The Journal of Chemical Physics

AIP Publishing

RESEARCH ARTICLE | OCTOBER 31 2023

Multicomponent solutions: Combining rules for multisolute osmotic virial coefficients

Hikmat Binyaminov ; Janet A. W. Elliott

J. Chem. Phys. 159, 164116 (2023)
https://doi.org/10.1063/5.0166482

The Journal
of Chemical Physics

ARTICLE

pubs.aip.org/aip/jcp

Multicomponent solutions: Combining rules for multisolute osmotic virial coefficients

Cite as: J. Chem. Phys. 159, 164116 (2023); doi: 10.1063/5.0166482
Submitted: 5 July 2023 • Accepted: 26 September 2023 •
Published Online: 31 October 2023

View Online Export Citation CrossMark

Hikmat Binyaminov and **Janet A. W. Elliott**[a)]

AFFILIATIONS

Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta T6G 1H9, Canada

[a)] **Author to whom correspondence should be addressed:** janet.elliott@ualberta.ca

ABSTRACT

This paper presents an exploration of a specific type of a generalized multicomponent solution model, which appears to be first given by Saulov in the current explicit form. The assumptions of the underlying theory and a brief derivation of the main equation have been provided preliminarily for completeness and notational consistency. The resulting formulae for the Gibbs free energy of mixing and the chemical potentials are multivariate polynomials with physically meaningful coefficients and the mole fractions of the components as variables. With one additional assumption about the relative magnitudes of the solvent–solute and solute–solute interaction exchange energies, combining rules were obtained that express the mixed coefficients of the polynomial in terms of its pure coefficients. This was done by exploiting the mathematical structure of the asymmetric form of the solvent chemical potential equation. The combining rules allow one to calculate the thermodynamic properties of the solvent with multiple solutes from binary mixture data only (i.e., each solute with the solvent), and hence, are of practical importance. Furthermore, a connection was established between the osmotic virial coefficients derived in this work and the original osmotic virial coefficients of Hill found by employing a different procedure, illustrating the equivalency of what appears to be two different theories. A validation of the combining rules derived here has been provided in a separate paper where they were successfully used to predict the freezing points of ternary salt solutions of water.

I. INTRODUCTION

There exist numerous nonideal solution theories in literature. We will not attempt a review here but rather list a few that are essential for providing a context for the approach presented in this paper. Regular solution theory[1] is the most widely used nonideal solution theory mainly because of its simplicity. It considers only pairwise interactions and assumes random mixing. The quasi-chemical treatment[1] improves on regular solution theory by considering pairwise additive interactions without assuming random mixing (we restrict the usage of the term "regular" to the former case; Guggenheim's definition includes the quasi-chemical treatment as part of regular solution theory). The quasi-chemical approach provides a good balance of accuracy and simplicity. It is also commonly used, especially to capture the phase-change behavior of solutions more accurately. By accounting for complex interactions in larger particle groups, formally exact theories[2–4] avoid all approximations, but they are rarely used in practice due to their complexity and/or for being computationally expensive.

Most solution models have been originally developed for binary mixtures and then extended to multicomponent mixtures. This extension is straight forward in the context of simple theories, such as regular solution theory, but it is challenging in more accurate frameworks since they are complicated by nature. Hence, it is not surprising that explicit multicomponent formulae are usually not provided even when the existence of such an extension is obvious. Consequently, in practice, while the properties of a binary mixture can be calculated with any theory depending on the required accuracy, only the simple models can be effectively used for multicomponent solutions. In contrast, it is rare for mixtures in nature as well as in industrial processes to only contain two components. Consider, for example, solutions studied in biology,[5–7] geology,[8] oil processing,[9,10] metallurgy,[11–14] and atmospheric physics.[15–20] Therefore, it is desirable to have an accurate multicomponent solution model with manageable complexity so that practically-useful equations can be developed.

In this work, we look at a solution made of an arbitrary number of components by considering interactions in groups consisting of an arbitrary number of particles, but nevertheless, we assume random mixing. This approach can be viewed as an extension of regular solution theory in a specific way: while the quasi-chemical treatment addresses the nonrandomness of mixing in a system with pairwise additive interactions, the method employed herein accounts for multi-body interactions with random mixing. The objectives of this paper are twofold: (i) to give a theoretical basis for the virial equation for the change in the chemical potential of the solvent (i.e., multisolute osmotic virial equation) and derive its coefficients from first principles, and (ii) to provide a method of estimating the chemical potential of a solvent in the presence of multiple solutes by deriving combining rules for the mixed coefficients.

In the first part of this paper, we list the assumptions of the present model and rederive the corresponding equation for the Gibbs free energy of mixing. Except for the structure of the arguments, some nuances in the assumptions, and notational differences, our formulation and the resulting equation are the same as what was obtained before by Saulov[21] (and by Kakuda *et al.*[22] for special cases). We give a brief, equivalent derivation here for the sake of completeness and consistency of notation with what follows.

It should be noted that the idea of combining rules for the multisolute osmotic virial equation is not new. To our knowledge, the first such concept was put forward by Elliott *et al.*[23] who derived the arithmetic-average combining rule for the second-order mixed coefficient in the regular solution theory framework. A geometric-average combining rule was also proposed for the cubic mixed coefficient in the same work. These combining rules were then used to predict the osmolality of solutions of interest in biology.[23–26] Later, a modified version of the multisolute osmotic virial equation with the same combining rules was applied to solutions containing salts as well.[25–27] In all cases, good agreement was achieved between the predictions of the model and independent experimental measurements. The present work aims to derive generalized combining rules in a specific solution theory framework.

This paper is intended as a standalone work on the theoretical exploration of the multisolute osmotic virial equation and corresponding combining rules. We note that, however, we have applied an extension of this model to dissociating solutes in another paper.[28] There, the predictions of the present model for freezing points of 11 ternary inorganic salt solutions of water were found to be in excellent agreement with independent experimental measurements (root-mean square error of 0.45 K and close-to-zero mean bias for a total of 371 data points).

II. THE MODEL

A. Assumptions of the model

We consider a single-phase multicomponent mixture theory that accounts for interactions in arbitrarily large particle groups. The goal here is to find an explicit formula for the thermodynamic potential (or free energy) of the system in terms of the energies of the particle groups. We are interested in solutions formed at constant temperature and pressure with fixed amounts of each component, hence, the Gibbs free energy, G, is the thermodynamic potential of the system. The assumptions of the model are listed below:

(i) The total nonconfigurational Gibbs free energy of the system can be written as the sum of the interaction energies of all particle groups of a chosen size. The selection of this size is based on assumption (ii).

