaric (P_O) data, even if the solute and solvent compositions are identical, but will correspond to the KBIs obtained from an isothermal isobaric analysis at the same composition and the higher pressure of $P_O + \Pi$. These differences may or may not be important depending on the exact application.[6,15]

The primary use for the above expressions is 2-fold. First, one can apply the expressions provided in eqs 5−7, 14, and 17 to help interpret the experimental data concerning osmotic pressure changes, or eqs 21, 23, 25, and 27 to help interpret changes in equilibrium constants, in terms of the distribution functions between molecules provided in eq 2. Hence, one can develop a link between the experimental thermodynamic data and the relative distributions of the various species in solution. Second, one can reverse the whole process and relate the solution distributions, obtained from theory or simulation, to compare with available experimental data or to make predictions concerning the thermodynamic behavior of the solutions.

METHODS

Molecular Dynamics Simulations. All molecular dynamics simulations were performed using the KBFF models (http://kbff.chem.k-state.edu),[35,42,43] together with the SPC/E water model,[44] as implemented in the GROMACS 4.0.5 package.[45] All simulations were performed at 300 K and the pressure of interest ($P = P_O = 1$ bar or $P = P_O + \Pi$) using the weak coupling technique to modulate the temperature and pressure with relaxation times of 0.1 and 0.5 ps,[46] respectively. A time-step of 2 fs was used and the bond lengths were constrained using the Lincs (solutes) and Settle (water) algorithms.[47,48] The particle mesh Ewald technique was used to evaluate electrostatic interactions with a grid resolution of 0.1 nm.[49] A real space convergence parameter of 3.5 nm^{-1} was used in combination with twin range cutoffs of 1.0 and 1.5 nm and a nonbonded update frequency of 10 steps. Random initial configurations of molecules in a cubic box were used to study the closed systems. Initial configurations of the different solutions were generated from a cubic box ($L \approx 6.0$ nm) of equilibrated water molecules by randomly replacing waters with solutes until the required concentration was attained. The steepest descent method was then used to perform 100 steps of energy minimization. This was followed by extensive equilibration, which was continued until the rdf's displayed no drift with time (typically 5 ns). Total simulation times were in the 25−50 ns range, and the final 25−30 ns were used for calculating ensemble averages. Configurations were saved every 0.1 ps for the calculation of various properties. Errors ($\pm 1\sigma$) in the simulation data were estimated by using five block averages.

Osmotic Simulations. There are several simulation techniques available to study osmotic systems. Here, we take a very simple physical approach. Simulations of systems extended in the z direction ($6 \times 6 \times 24$ nm) were performed which included a series of Lennard-Jones (LJ) particles to act as two semipermeable walls separating the bulk solution from a central semiopen region of interest. The LJ "walls" were separated by a z distance of 12 nm, and all solutes were placed in the central region between the two walls. The parameters for the LJ particles were taken to be 0.3 nm and 0.02 kJ/mol, and

each wall was constructed of 20 × 20 particles separated by 0.3 nm in both the x and y directions. The walls were held fixed during the simulations, and all interactions between the LJ particles and between the LJ particles and the solvent were excluded. Periodic boundary conditions were applied in all directions. Anisotropic pressure coupling was used to keep the x and y box lengths fixed and to maintain a constant pressure in the z direction. All other simulation conditions were the same as for the closed systems. The osmotic pressure was then obtained by determining the pressure on the walls provided by the nondiffusible components.[32]

Several technical issues can arise with such a system setup. First, the presence of the walls could affect the solute distribution and/or the pressure profile for the central region. This issue is discussed in the Results section. Second, the use of a finite bulk solvent region acting as the chemical potential bath leads to a drop in the pure solvent pressure, as the solvent moves into the central (low μ_1) region, when one simply couples $P_{zz} = P_T$ to a barostat at 1 bar. Hence, the outside pressure displayed by the pure solvent region will be less than 1 bar, and therefore, the system, while providing the same solvent chemical potential inside and outside the open region, will correspond to different constant solvent chemical potentials for each solute concentration. This makes it difficult to follow the equilibrium line where μ_1 is held constant, at 1 bar for instance, as performed experimentally. Fortunately, this is easy to correct if we note that the total or reference pressure (P_T), the inside pressure (P_I), and the outside pressure (P_O) are related by

$$P_T V_T = P_O V_O + P_I V_I \tag{28}$$

Hence, given the measured osmotic pressure, $\Pi = P_I - P_O$, one can then determine the outside (and thereby inside) pressure according to

$$P_O = P_T + \Pi(V_I/V_T) \tag{29}$$

To ensure that the solvent chemical potential remains at the same constant value as the solute concentration is increased, one needs to adjust the reference pressure to raise the outside pressure to the desired value of 1 bar. The above equation can be used to predict the value of P_T that is consistent with $P_O = 1$ bar, assuming the osmotic pressure is independent of P_T, and the whole processes can be iterated (2−3 cycles) to consistency. The same process was performed for the NaCl simulations where μ_1 and μ_3 were held constant, except that the target outside pressure was the osmotic pressure obtained from the NaCl solute simulations. While these adjustments are usually small, they can also be important.[6,15]