(ii) For each particle, there exists a neighborhood, and the interactions of this particle with particles outside of this neighborhood can be neglected. The neighborhood is defined by the smallest spherical volume encapsulating a given number of particles. Note that negligible interaction outside of this sphere does not mean zero interaction, but instead, implies that the difference in the long-range interactions between the pure substance and the mixture is negligible such that the associated energy remains unchanged upon mixing.

(iii) Each particle is directly surrounded by a certain constant number of other particles, which is called the coordination number. The coordination number is well defined for a solid lattice but should be thought of as an average of many possible arrangements for liquids. The fluctuations around this average can be shown to have insignificant effect on the local geometry in liquids. No assignment of this constant is needed in our calculations.

(iv) All particles are of similar size so that, when a particle at a site is replaced by a different type of particle, the coordination number can be assumed to be unchanged.

(v) The particles are randomly distributed even though the exchange interaction energies are not zero. In other words, all particles have an equal probability of occupying any given site, which is also the same for all sites, independent of the occupancy of the neighboring sites.

Assumption (v) is also known as the Braggs–Williams approximation,[2] and it is the main simplification made in the present treatment.

B. Nonconfigurational Gibbs free energy of random mixing

Here, it is convenient to work with a modified Gibbs free energy function that excludes the terms due to configurational entropy. This contribution can be separately calculated and added later. Consider n-tuples as the interacting particle groups (i.e., groups of n particles; smaller or larger groups are not used in the formulation) at constant pressure and temperature. Then, based on assumptions (i) and (ii), the nonconfigurational Gibbs free energy, G^{nc}, of a system made by mixing r types of components can be expressed as the sum of energies of all particle groups:

$$G^{\mathrm{nc}} = \sum_{\ell} g_{\mathbf{A}_\ell} N_{\mathbf{A}_\ell} \tag{1}$$

where $g_{\mathbf{A}_\ell}$ and $N_{\mathbf{A}_\ell}$ are the nonconfigurational Gibbs free energy and the number of a certain type of fixed composition n-tuples denoted by the index $\mathbf{A}_\ell$, respectively. The subscript ℓ is an incrementing label pointing to each unique element of $\mathbf{A}$ (see the next paragraph). Each $g_{\mathbf{A}_\ell}$ is assumed to be averaged over all possible spatial configurations of particle groups having the same composition. The main idea of the present treatment is to not search for a decomposition of $g_{\mathbf{A}_\ell}$ in terms of the energies of smaller particle groups if not necessary, and instead, treat it as a fundamental quantity. This approach reduces mathematical complexity and avoids needless approximations that may be present in other approaches, such as the quasi-chemical treatment, where the pairwise additivity of interaction energies had to be assumed. Note that this does not mean that the method presented here is superior to others. The choice is rather a matter of what one wishes to calculate/achieve using the model. For the purposes of this article, viewing the n-tuple energy as irreducible will suffice.

In Eq. (1), $\mathbf{A}$ denotes the set of all multisets of length n, chosen from r types (i.e., all the different n-tuples that can be made of n particles chosen from r particle types or components), and ℓ is used to label its elements. One such element represents a unique n-tuple type (and vice versa) in terms of its composition, and each $\mathbf{A}_\ell$ is permutation-invariant (i.e., no ordering is assumed). From basic combinatorics, the number of such multisets (i.e., the number of elements of $\mathbf{A}$) can be found to be[29]

$$|\mathbf{A}| = \binom{n+r-1}{n} = \frac{(n+r-1)!}{n!(r-1)!} \tag{2}$$

meaning that $\ell \in [1, |\mathbf{A}|]$ (note that $0! = 1$). We use natural numbers from 1 to r to label the components. For example, for $n = 2$ and $r = 3$ we have

$$G^{\text{nc}} = g_{11}N_{11} + g_{22}N_{22} + g_{33}N_{33} + g_{12}N_{12} + g_{13}N_{13} + g_{23}N_{23} \tag{3}$$

for $n = 3$ and $r = 2$ we have

$$G^{\text{nc}} = g_{111}N_{111} + g_{222}N_{222} + g_{112}N_{112} + g_{122}N_{122} \tag{4}$$

and so on, as desired. The complicated index notation we adopt here may seem unnecessary, however, its use has the benefit of making certain formulae in the following derivation much shorter and easier to comprehend.

Using assumptions (iii) and (iv) and taking z_n as the number of n-tuples sharing each particle, which only depends on n and the coordination number, the number of particles of type j and the number of n-tuples can be related by simple counting as follows:

$$N_j = \frac{1}{z_n}\sum_{\ell} \xi_{j|\mathbf{A}_\ell} N_{\mathbf{A}_\ell} \tag{5}$$

where $0 \le \xi_{j|\mathbf{A}_\ell} \le n$ is the multiplicity of component j in the multiset $\mathbf{A}_\ell$. Noting that $\sum_{j=1}^{r} \xi_{j|\mathbf{A}_\ell} = n$ for any $\mathbf{A}_\ell$, we can sum the left- and right-hand sides of Eq. (5) over j and obtain

$$\sum_{\ell} N_{\mathbf{A}_\ell} = \frac{z_n}{n}\sum_{j=1}^{r} N_j \tag{6}$$

which relates the total number of n-tuples to the total number of particles of individual species. As it can be inferred from Eq. (6), the factor z_n/n is the number of times we overcount each particle when counting the particle groups. If we multiply both sides of Eq. (5) by the pure n-tuple energy, $g_{j...j}$, before summing its left- and right-hand sides over j, we instead get

$$\frac{z_n}{n}\sum_{j=1}^{r} g_{j\ldots j}N_j - \sum_{\ell}\left(\frac{1}{n}\sum_{j=1}^{r} \xi_{j|\mathbf{A}_\ell} g_{j\ldots j}\right)N_{\mathbf{A}_\ell} = 0 \tag{7}$$

Equation (7) can be added to Eq. (1) and the result can be rearranged to give

$$G^{\text{nc}} = \frac{z_n}{n}\sum_{j=1}^{r} g_{j\ldots j}N_j + \sum_{\ell}\left(g_{\mathbf{A}_\ell} - \frac{1}{n}\sum_{j=1}^{r} \xi_{j|\mathbf{A}_\ell} g_{j\ldots j}\right)N_{\mathbf{A}_\ell} \tag{8}$$

Notice that the first term on the right-hand side of Eq. (8) is the nonconfigurational Gibbs free energy of the system when only the like particles interact. It has the same value as if we calculated the total system energy with each pure-component subsystem existing far apart.