Analysis of the Simulation Data. The primary analysis involved the determination of the KBIs from the simulations. This was achieved in two ways. The first involved the usual integration of the corresponding rdf. The KBIs are defined in systems open to all components, and hence one cannot integrate over the full volume. Hence, the integration was truncated at a distance R from the central particle, where R is the distance at which the rdf approaches unity.[50] This also provides a distance dependent KBI which can be used to determine contributions from various solvation shells

$$G_{ij}(R) = \int_0^R [g_{ij}^{(2)}(r_{12}) - 1]\, dr_{12} \tag{30}$$

For this work, we used a final value of $R = 1.5$ nm. KBIs were only determined from the closed systems at the equivalent state point. The main advantage of this approach is that the rdf's and KBIs provide information concerning the "structure" of the solution surrounding the central i particle. The second approach involves the direct application of eq 3 and the determination of the appropriate particle number fluctuations. The closed systems were analyzed by considering a series of reference volumes centered on a randomly chosen origin and then averaging over these volumes. The reference volumes were chosen as cubes of length 3 nm, and approximately 10 000 origins were used. The advantage of this approach lies in the large number of origins which can be used, which greatly improves the statistical significance, but with the loss of structural information. The determination of G_{222} was performed using the particle number fluctuations and eq 4.

A more physical analysis of solute association was also performed. The most prominent interaction between the Gly_n solutes involved direct association between the N and C terminal groups, as evidenced by the atom based rdf's. Hence, a solute dimer (and trimer etc.) was defined by considering the contact distance between the nitrogen and the midway point between both oxygens of the carboxylate groups. If this distance was less than 0.5 nm, the first minimum in the rdf between these two groups, then the two solutes were considered to be associated. An iterative procedure was then applied to determine the number of solutes in each solute cluster.

RESULTS

In this section, we analyze both the experimental and simulated data for binary mixtures of water containing various concentrations of NaCl, Gly, Gly_2, and Gly_3 as solutes. In addition, simulated data for ternary mixtures of water with a solute and cosolvent are also examined. In performing the osmotic simulations, one has little control over the exact concentration of the diffusible components. Hence, while we will constantly refer to systems at 3 m Gly or 6 m NaCl etc., it should be remembered that these are approximate concentrations (to within 10%). The exact concentrations can be found in the various tables. Furthermore, the statistical noise associated with KBIs increases as the concentrations of the components decreases.[51] Therefore, in many situations, we have chosen to analyze only the simulations at high solute and cosolvent concentrations and to use the highest possible concentrations of both solute and cosolvent.

Before continuing with the present analysis, it is important to ensure there were no significant artifacts in the osmotic simulations. This is unlikely due to the fact that there is no significant desolvation process for the solutes at the walls, although effects on solute−solute distributions are still possible. The pressure and density profiles for the 6 m NaCl osmotic system are displayed in Figure 1. The pressure profile, $P(z)$, was determined using the approach outlined in previous work on surface tension.[52] However, here we approximated the pressure contributions using a simple Coulomb plus LJ potential truncated at 1.5 nm for the molecular virial, primarily due to the excessive cost involved with calculating the contributions using the full Ewald potential. Hence, the pressures do not exactly match the pressure determined during the simulation. Nevertheless, it seems clear that neither the density nor pressure profiles indicate any surface effects beyond a few molecular diameters.

The experimental and simulated osmotic pressures are displayed in Figure 2. The force fields used here performed reasonably well at low solute concentrations but displayed some deviation from experimental results at higher solute concen-

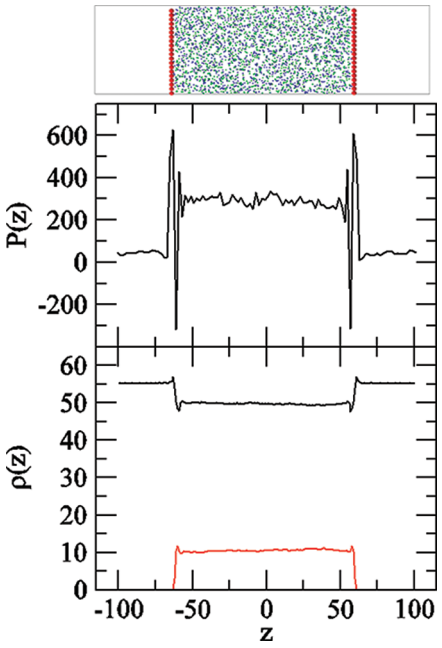

Figure 1. Pressure and concentration profiles obtained from the simulation of the 6 m NaCl osmotic system at 300 K. The top panel shows a snapshot from the simulation with water molecules removed. The LJ spheres comprising the "walls" are displayed in red. The sodium ions (blue) and chloride ions (green) are confined to the central inside region. The central panel displays the pressure profile in units of bar. The lower panel displays the molar concentrations of water (black) and ions (red).

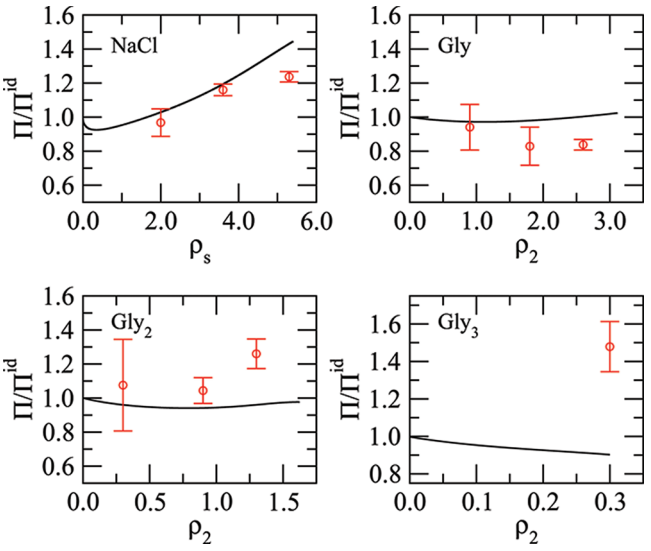

Figure 2. Experimental and simulated osmotic pressures at 300 K as a function of solute molarity. Data are displayed as Π/Π^{id} where solid lines correspond to the experimental data and symbols indicate simulated results. Experimental data taken from refs 54−58.

trations. It also appears that, even if the force fields were perfect, the estimated errors are such that one could not distinguish between the real experimental data and the ideal data provided by the van't Hoff curves, using the current simulation times. This picture changes somewhat when the focus is shifted to the KBIs, as we shall see later.

One of the goals of this work is to investigate the thermodynamics of open (and closed) systems in terms of the KBIs. The presence of the walls and the subsequent loss of periodicity hinder the determination of the KBIs for the inside region. To circumvent this problem, we have performed additional isothermal isobaric simulations at the reference pressure of P_O = 1 bar and also at a pressure of $P_O + \Pi$, using the solute and solvent concentrations obtained for the inside region. The solute−solute rdf's obtained for all three systems are displayed in Figure 3. They clearly show that the rdf's are

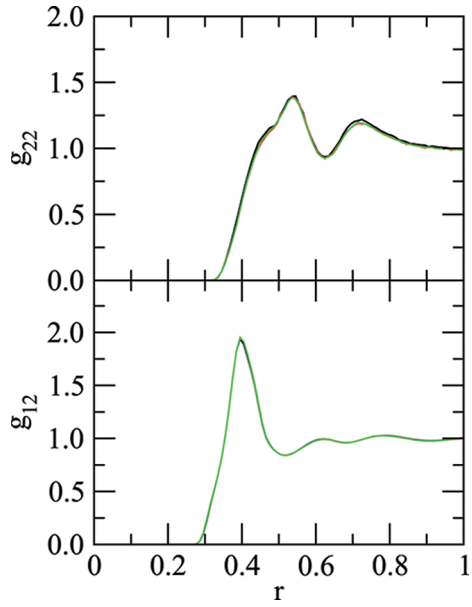

Figure 3. Solute−solute (g_{22}) and solute−solvent (g_{12}) rdf's as a function of ensemble and pressure. Data are presented for 3 m Gly as a solute, but similar observations are found for the Gly_2 and Gly_3 systems. Curves correspond to the osmotic simulation (black) and closed systems with $P = P_O$ = 1 bar (red) and $P = P_O + \Pi$ = 53 bar (green).

identical within the precision of the simulations. Hence, while the KBIs will vary with composition, they appear to be relatively insensitive to pressure, i.e., $G_{22}(T, m_2, P_O) \approx G_{22}(T, m_2, P_O + \Pi)$. This is to be expected for the relatively low pressures exhibited in the current osmotic systems. Consequently, we have obtained all of the KBIs presented here from the corresponding closed system simulations.

The experimental and simulated fluctuating quantities are provided in Tables 1 and 2 and Figure 4. The values of G_{22} for all solutes start positive and decrease with increasing solute concentration. Hence, there is a tendency for solute self-association at low solute concentrations which increases as one moves from Gly to Gly_2 to Gly_3. This behavior has been observed before in closed systems where we used the isobaric isothermal results to investigate possible group contributions to the observed association behavior.[53] A comparison of the closed (isothermal isobaric) and open (osmotic) results indicates that the G_{22} values are essentially the same, to within the typical precision of the data, which is to be expected considering the negligible pressure dependence exhibited by the rdf's in Figure 3. The simulated values of G_{22} are also provided in Table 3 and Figure 4. There is not perfect agreement with experimental results. The trends in G_{22} with composition appear to be correct, and one observes a general agreement in sign. Fortunately, unlike the raw osmotic pressure data, it does appear possible to distinguish the G_{22} values from their ideal