From the second term on the right-hand side of Eq. (8), we can naturally define the nonconfigurational Gibbs free energy change of formation of a particle group by viewing the mixing process as the creation of a particular n-tuple from the "disassembly" and "reassembly" of the appropriate pure-component n-tuples:

$$w_{\mathbf{A}_\ell} = g_{\mathbf{A}_\ell} - \frac{1}{n}\sum_{j=1}^{r} \xi_{j|\mathbf{A}_\ell} g_{j\ldots j} \tag{9}$$

Clearly, $w_{j...j} = 0$ for all j, however, we will not make this substitution in some cases to preserve the symmetry of certain expressions. With this notation, the nonconfigurational Gibbs free energy of the system can be expressed as

$$G^{\text{nc}} = \frac{z_n}{n}\sum_{j=1}^{r} g_{j\ldots j}N_j + \sum_{\ell} w_{\mathbf{A}_\ell} N_{\mathbf{A}_\ell} \tag{10}$$

Based on the random-mixing assumption, (v), the probability of a randomly chosen n-tuple having the composition $\mathbf{A}_\ell$ can be found with the help of combinatorics:

$$\mathcal{P}_{\mathbf{A}_\ell} = \frac{n!}{\left(\sum_{j=1}^{r} N_j\right)^{n}} \prod_{j=1}^{r} \frac{(N_j)^{\xi_{j|\mathbf{A}_\ell}}}{\xi_{j|\mathbf{A}_\ell}!} \tag{11}$$

which is the probability mass function of the multinomial distribution.[29] Here, it is assumed that each N_j is much larger than n so that the probability of each consecutive inclusion of a certain type of species in the group is independent of its existing fraction in that group. Using Eqs. (6) and (11), we can calculate the expected or average number of n-tuples of a given composition in the mixture as

$$\langle N_{\mathbf{A}_\ell} \rangle = \mathcal{P}_{\mathbf{A}_\ell}\sum_{\ell} N_{\mathbf{A}_\ell} = \frac{z_n n!}{n\left(\sum_{j=1}^{r} N_j\right)^{n-1}} \prod_{j=1}^{r} \frac{(N_j)^{\xi_{j|\mathbf{A}_\ell}}}{\xi_{j|\mathbf{A}_\ell}!} \tag{12}$$

Substituting Eq. (12) into Eq. (10), we can find the expected nonconfigurational Gibbs free energy of the entire system:

$$\langle G^{\text{nc}}\rangle = \frac{z_n}{n}\sum_{j=1}^{r} g_{j\ldots j}N_j + \frac{z_n}{n\left(\sum_{j=1}^{r}N_j\right)^{n-1}} \times \sum_{\ell} w_{\mathbf{A}_\ell}\, n!\prod_{j=1}^{r}\frac{(N_j)^{\xi_{j|\mathbf{A}_\ell}}}{\xi_{j|\mathbf{A}_\ell}!} \tag{13}$$

Used here to distinguish the exact value from the expected value, we drop the angle brackets in the remainder of the text. Applying the multinomial theorem to Eq. (13), we can notationally simplify and rewrite it as

$$G^{\text{nc}} = \frac{z_n}{n}\sum_{j=1}^{r} g_{j\ldots j}N_j + \frac{z_n}{n\left(\sum_{j=1}^{r}N_j\right)^{n-1}} \times \sum_{i_1,i_2\ldots i_n=1}^{r} w_{i_1 i_2\ldots i_n}N_{i_1}N_{i_2}\ldots N_{i_n} \tag{14}$$

where we use a single sum notation for a series of sums repeated n times for each element (i.e., each index) from the set of variables $\{i_1, i_2\ldots i_n\}$, each starting from 1 going up to r. Since the individual indices appear explicitly and the second sum is symmetric in Eq. (14), this explicit notation will prove more useful in the remainder of the text compared to the notation used in Eqs. (1)–(13). Note that the choice to label the variable indices in an increasing order is a convention we make here and follow throughout the text.

Equation (13) has been previously obtained by Saulov[21] for the general case and by Kakuda *et al.*[22] for up to $n = 4$ and $r = 4$ using similar arguments and assumptions. Both articles have received minimal attention in the solution thermodynamics literature and are not particularly well known.

C. Chemical potential

To find the chemical potential of component j, μ_j, we take the derivative of the total Gibbs free energy with respect to the number of that component, N_j, at constant temperature, T, pressure, P, and the fixed numbers of the remaining species:

$$\mu_j = \left(\frac{\partial G}{\partial N_j}\right)_{T,P,N_i\neq N_j} = \left(\frac{\partial G^{\text{nc}}}{\partial N_j}\right)_{T,P,N_i\neq N_j} + \left(\frac{\partial G^{\text{c}}}{\partial N_j}\right)_{T,P,N_i\neq N_j} = \mu_j^{\circ} + \mu_j^{\text{ex}} + k_{\text{B}}T\ln x_j \tag{15}$$

where G^{c} is the configurational part of the Gibbs free energy ($G = G^{\text{nc}} + G^{\text{c}}$), $\mu_j^{\circ} = z_n g_{j\ldots j}/n$ is the chemical potential of the pure substance (by the definition of chemical potential), k_{B} is the Boltzmann constant, and x_j is the mole fraction of component j in the solution. Termed the excess chemical potential of species j, μ_j^{ex} is the derivative of the second term on the right-hand side of Eq. (14). The term $k_{\text{B}}T\ln x_j$ in Eq. (15) is the derivative of G^{c} under the random-mixing assumption and its derivation can be found in any standard textbook on the topic (e.g., see Guggenheim[1]).

Next, we explicitly evaluate μ_j^{ex} in Eq. (15) using the expression through which it is defined:

$$\begin{aligned}\mu_j^{\text{ex}} &= \frac{\partial}{\partial N_j}\left[\frac{z_n}{n\left(\sum_{k=1}^{r}N_k\right)^{n-1}}\sum_{i_1,i_2\ldots i_n=1}^{r} w_{i_1 i_2\ldots i_n}N_{i_1}N_{i_2}\ldots N_{i_n}\right]\\ &= \frac{z_n}{nN_j\left(\sum_{k=1}^{r}N_k\right)^{n-1}}\sum_{i_1,i_2\ldots i_n=1}^{r}\left(\sum_{k=1}^{n}\delta_{ji_k}\right)w_{i_1 i_2\ldots i_n}\\ &\quad\times N_{i_1}N_{i_2}\ldots N_{i_n} - \frac{z_n(n-1)}{n\left(\sum_{k=1}^{r}N_k\right)^{n}}\\ &\quad\times\sum_{i_1,i_2\ldots i_n=1}^{r} w_{i_1 i_2\ldots i_n}N_{i_1}N_{i_2}\ldots N_{i_n}\end{aligned} \tag{16}$$

where δ_{ij} is the Kronecker delta ($\delta_{ij} = 1$, if $i = j$; $\delta_{ij} = 0$, otherwise). We can expand the innermost sum in Eq. (16), reindex each resulting term, and recollect the terms to get

$$\begin{aligned}\mu_j^{\text{ex}} &= \frac{z_n}{nN_j\left(\sum_{k=1}^{r}N_k\right)^{n-1}}\left(n\sum_{i_1,i_2\ldots i_n=1}^{r}\delta_{ji_1}w_{i_1 i_2\ldots i_n}N_{i_1}N_{i_2}\ldots N_{i_n}\right)\\ &\quad - \frac{z_n(n-1)}{n\left(\sum_{k=1}^{r}N_k\right)^{n}}\sum_{i_1,i_2\ldots i_n=1}^{r} w_{i_1 i_2\ldots i_n}N_{i_1}N_{i_2}\ldots N_{i_n}\end{aligned} \tag{17}$$