dx.doi.org/10.1021/ct300079v | J. Chem. Theory Comput. 2012, 8, 3493−3503

Table 1. Experimental and Simulated Binary Osmotic Virial Coefficients and KB Integrals[a]

system		B_2 / M^{-1}	B_3 / M^{-2}	B_4 / M^{-3}	G_{22}^{∞} / M^{-1}	$G_{22}^{\infty\prime}$ / M^{-2}	G_{222}^{∞} / M^{-2}
Gly	exptl	−0.104	0.082	−0.011	0.106	−0.071	−0.050
	MD	−0.260	0.075		0.260	−0.007	0.128
Gly_2	exptl	−0.361	0.484	−0.142	0.330	−0.354	−0.093
	MD	−0.529	1.051		0.529	−0.774	−0.211
Gly_3	exptl	−0.710	1.445		0.973	−0.941	0.067

[a]Obtained from a fit to eq 1 with B_1 = 1 at 298.15 K. Experimental data taken from refs 54−56.

Table 2. Summary of the Osmotic Molecular Dynamics Simulations[a]

	out		in					
system	ρ_1	ρ_s	ρ_1	ρ_2	ρ_s	Π	Π^{id}	P_T
			T, μ_1					
2.0 m NaCl	55.2		53.7		1.9	96	95	57
4.0 m NaCl	55.2		51.9		3.6	207	180	116
6.0 m NaCl	55.2		49.7		5.3	325	264	188
1.0 m Gly	55.2		53.3	0.9		21	22	10
2.0 m Gly	55.2		51.3	1.8		37	45	22
3.0 m Gly	55.2		49.3	2.6		52	65	33
0.3 m Gly_2	55.2		54.0	0.3		8	7	4
1.0 m Gly_2	55.2		51.4	0.9		23	22	10
1.5 m Gly_2	55.2		49.8	1.3		41	32	15
0.3 m Gly_3	55.2		53.3	0.3		11	7	5
			T, μ_1, μ_3					
3.0 m Gly/ 6.0 m NaCl	49.4	5.5	40.4	2.6	5.8	68	65	369
1.5 m Gly_2/ 6.0 m NaCl	47.8	5.6	43.7	1.3	5.5	54	32	358
0.3 m Gly_3/ 6.0 m NaCl	49.2	5.5	47.6	0.3	5.2	11	7	333
			T, μ_1, ρ_3					
3.0 m Gly/ 6.0 m NaCl	55.9		41.1	2.6	5.1	348	319	186
1.5 m Gly_2/ 6.0 m NaCl	55.4		42.8	1.3	5.2	399	291	205
0.3 m Gly_3/ 6.0 m NaCl	55.3		47.9	0.3	5.2	342	267	185

[a]Mixtures of Gly_n (2) and water (1) in the presence and absence of NaCl (3) at 300 K. Number densities in units of M (ρ_3 = 2 ρ_s). Pressures are in units of bar. Typical standard deviations for the measured osmotic pressures were 4 bar. The ideal osmotic pressure is given by $\beta\Pi^{id} = \rho$, where ρ is the total number density of all non-diffusable species. P_T is the total external pressure applied to each system.

values (G_{22} = 0). Furthermore, the infinite dilution KBIs obtained from a fit of the simulated osmotic pressures appear to be reasonable (see Table 1), which is probably a reflection that the largest disagreement only occurs for high solute concentrations. This is potentially important for applications in force field design, as it allows one to determine if one has a correct balance between the solute−solute, solute−solvent, and solvent−solvent distributions.

Also displayed in Figure 4 are the G_{222} values. The G_{222} values quantify the role of triplet distributions toward the thermodynamic behavior of the mixture and should be zero for ideal solutions. The experimental data suggest that triplet correlations become increasingly important for Gly_3 and display a strong dependence on concentration. In contrast, the relatively small negative values of G_{222} for Gly and Gly_2 suggest a focus on dimer association for most solute concentrations.