Cancelling out the terms in Eq. (17) and writing it in mole fraction units, we have

$$\begin{aligned}\mu_j^{\text{ex}} &= z_n\sum_{i_2,i_3\ldots i_n=1}^{r} w_{ji_2\ldots i_n}x_{i_2}x_{i_3}\ldots x_{i_n}\\ &\quad - \frac{z_n(n-1)}{n}\sum_{i_1,i_2\ldots i_n=1}^{r} w_{i_1 i_2\ldots i_n}x_{i_1}x_{i_2}\ldots x_{i_n}\end{aligned} \tag{18}$$

Equation (18) is the most general form of the excess chemical potential given the current assumptions. It can be substituted into Eq. (15) giving the equation for the full chemical potential. However, in the next few steps of the derivation, it will be more convenient to work with the excess chemical potential, thus we choose to make this substitution later.

III. SOLVENT-DOMINANT MIXTURES

In many practical applications, multicomponent mixtures are often those consisting of a solvent as the main medium with added solutes in small amounts (not necessarily dilute). Therefore, it is valuable to have a method of estimating the chemical potential of the solvent from binary solution data only. Here, with one additional assumption about the exchange interaction energies of the components, we provide a derivation of such a method based on Eq. (18).

In the remainder of this paper, we will denote the solvent by 1 and the solutes by $2, 3, \ldots, r$.

A. Asymmetric form of Eq. (18) for the solvent

Here, we work with Eq. (18) written for the solvent (i.e., $j = 1$). It is useful to write this equation in terms of the mole fractions of the solutes only. This can be done by repeatedly substituting $x_1 = 1 - \sum_{j=2}^{r} x_j$ in the expansion over each i_j and simplifying the results by combining the like terms. Below we give the details of substitution for x_1 in the expansion over i_1:

Step (1). Separate the term containing x_1 in the sum over i_1 in Eq. (18) from the rest and substitute for x_1 to get

$$\begin{aligned}\mu_1^{\text{ex}} = z_n &\sum_{i_2,i_3\ldots i_n=1}^{r} w_{1i_2\ldots i_n} x_{i_2} x_{i_3}\ldots x_{i_n} - \frac{z_n(n-1)}{n}\\ &\times \sum_{i_2,i_3\ldots i_n=1}^{r} w_{1i_2\ldots i_n}\left(1 - \sum_{j=2}^{r} x_j\right) x_{i_2} x_{i_3}\ldots x_{i_n}\\ &- \frac{z_n(n-1)}{n} \sum_{\substack{i_2,i_3\ldots i_n=1\\ i_1=2}}^{r} w_{i_1 i_2\ldots i_n} x_{i_1} x_{i_2}\ldots x_{i_n}\end{aligned} \tag{19}$$

Step (2). Split the expression $(1 - \sum_{j=2}^{r} x_j)$ in Eq. (19) into two terms to get

$$\begin{aligned}\mu_1^{\text{ex}} = z_n &\sum_{i_2,i_3\ldots i_n=1}^{r} w_{1i_2\ldots i_n} x_{i_2} x_{i_3}\ldots x_{i_n}\\ &- \frac{z_n(n-1)}{n} \sum_{i_2,i_3\ldots i_n=1}^{r} w_{1i_2\ldots i_n} x_{i_2} x_{i_3}\ldots x_{i_n}\\ &+ \frac{z_n(n-1)}{n} \sum_{\substack{i_2,i_3\ldots i_n=1\\ i_1=2}}^{r} w_{1i_2\ldots i_n} x_{i_1} x_{i_2}\ldots x_{i_n}\\ &- \frac{z_n(n-1)}{n} \sum_{\substack{i_2,i_3\ldots i_n=1\\ i_1=2}}^{r} w_{i_1 i_2\ldots i_n} x_{i_1} x_{i_3}\ldots x_{i_n}\end{aligned} \tag{20}$$

Step (3). Collect the terms over the same sums to get

$$\begin{aligned}\mu_1^{\text{ex}} = \frac{z_n}{n} &\sum_{i_2,i_3\ldots i_n=1}^{r} w_{1i_2\ldots i_n} x_{i_2} x_{i_3}\ldots x_{i_n} - \frac{z_n(n-1)}{n}\\ &\times \sum_{\substack{i_2,i_3\ldots i_n=1\\ i_1=2}}^{r} (w_{i_1 i_2\ldots i_n} - w_{1i_2\ldots i_n}) x_{i_1} x_{i_2}\ldots x_{i_n}\end{aligned} \tag{21}$$

If we now substitute $x_1 = 1 - \sum_{j=2}^{r} x_j$ in the expansion over i_2 in Eq. (21) and follow the same steps (see the supplementary material), we get

$$\begin{aligned}\mu_1^{\text{ex}} = \frac{z_n}{n} &\sum_{i_3,i_4\ldots i_n=1}^{r} w_{11i_3\ldots i_n} x_{i_3} x_{i_4}\ldots x_{i_n} - \frac{z_n(n-2)}{n}\\ &\times \sum_{\substack{i_3\ldots i_n=1\\ i_2=2}}^{r} (w_{1i_2\ldots i_n} - w_{11i_3\ldots i_n}) x_{i_2} x_{i_3}\ldots x_{i_n}\\ &- \frac{z_n(n-1)}{n} \sum_{\substack{i_3,i_4\ldots i_n=1\\ i_1,i_2=2}}^{r} (w_{i_1 i_2\ldots i_n} - 2w_{1i_2\ldots i_n}\\ &+ w_{11i_3\ldots i_n}) x_{i_1} x_{i_2}\ldots x_{i_n}\end{aligned} \tag{22}$$

Similarly, if substitute $x_1 = 1 - \sum_{j=2}^{r} x_j$ in the expansion over i_3 in Eq. (22) (see the supplementary material), we get

$$\begin{aligned}\mu_1^{\text{ex}} = \frac{z_n}{n} &\sum_{i_4,i_5\ldots i_n=1}^{r} w_{111i_4\ldots i_n} x_{i_4} x_{i_5}\ldots x_{i_n} - \frac{z_n(n-3)}{n}\\ &\times \sum_{\substack{i_4\ldots i_n=1\\ i_3=2}}^{r} (w_{11i_3\ldots i_n} - w_{111i_4\ldots i_n}) x_{i_3} x_{i_4}\ldots x_{i_n}\\ &- \frac{z_n(2n-3)}{n} \sum_{\substack{i_4,i_5\ldots i_n=1\\ i_2,i_3=2}}^{r} (w_{1i_2\ldots i_n} - 2w_{11i_3\ldots i_n}\\ &+ w_{111i_4\ldots i_n}) x_{i_2} x_{i_3}\ldots x_{i_n} - \frac{z_n(n-1)}{n}\\ &\times \sum_{\substack{i_4,i_5\ldots i_n=1\\ i_1,i_2,i_3=2}}^{r} (w_{i_1 i_2\ldots i_n} - 3w_{1i_2\ldots i_n} + 3w_{11i_3\ldots i_n}\\ &- w_{111i_4\ldots i_n}) x_{i_1} x_{i_2}\ldots x_{i_n}\end{aligned} \tag{23}$$

Noticing the emerging pattern in this series, the result after n steps can be cast into the following compact expression:

$$\begin{aligned}\mu_1^{\text{ex}} = &\sum_{m=2}^{n} \frac{z_n(1-m)}{n}\binom{n}{m}\\ &\times \sum_{i_1,i_2\ldots i_m=2}^{r}\left[\sum_{k=0}^{m}(-1)^k\binom{m}{k} w_{1\ldots 1 i_{k+1}\ldots i_m}\right] x_{i_1} x_{i_2}\ldots x_{i_m}\end{aligned} \tag{24}$$

where $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ and $\binom{m}{k} = \frac{m!}{k!(m-k)!}$ are the binomial coefficients. The number of 1's in the subscript of w is equal to $n - m + k$ while the remaining $m - k$ number of indices stay free. Note that more than one choice exists to label these free indices when $k \geq 1$; the choice from $k+1$ up to m is arbitrary. Equation (24) is an asymmetric version of Eq. (18) for the excess chemical potential of the solvent. It is a multivariate polynomial containing all possible configurations of powers from 2 to n of solute mole fractions only ($r-1$ variables), compared to powers of $n-1$ and n in Eq. (18), which also includes the solvent mole fraction as a variable (r variables). Below we give examples of each element in the main sum in Eq. (24) for $n = 4$:

for $n = 4$ and $m = 4$, we have

$$\begin{aligned}
&-\frac{3z_n}{4}\sum_{i_1,i_2,i_3,i_4=2}^{r}\left[\sum_{k=0}^{4}(-1)^k\binom{4}{k}w_{1\ldots 1 i_{k+1}\ldots i_4}\right]x_{i_1}x_{i_2}x_{i_3}x_{i_4}\\
&\quad= -\frac{3z_n}{4}\sum_{i_1,i_2,i_3,i_4=2}^{r}\big(w_{i_1i_2i_3i_4} - 4w_{1i_2i_3i_4} + 6w_{11i_3i_4}\\
&\qquad - 4w_{111i_4} + w_{1111}\big)x_{i_1}x_{i_2}x_{i_3}x_{i_4}
\end{aligned} \tag{25}$$

for $n = 4$ and $m = 3$, we have

$$\begin{aligned}
&-2z_n\sum_{i_1,i_2,i_3=2}^{r}\left[\sum_{k=0}^{3}(-1)^k\binom{3}{k}w_{1\ldots 1 i_{k+1}\ldots i_3}\right]x_{i_1}x_{i_2}x_{i_3}\\
&\quad= -2z_n\sum_{i_1,i_2,i_3=2}^{r}\big(w_{1i_1i_2i_3} - 3w_{11i_2i_3} + 3w_{111i_3} - w_{1111}\big)x_{i_1}x_{i_2}x_{i_3}
\end{aligned} \tag{26}$$

for $n = 4$ and $m = 2$, we have

$$\begin{aligned}
&-\frac{3z_n}{2}\sum_{i_1,i_2=2}^{r}\left[\sum_{k=0}^{2}(-1)^k\binom{2}{k}w_{1\ldots 1 i_{k+1}\ldots i_2}\right]x_{i_1}x_{i_2}\\
&\quad= -\frac{3z_n}{2}\sum_{i_1,i_2=2}^{r}\big(w_{11i_1i_2} - 2w_{111i_2} + w_{1111}\big)x_{i_1}x_{i_2}
\end{aligned} \tag{27}$$

B. Derivation of combining rules

If we hope to obtain any type of combining rules (i.e., a model for a multisolute mixture that uses only single-solute–solvent binary information), we need to express the coefficients of the mixed terms (i.e., the terms containing the powers of more than one of $\{x_2, x_3, \ldots, x_r\}$) of any order in terms of the coefficients of the pure terms (i.e., the terms containing the powers of only one of $\{x_2, x_3, \ldots, x_r\}$).

Note that Eq. (24), written for a multicomponent mixture, inevitably contains terms with coefficients corresponding to purely solute–solute interactions (e.g., terms with w_{223}, w_{2345}, etc.). Clearly, these coefficients are not expressible in terms of the solvent–solute binary coefficients, even with any additional (physically meaningful) approximations. Therefore, to continue further, we make the following simplifying assumption:

(vi) The interaction exchange energies of the solute species with one another are negligible compared to the interaction exchange energies of the solute species and the solvent species. That is, the nonconfigurational Gibbs free energy change of formation of a particle group made purely from solutes is zero. This is equivalent of assuming that a mixture made of any combination of the solute components only behaves ideally.

Under assumption (vi), if one type of solute species is replaced by another type of solute species in a certain particle group, the change of the interaction energy of this group does not depend on the types of the other solute species present in the group: it only depends on the number of solvent particles in the group and the types of particles being exchanged. It is important to note that no restriction is imposed on the concentrations of the species.