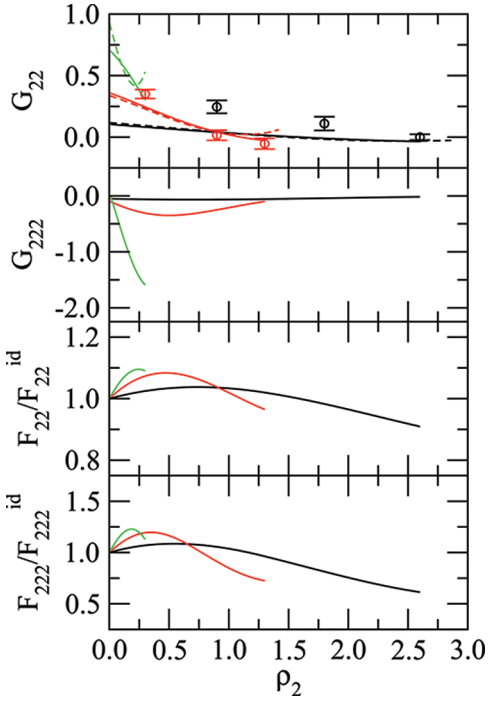

Figure 4. Experimental and simulated KBIs and solute fluctuations for Gly (black), Gly_2 (red), and Gly_3 (green) solutes as a function of solute molarity at 300 K. The values of G_{22} and G_{222} are in units of M^{-1} and M^{-2}, respectively. The colors correspond to the different solutes investigated here. Solid lines correspond to the current analysis of the experimental osmotic data, while dashed lines were obtained from an analysis of the corresponding experimental isothermal isobaric data.[54−56] The fluctuating quantities F_{22} and F_{222} are given by eqs 3 and 4 with ideal values of $F_{22}^{id} = F_{222}^{id} = \rho_2$. Symbols represent the simulated data.

The particle number fluctuations (eq 3) are also displayed in Figure 4. The data display both positive and negative deviations from ideal behavior with significant deviations even at low solute concentrations. The finite values for F_{222} also indicate that the number fluctuations are not characterized by a symmetric distribution; i.e., they are non-Gaussian. Finally, we attempted to determine F_{222} from our simulations. Even for the highest (most statistically reliable) solute concentration the value of F_{222} was found to be −0.02(30) for 3 m Gly, which is essentially meaningless using the current simulation times of 25 ns or so.

The previous analysis has centered upon the KBIs. We have argued that these are the most relevant quantities relating molecular distributions to the corresponding thermodynamics and can provide an interpretation of solute association. However, it is more typical to analyze simulation results in terms of molecular association defined by some simple distance

Table 3. Simulated KB Integrals and Preferential Interactions[a]

system	in			KBIs					
	ρ_1	ρ_2	ρ_s	G_{22}	G_{22}*	G_{23}	G_{21}	Γ_{23}	Π
				T, μ_1					
1.0 m Gly	53.3	0.9		246(52)	36		−53(2)		21
2.0 m Gly	51.3	1.8		110(56)	−33		−56(5)		37
3.0 m Gly	49.3	2.6		0(23)	−91		−51(3)		52
0.3 m Gly_2	54.0	0.3		350(37)			−83(1)		8
1.0 m Gly_2	51.4	0.9		16(41)			−83(3)		23
1.5 m Gly_2	49.8	1.3		−53(44)			−80(5)		41
0.3 m Gly_3	53.3	0.3		237(376)			−124(13)		11
				T, μ_1, μ_3					
3.0 m Gly/6.0 m NaCl	40.4	2.6	5.8	−155(2)	38	24(1)	−53(3)	1.22(5)	68
1.5 m Gly_2/6.0 m NaCl	43.7	1.3	5.5	−261(90)		−2(8)	−80(11)	1.12(12)	54
0.3 m Gly_3/6.0 m NaCl	47.6	0.3	5.2	−669(214)		−65(9)	−117(8)	0.65(13)	11

[a]Mixtures of Gly_n (2) and water (1) in the presence and absence of NaCl (3) at 300 K. Number densities in units of M ($\rho_3 = 2\ \rho_s$). Pressures in units of bar. KBIs in units of cm^3/mol. G_{22}* corresponds to the integration of G_{22} to the first minimum in the solute−solute rdf (0.62 nm for Gly).

criteria. This is also likely to be more relevant to spectroscopic data for protein association, for example. To investigate the similarities and differences between these two viewpoints, we have analyzed the degree of association of the solutes in our simulations and investigated the effect of salt on these distributions. A detailed examination of all the solute atom−atom rdf's indicated that the only significant interaction leading to dimer or higher aggregate formation occurred between the zwitterionic N and C terminal groups, and so the first minimum in this rdf was used to define the degree of solute aggregation. The results are presented in Table 3 and Figures 5 and 6.

Figure 5 displays the fraction of solute molecules observed in aggregates containing n solute molecules during the simulations. The predominant solute form was the monomer for all solutes at all concentrations. However, as the solute concentration increases, it becomes more difficult to find isolated solute molecules, indicating a potential difficulty one encounters when applying such simple models to concentrated solutions. Figure 5 also displays the equilibrium constants for association ($\ln K_n$) as a function of solute concentration. The equilibrium constant data display an increase in solute association with solute concentration for both Gly and Gly_2. This is the opposite trend to that indicated by the previous analysis on the KBIs. However, both approaches agree that solute association (dimer or trimer) increases from Gly to Gly_3. Of course, the difference between the two approaches can be reconciled when one considers that K_n will increase with solute concentration, even if there is no net affinity between the solutes, simply because one has more solutes per unit volume. Comparison with eq 21 indicates that solute association with the dimer or trimer (G_{A2}) must therefore be larger than n times the solute association with the monomer (G_{M2}).

The addition of salt had a dramatic effect on the solute association. This was demonstrated by both a significant drop in G_{22}, as indicated in Table 3, and a drop in the equilibrium constants, as shown in Figure 5. However, the underlying story was much more complicated. First, the fraction of solute molecules in either the monomer, dimer, or trimer form is increased in the presence of salt. This appears to result from a decrease in the number of high n aggregates. Second, the equilibrium constant drops in the presence of salt primarily because the monomer concentration increases. Third, as the total concentration is decreased, the fraction of monomer will naturally increase. Hence, the monomer fraction is largest for

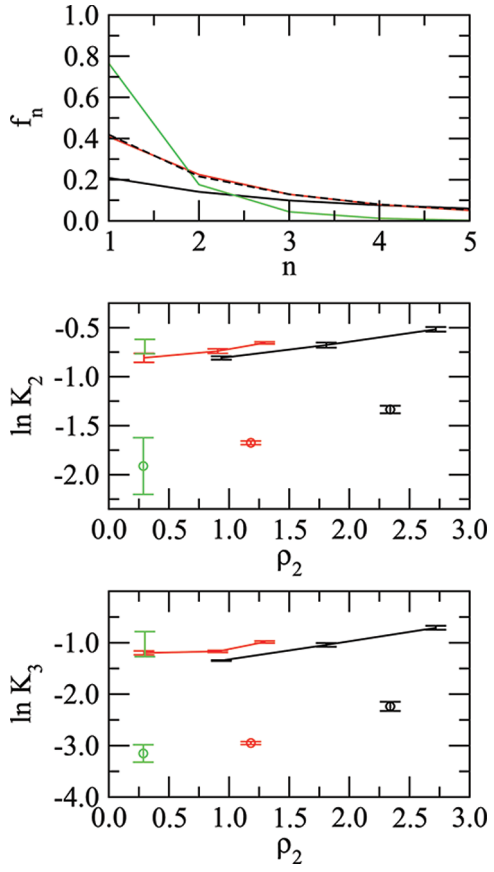

Figure 5. The fraction of solute molecules in an aggregate of n solute molecules (top) as a function of aggregate size. The equilibrium constants for dimer (middle) and trimer (bottom) formation as a function of solute molarity. See text for definitions. The solid curves correspond to 3.0 m Gly (black), 1.5 m Gly_2 (red), and 0.3 m Gly_3 (green), while the symbols and dashed curve represents the same solutes in 6.0 m NaCl.

Gly_3 at the concentrations displayed in Figure 5, even though Gly_3 displays the largest equilibrium constant for dimer or trimer formation at equivalent solute concentrations.

The change in solute association could be attributable to either a general salt screening of the large dipole−dipole interactions between the solutes, or specific binding of anions

dx.doi.org/10.1021/ct300079v | J. Chem. Theory Comput. 2012, 8, 3493−3503

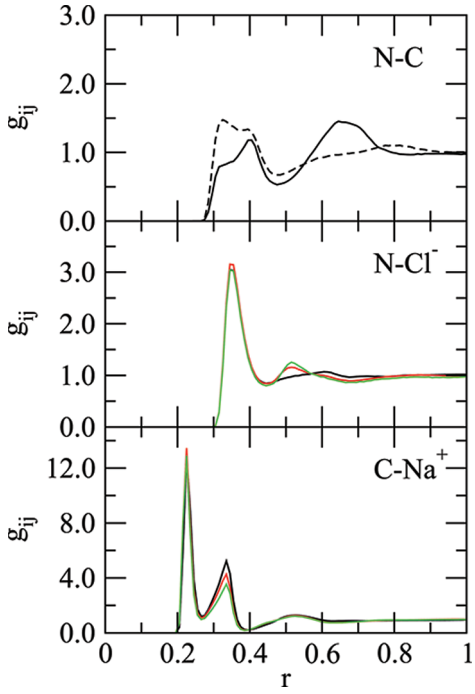

Figure 6. Solute−solute and solute−ion atom based rdf's. The N terminus to C terminus rdf for 3 m Gly in the presence (solid) and absence (dashed) of 6 m Nacl (top). The N terminus to chloride (center) and the C terminus to sodium (bottom) rdf's for various 3.0 m Gly (black), 1.5 m Gly_2 (red), and 0.3 m Gly_3 (green) concentrations in the presence of 6 m NaCl (bottom).