We can arrive at the mathematical implications of assumption (vi) with the following thought experiment: Consider some number of copies of the same n-tuple containing at least one solvent particle, that is, consider all copies having the same composition as one system. If we swap one solute particle of a certain type from one of the n-tuples in this system with another solute particle of a different type from a different n-tuple in this system, then the nonconfigurational Gibbs free energy of the entire system will not change because of assumption (vi). Clearly, we can repeat this particle swapping any number of times without affecting the total (nonconfigurational) energy of the system. Now, to ensure that we can reduce this system of copies to a sum of n-tuples each containing the same number of solvent particles and only one type of solute particles (i.e., binary groups), we can make the initial system such that the number of the copies is equal to the number of the solute particles in one n-tuple. Consistent with the notation in Eq. (24), we can mathematically summarize this realization as a decomposition of g in the following way:

$$g_{1\ldots 1 i_{k+1}\ldots i_m} = \frac{1}{m-k}\sum_{j=k+1}^{m} g_{1\ldots 1 i_j\ldots i_j} \tag{28}$$

where the number of 1's and the number of the variable indices in the subscript of g are preserved from the left side of the equation to the right side of the equation. Like in Eq. (24), each variable index here can only represent a solute: $i_j \in [2, r]$. The case $m = k$ corresponds to the pure n-tuple energy, so no decomposition exists in this case, and we do not need to worry about the division by zero.

Using Eq. (9), it can be checked that this representation of g in terms of the binary group energies implies the same decomposition for w (see the supplementary material). That is, we can write

$$w_{1\ldots 1 i_{k+1}\ldots i_m} = \frac{1}{m-k}\sum_{j=k+1}^{m} w_{1\ldots 1 i_j\ldots i_j} \tag{29}$$

As an example, the following series of equalities holds for $n = 4$, $m = 4$, and $k = 2$:

$$\begin{cases}
w_{1123} = \frac{1}{2}(w_{1122} + w_{1133})\\
w_{1134} = \frac{1}{2}(w_{1133} + w_{1144})\\
w_{1124} = \frac{1}{2}(w_{1122} + w_{1144})\\
w_{1125} = \frac{1}{2}(w_{1122} + w_{1155})\\
\vdots
\end{cases} \tag{30}$$

Substituting Eq. (29) into Eq. (24), after some reindexing and regrouping (see the supplementary material), the equation for the excess chemical potential can be expressed as

$$\mu_1^{\mathrm{ex}} = \sum_{m=1}^{n} \sum_{i_1,i_2\ldots i_m=2}^{r} \frac{1}{m}\left(\sum_{j=1}^{m} Q_{i_j}^{n,m}\right) x_{i_1} x_{i_2} \ldots x_{i_m} \tag{31}$$

where

$$Q_j^{n,m} = \frac{z_n(1-m)}{n}\binom{n}{m}\sum_{k=0}^{m}(-1)^k\binom{m}{k} w_{1\ldots 1j\ldots j} \tag{32}$$

$Q_j^{n,m}$ is the mth order pure coefficient of the excess chemical potential expression of solute j in a degree-n expansion, and it can be obtained by setting $x_i = 0$ for all i except $i = j$ in Eq. (24). Note that the lower bound of the main sum in Eq. (31) is switched from $m = 2$ to $m = 1$ without affecting the result because the first-order pure coefficients are zero by definition (i.e., $Q_j^{n,1} = 0$ for any n).

In Eq. (32), the number of 1's and j's in the subscript of w changes from one term of the sum to another as k varies. For instance, the fourth- and third-order pure coefficients of component 2 in a degree-4 expansion are

$$\begin{aligned} Q_2^{4,4} = &-\frac{3z_n}{4} w_{2222} + 3z_n w_{1222} - \frac{9z_n}{2} w_{1122} \\ &+ 3z_n w_{1112} - \frac{3z_n}{4} w_{1111} \end{aligned} \tag{33}$$

and

$$Q_2^{4,3} = -2z_n w_{1222} + 6z_n w_{1122} - 6z_n w_{1112} + 2z_n w_{1111} \tag{34}$$

respectively.

Substituting the expression for μ_1^{ex} given by Eq. (31) into Eq. (15) (written for the solvent), we can calculate μ_1. However, if we want to have a polynomial expression for μ_1 as well, we need to approximate $\ln(x_1)$ [which will be present in Eq. (15) when it is written for the solvent] by a suitable polynomial expression first. This can be done by substituting $x_1 = 1 - \sum_{j=2}^{r} x_j$ into $\ln(x_1)$ and writing its Taylor series expansion around $\sum_{j=2}^{r} x_j = 0$ up to the nth term:

$$\begin{aligned} \ln(x_1) = \ln\left(1 - \sum_{j=2}^{r} x_j\right) &\approx -\sum_{j=2}^{r} x_j - \frac{1}{2}\left(\sum_{j=2}^{r} x_j\right)^2 \\ &- \frac{1}{3}\left(\sum_{j=2}^{r} x_j\right)^3 - \cdots - \frac{1}{n}\left(\sum_{j=2}^{r} x_j\right)^n \\ &= -\sum_{i_1=2}^{r} x_{i_1} - \sum_{i_1,i_2=2}^{r} \frac{1}{2} x_{i_1} x_{i_2} - \sum_{i_1,i_2,i_3=2}^{r} \frac{1}{3} x_{i_1} x_{i_2} x_{i_3} \\ &- \cdots - \sum_{i_1,i_2\ldots i_n=2}^{r} \frac{1}{n} x_{i_1} x_{i_2} \ldots x_{i_n} \\ &= -\sum_{m=1}^{n} \sum_{i_1,i_2\ldots i_m=2}^{r} \frac{1}{m} x_{i_1} x_{i_2} \ldots x_{i_m} \end{aligned} \tag{35}$$

Chemical potential is more useful in practice when it is calculated with respect to a reference state. For a solvent with added solutes, one convenient reference state is the pure solvent system at the same temperature and pressure for which the chemical potential is μ_1°. It is also convenient to express this change in chemical potential in units of $k_B T$ to have a nondimensional equation. Thus, substituting the results from Eq. (31) and Eq. (35) into Eq. (15) (written for the solvent) and rearranging, we write the following:

$$\begin{aligned} \frac{\mu_1 - \mu_1^\circ}{k_B T} = &-\sum_{m=1}^{n} \sum_{i_1,i_2\ldots i_m=2}^{r} \frac{1}{m}\left[1 - \frac{1}{k_B T}\left(\sum_{j=1}^{m} Q_{i_j}^{n,m}\right)\right] \\ &\times x_{i_1} x_{i_2} \ldots x_{i_m} \end{aligned} \tag{36}$$