and/or cations with the solutes such that solute association is diminished. To investigate further, we present the relevant rdf's in Figure 6 and have also determined the corresponding preferential interaction coefficients between the solutes and NaCl, which are displayed in Table 3. The solute−solute N to C terminal rdf is changed on the addition of NaCl. The first peak is decreased, and the second peak increased in the presence of NaCl. The increased second solvation shell probability appeared to correspond to the binding of multiple Gly solutes with a shared sodium ion via their carboxylate groups. This also had an effect (−91 to 38 cm^3/mol) on the value of G_{22} truncated after the first solvation shell (denoted as G_{22}*), suggesting an increase in solute−solute contacts at short range, which must be compensated by changes at large distances. The first shell coordination numbers for the N to C termini were 0.74 and 0.48 for 3 m Gly in the absence and presence of 6 m NaCl, respectively. The values of Γ_{23} were all positive, indicating a net thermodynamic binding of salt ions with the solutes. However, the values of G_{21} were consistently larger than G_{23}, suggesting that the greater effect was due to water exclusion from the solutes rather than ion binding. Interestingly, the G_{21} values were the same in the presence and absence of 6 m NaCl. The rdf's between the ions and the terminal groups displayed in Figure 6 also support a role for ion binding. First shell coordination numbers were found to be 1.14 and 0.67 for the chloride and sodium ions, respectively, and were essentially the same for all three solutes. However, the net ion first shell coordination of 1.81 was significantly higher than that provided by the corresponding thermodynamic quantities (G_{23} or Γ_{23}). Hence, changes in solute association on the addition of salt appear to be distance dependent.

CONCLUSIONS

Expressions have been provided for the analysis of binary and ternary open and closed systems using the KB theory of solutions and the corresponding KB integrals. The KBIs provide an alternative to the cluster integrals in the MM expressions, which are much easier to determine from simulations of concentrated solutions. The expressions have been illustrated using both experimental and simulation data for small Gly, Gly_2, and Gly_3 zwitterionic peptide solutes in the presence and absence of NaCl. Two measures of solute association were investigated and found to provide different viewpoints of the association process. A thermodynamic measure of solute association is provided by G_{22}, and this is aided by the additional information concerning triplet correlations provided by G_{222}. The experimental and simulation data indicated that solute association prevails at low concentrations and increases within the series Gly < Gly_2 < Gly_3. In addition, solute association decreases with increasing solute concentration for all of the solutes. A more physical measure of solute association was investigated and expressed in terms of equilibrium constants for dimer and trimer formation. Here, an increase in the equilibrium constants was observed on increasing the solute concentration, in contrast to the thermodynamic measure of association. The differences arise as the thermodynamic measure includes changes to the solute distribution over all distances, while the physical measure focuses primarily on the first solvation shell. The addition of salt to solutions of Gly_n solutes reduces the values of G_{22} and the equilibrium constants for association. Further analysis of 3 m Gly solutions indicated that this was a consequence of the disruption of larger aggregates, leading to an increase in the number of monomers, dimers, and trimers. The overall global (long-range) effect was clearly solute disassociation as indicated by the decrease of G_{22} in the presence of NaCl, whereas solute association increased at the local (first shell) level. This suggests an overall salt screening effect that includes a local increase in dimer and trimer formation due to the binding of sodium ions with multiple solute carboxylate groups. It should be noted that, while the value of G_{22} is the most thermodynamically relevant quantity, a clear physical interpretation is often difficult, as it probes changes in the solute−solute distribution over multiple solvation shells. In contrast, the physical picture of association is quite clear but often subjective and not necessarily thermodynamically relevant. The present results therefore illustrate the advantages of a combination of KB theory and computer simulations data provide for the interpretation of complex solution behavior.

AUTHOR INFORMATION

Corresponding Author

*Tel.: 785-532-5109. Fax: 785-532-6666. E-mail: pesmith@ksu.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The project described was supported by grant R01GM079277 (P.E.S.) from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Science or the National Institutes of Health.

DEDICATION

Paul Smith would like to dedicate this article to Wilfred on this celebration of his 65th birthday. I remember with great fondness my time spent as a postdoc at the ETH in the early 1990's. As everyone knows, Wilfred is an incredibly productive researcher and always kept very busy. However, I discovered quite early that the best way to get Wilfred's full attention was to flash an equation in front of him. That would usually start a thought provoking conversation. I really enjoyed those discussions during which his love of science clearly shone through. Happy birthday, Wilfred. I hope you enjoy the equations presented here.

REFERENCES

(1) McMillan, W. G.; Mayer, J. E. *J. Chem. Phys.* **1945**, *13*, 276−305.
(2) Hill, T. L. *J. Am. Chem. Soc.* **1957**, *79*, 4885−4890.
(3) Hill, T. L. *J. Chem. Phys.* **1959**, *30*, 93−97.
(4) O'Connell, J. P. *Mol. Phys.* **1971**, *20*, 27−33.
(5) Anderson, C. F.; Courtenay, E. S.; Record, M. T. *J. Phys. Chem. B* **2002**, *106*, 418−433.
(6) Curtis, R. A.; Newman, J.; Blanch, H. W.; Prausnitz, J. M. *Fluid Phase Equilib.* **2001**, *192*, 131−153.
(7) Timasheff, S. N. *Adv. Protein Chem.* **1998**, *51*, 355−432.
(8) Cannon, J. G.; Anderson, C. F.; Record, M. T. *J. Phys. Chem. B* **2007**, *111*, 9675−9685.
(9) Lenhoff, A. M. *AIChE J.* **2003**, *49*, 806−812.
(10) Moon, Y. U.; Curtis, R. A.; Anderson, C. O.; Blanch, H. W.; Prausnitz, J. M. *J. Solution Chem.* **2000**, *29*, 699−717.
(11) Moon, Y. U.; Anderson, C. O.; Blanch, H. W.; Prausnitz, J. M. *Fluid Phase Equilib.* **2000**, *168*, 229−239.
(12) Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774−777.
(13) Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press: New York, 2006.
(14) Smith, P. E. *J. Phys. Chem. B* **2004**, *108*, 18716−18724.
(15) Cabezas, H.; O'Connell, J. P. *Ind. Eng. Chem. Res.* **1993**, *32*, 2892−2904.
(16) Terdale, S. S.; Dagade, D. H.; Patil, K. J. *J. Phys. Chem. B* **2006**, *110*, 18583−18593.
(17) Blanco, M. A.; Sahin, E.; Li, Y.; Roberts, C. J. *J. Chem. Phys.* **2011**, *134*, 225103.
(18) Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2001**, *105*, 11513−11522.
(19) Shimizu, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1195−1199.
(20) Shimizu, S.; Matubayasi, N. *Chem. Phys. Lett.* **2006**, *420*, 518−522.
(21) Shulgin, I. L.; Ruckenstein, E. *J. Chem. Phys.* **2005**, *123*, 054909.
(22) Schurr, J. M.; Rangel, D. P.; Aragon, S. R. *Biophys. J.* **2005**, *89*, 2258−2276.
(23) Perry, R. L.; O'Connell, J. P. *Mol. Phys.* **1984**, *52*, 137−159.
(24) Ben-Naim, A. *J. Chem. Phys.* **1975**, *63*, 2064−2073.
(25) Gee, M. B.; Smith, P. E. *J. Chem. Phys.* **2009**, *131*, 165101.
(26) Jiao, Y.; Smith, P. E. *J. Chem. Phys.* **2011**, *135*, 014502.
(27) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
(28) Paulsen, M. D.; Anderson, C. F.; Record, M. T. *Biophys. J.* **1988**, *53*, A483.
(29) Lynch, G. C.; Perkyns, J. S.; Pettitt, B. M. *J. Comput. Phys.* **1999**, *151*, 135−145.
(30) Murad, S.; Powles, J. G. *J. Chem. Phys.* **1993**, *99*, 7271−7272.
(31) Powles, J. G.; Murad, S.; Holtz, B. *Chem. Phys. Lett.* **1995**, *245*, 178−182.
(32) Luo, Y.; Roux, B. *J. Phys. Chem. Lett.* **2010**, *1*, 183−189.
(33) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.
(34) Kusalik, P. G.; Patey, G. N. *J. Chem. Phys.* **1987**, *86*, 5110−5116.
(35) Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *119*, 11342−12349.
(36) Smith, P. E. *J. Phys. Chem. B* **2006**, *110*, 2862−2868.
(37) Smith, P. E. *Biophys. J.* **2006**, *91*, 849−856.
(38) Zhou, H. X.; Dill, K. A. *Biochemistry* **2001**, *40*, 11289−11293.
(39) Zhou, H. X.; Rivas, G. N.; Minton, A. P. *Ann. Rev. Biophys.* **2008**, *37*, 375−397.
(40) Timasheff, S. N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9721−9726.
(41) Aburi, M.; Smith, P. E. *J. Phys. Chem. B* **2004**, *108*, 7382−7388.
(42) Kang, M.; Smith, P. E. *J. Comput. Chem.* **2006**, *27*, 1477−1485.
(43) Ploetz, E. A.; Bentenitis, N.; Smith, P. E. *Fluid Phase Equilib.* **2010**, *290*, 43−47.
(44) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(45) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(46) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(47) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(48) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.
(49) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(50) Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2003**, *107*, 3891−3898.
(51) Matteoli, E.; Lepori, L. *J. Chem. Phys.* **1984**, *80*, 2856−2863.
(52) Alejandre, J.; Tildesley, D. J.; Chapela, G. A. *J. Chem. Phys.* **1995**, *102*, 4574−4583.
(53) Kang, M.; Smith, P. E. *Int. J. Thermophys.* **2010**, *31*, 793−804.
(54) Ellerton, H. D.; Dunlop, P. J.; Mulcahy, D. E.; Reinfelds, G. *J. Phys. Chem.* **1964**, *68*, 398−402.
(55) Smith, E. R. B.; Smith, P. K. *J. Biol. Chem.* **1940**, *135*, 273−279.
(56) Venkatesu, P.; Lee, M. J.; Lin, H. M. *J. Chem. Thermodyn.* **2007**, *39*, 1206−1216.
(57) Robinson, R. A.; Stokes, R. H. *Electrolyte Solutions*; Butterworths: London, 1959.
(58) Sohnel, O.; Novotny, P. *Densities of Aqueous Solutions of Inorganic Substances*; Elsevier: Amsterdam, 1985.