Because the quantity $(\mu_1 - \mu_1^\circ)/k_B T$ appears frequently in calculations, it has been given different names in various forms. Some notable ones are the relative activity[1] of the solvent, defined as

$$a_1 = \exp\left(\frac{\mu_1 - \mu_1^\circ}{k_B T}\right) \tag{37}$$

and the osmolality[25,30] of the solvent, defined as

$$\pi = -\frac{\mu_1 - \mu_1^\circ}{k_B T M_1} \tag{38}$$

where M_1 is the solvent's molar mass. A more directly related quantity is given as

$$\pi^+ = -\frac{\mu_1 - \mu_1^\circ}{k_B T} \tag{39}$$

which is usually referred to as the osmole fraction[25,30] of the solvent. Because of this simple relationship, we choose to express the change in chemical potential as osmole fraction. Thus, after distributing the 1 inside the brackets evenly into m terms and introducing nondimensional coefficients, we can rewrite Eq. (36) as

$$\pi^+ = \sum_{m=1}^{n} \sum_{i_1,i_2\ldots i_m=2}^{r} \frac{1}{m}\left(\sum_{j=1}^{m} \hat{Q}_{i_j}^{n,m}\right) x_{i_1} x_{i_2} \ldots x_{i_m} \tag{40}$$

where $\hat{Q}_j^{n,m}$ is defined in terms of $Q_j^{n,m}$ through the relation

$$\begin{aligned} \hat{Q}_j^{n,m} &= \frac{1}{m} - \frac{Q_j^{n,m}}{k_B T} \\ &= \frac{1}{m} + \frac{z_n(m-1)}{n k_B T}\binom{n}{m}\sum_{k=0}^{m}(-1)^k\binom{m}{k} w_{1\ldots 1j\ldots j} \end{aligned} \tag{41}$$

and it can be described as the mole-fraction-based mth order pure osmotic virial coefficient of component j in a degree-n expansion. Equation (40) is a multivariate polynomial in $r - 1$ variables for the change of the chemical potential of the solvent in the presence of solutes. It is a form of the multisolute osmotic virial equation derived using assumptions (i) through (vi). The physical meaning of each coefficient of the polynomial can be inferred with the help of Eqs. (41) and (9). One can also obtain[2,4] other thermodynamic properties of the solvent, such as its molar volume, heat content, etc., by suitable differentiations of π^+, which are not discussed here.

The arithmetic-average combining rules for the mixed-term coefficients follow naturally from the structure of Eq. (40). The usefulness of such mixing rules is obvious: if one has data for the chemical potential change of the solvent as a function of solute concentration with each solute separately, then the chemical potential of the solvent in the presence of any combination of these solutes can be predicted. In practice, this can be done by fitting a polynomial to each binary solution dataset and taking the arithmetic average of the appropriate coefficients to get the coefficients of the cross terms. Alternatively, if no experimental data are available for certain binaries, the pure osmotic virial coefficients can be found from a more sophisticated model or molecular dynamics simulations.

The combining rules derived above based on assumption (vi) neglect effects expected when the solutes are ions (e.g., solute dissociation, charge screening). However, previous versions of the multisolute osmotic virial equation have been successfully adapted for use with salt solutes by incorporation of a single additional parameter, $(k_{\rm d})_j$ for each salt j, that is obtained empirically from fitting to experimental data [for each salt, replace x_j with $(k_{\rm d})_j x_j$ in the virial equation].[27] The parameter $(k_{\rm d})_j$ is called the "dissociation constant of solute j" even though it empirically accounts for more electrolyte effects than just dissociation. As noted earlier, in another paper, we have empirically incorporated the dissociation constant into the solution theory presented in this paper and investigated whether the multisolute osmotic virial equation with the new combining rules derived here can describe salt solutions.[28] There it was found that the predictions of the extended multisolute osmotic virial equation were accurate for 11 ternary aqueous salt solutions and that the new combining rules had superior performance to those developed previously by our group. In that work, the data for most single salt–water solutions were fitted well by only quadratic or cubic polynomials [i.e., requiring only $(k_{\rm d})_j$, $\hat{Q}_j^{n,2}$, and sometimes $\hat{Q}_j^{n,3}$ for each salt; see Sec. IV for interpretation of n].

IV. DISCUSSION

In this section, we only discuss the derived polynomial model written for a single-solute–solvent system and its coefficients. Therefore, the validity of assumption (vi) and the resulting combining rules presented in Sec. III are not required here.

A. Connection to the original single-solute osmotic virial coefficients

The single-solute osmotic virial equation was first derived by McMillan and Mayer[3] from first principles, which expresses the osmolality of a solution in terms of the molar concentration of a solute. Later, Hill[4] obtained the same result in molality and mole fraction units by choosing different sets of independent variables (i.e., different ensembles). Both methods are formally exact and necessarily equivalent through a suitable change of variables (only true if the virial expansions are kept as infinite sums; not true when they are truncated). Here, we show that one can obtain the pure osmotic virial coefficients derived in this work by applying the random-mixing assumption to an exact model.

Let us look at the quadratic and cubic coefficients of a two-component system, since these are the coefficients provided in Hill's original paper. The extension of the argument to higher-order coefficients is straight forward and will be omitted here. For a single-solute solution, the polynomial expression is simply

$$\pi^+ = x_2 + C_2 x_2^2 + C_3 x_2^3 \dots \tag{42}$$

and the corresponding second- and third-order coefficients from the work of Hill are

$$C_2 = N\left[\frac{1}{2} - \exp\left(-\frac{\omega_2}{k_{\rm B}T}\right)\right] \tag{43}$$

and

$$\begin{aligned} C_3 = {} & 4N^2\left[\frac{1}{2} - \exp\left(-\frac{\omega_2}{k_{\rm B}T}\right)\right]^2 \\ & - 2N^2\left[\frac{1}{3} - \exp\left(-\frac{\omega_2}{k_{\rm B}T}\right) + \exp\left(-\frac{\omega_3}{k_{\rm B}T}\right)\right] \end{aligned} \tag{44}$$

respectively, where $N = N_1 + N_2$ is the total number of particles in the system. We use here C's (as in the paper of Hill) instead of $\hat{Q}$'s since they refer to coefficients derived within different theories. The quantity ω_i in Eqs. (43) and (44) is defined as the total Gibbs free energy change for the following process: i number of systems, each made of $N-1$ solvent molecules and one solute molecule, being rearranged into one system made of $N-i$ solvent molecules with i number of solute molecules, plus $i-1$ number of pure solvent systems each with N solvent molecules. Approximating the entropy term that is present in ω_i with a random-mixing entropy term, the exact configurational entropy change for this process can be written explicitly, and ω_i can be expressed as

$$\omega_i = \omega_i^{\rm nc} - k_{\rm B}T\left[\ln\binom{N}{i} - i\ln\binom{N}{1}\right] \tag{45}$$

where $\omega_i^{\rm nc}$ is the nonconfigurational part of Gibbs free energy change, and the second term is due to the configurational entropy change for the process described above. Substituting Eq. (45) into Eqs. (43) and (44), we get

$$C_2 = N\left[\frac{1}{2} - \left(\frac{1}{2} - \frac{1}{2N}\right)\exp\left(-\frac{\omega_2^{\rm nc}}{k_{\rm B}T}\right)\right] \tag{46}$$

and

$$\begin{aligned} C_3 = {} & 4N^2\left[\frac{1}{2} - \left(\frac{1}{2} - \frac{1}{2N}\right)\exp\left(-\frac{\omega_2^{\rm nc}}{k_{\rm B}T}\right)\right]^2 \\ & - 2N^2\left[\frac{1}{3} - \left(\frac{1}{2} - \frac{1}{2N}\right)\exp\left(-\frac{\omega_2^{\rm nc}}{k_{\rm B}T}\right)\right. \\ & \left. + \left(\frac{1}{6} - \frac{1}{2N} + \frac{1}{3N^2}\right)\exp\left(-\frac{\omega_3^{\rm nc}}{k_{\rm B}T}\right)\right] \end{aligned} \tag{47}$$

Retaining only the constant and linear terms in the Taylor series expansion for the exponentials, these equations reduce to

$$C_2 = \frac{1}{2} + \frac{1}{Nk_{\rm B}T}\binom{N}{2}\omega_2^{\rm nc} \tag{48}$$

and

$$C_3 = \frac{1}{3} + \frac{2}{Nk_\mathrm{B}T}\binom{N}{3}\left(\omega_3^{\mathrm{nc}} - 3\omega_2^{\mathrm{nc}}\right) \tag{49}$$

respectively. We can express the ω_i^{nc}'s in terms of the alternative parameters defined in this work, $w_{1...12...2}$'s, using their respective definitions. Doing so, with some algebraic manipulation (see the supplementary material), we find

$$C_2 = \frac{1}{2} + \frac{1}{Nk_\mathrm{B}T}\binom{N}{2}\left(w_{1\ldots122} - 2w_{1\ldots12}\right) = \hat{Q}_2^{N,2} \tag{50}$$

and

$$C_3 = \frac{1}{3} + \frac{2}{Nk_\mathrm{B}T}\binom{N}{3}\left(w_{1\ldots1222} - 3w_{1\ldots122} + 3w_{1\ldots12}\right) = \hat{Q}_2^{N,3} \tag{51}$$

As indicated, Eqs. (50) and (51) are precisely what we can obtain from Eq. (41) by taking $n = N$, which translates to setting $z_n = 1$ (each particle is shared by a single N-tuple, which is the entire system). At first, it may seem problematic to set $n = N_1 + N_2$ because we had to assume that n is much smaller than both N_1 and N_2 to be able to write Eq. (11). However, this is allowed due to the different system definitions employed: while we fix both N_1 and N_2 and seek the chemical potentials, Hill fixes the total number of sites, N, and the chemical potential difference, $\mu_2 - \mu_1$, to arrive at his osmotic virial coefficients. In other words, we look at a large but finite closed system which can only exchange heat and volume with a reservoir (a piston–cylinder device), and Hill considers an open system with a fixed number of sites in equilibrium with an infinite reservoir, while effectively treating the system as a single interacting N-tuple.

B. On the size of *n*

Based on the connection established above, a natural question arises: how big is the error due to the linear approximations of the exponentials in the osmotic virial coefficients of Hill? To answer this, let us consider a macroscopic system corresponding to $N \to \infty$ limit in Eqs. (46) and (47). Since we expect the osmotic virial coefficients to stay finite at this limit for any physical system far from criticality, we deduce from Eqs. (46) and (47) that $\lim_{N\to\infty} \omega_2^{\mathrm{nc}} = \lim_{N\to\infty} \omega_3^{\mathrm{nc}} = 0$. This means that the larger the system the more accurate are the linear approximations. That is, with increasing n, the present treatment becomes asymptotically close to the exact treatment with a random-mixing entropy term, applied to a macroscopic system. This is because, even if one does not assume random mixing when calculating the nonconfigurational Gibbs free energy of the system (like in the exact treatment), the larger the particle groups the more closely their number density will follow the multinomial distribution. In other words, provided that the system's size is much larger than the range of interactions between particles, by choosing a large n the errors introduced due to Eq. (11) can be made vanishingly small. This realization is not surprising because the assumption of a single-phase system is the same as the homogeneity of the system on large size scales (compared to the range of interactions of particles). Therefore, when compared to the exact treatment, the only approximation in the present model with large n is due to the use of the random-mixing entropy term.

Another valuable consequence to consider when n is large is the potential for relaxation of assumptions (iii) and (iv). Consider, for example, a two-component solution of protein in water, where the sizes of the molecules, and hence, their coordination numbers differ greatly. Naturally, the present approach would not be applicable in the case of small n because z_n would not be constant between the two types of molecules. However, by choosing a large n such that each n-tuple occupies a much bigger space than the protein molecule, the discrepancy in z_n can be made small. Also, note that when $n = N$ the formulation is independent of z_n as demonstrated in Subsection IV A. Therefore, we expect the present model with sufficiently large n would be a good approximation even for solutions containing particles of different sizes.

In practice, if polynomials are fitted to binary experimental data, the obtained coefficients would represent full interactions as well as nonrandom entropy of mixing, as given in the framework of the exact formulation [e.g., Eqs. (43) and (44)], which would also include empirical corrections due to the truncation of the polynomials. Therefore, no deliberate selection of n is required in this case (assume $n = N$ if interpreting the coefficients in our framework with a random-mixing entropy term). Notice that the full model contains terms up to Nth degree, however, truncating the polynomial after the quadratic or cubic term typically describes the data well.[25,28] It is important to note that this truncation does not imply that only small particle groups are considered, rather it implies that the higher order terms of the polynomial are numerically negligible.

There might be situations, however, where it is desirable to pick a value of n that is as small as possible while still capturing the nonideal behavior of the system accurately (e.g., due to computational cost when calculating the virial coefficients from molecular dynamics simulations). This is the same as finding the smallest system from which the osmotic virial coefficients of the macroscopic system can be inferred to within a desired accuracy. The decision of picking a suitable n should be guided by the nature and range of the interactions in a given system while considering factors such as the desired accuracy and cost of computation. In general, a solvent with salts would require a larger n compared to the same solvent with nonionic solutes because of the presence of long-range forces in the former.

V. CONCLUSION

In this work, we considered a general multicomponent solution model that assumes random mixing of the components at constant pressure and temperature, but accounts for complex interactions in arbitrarily large particle groups. Based on the model, the chemical potential of the mixture components could be expressed as a multivariate polynomial with mole fractions of the solutes as variables. It was demonstrated that, for a macroscopic system, the present model with $n = N$ can be alternatively derived from the formally exact approach of Hill if the random-mixing entropy term is used. This was done to highlight the equivalency of two seemingly different theoretical approaches. It was also concluded that, if the model is implemented with large particle groups, it does not require some of the restrictive assumptions needed when considering small particle groups.

Furthermore, using the developed equations, we provided a theoretical basis for multisolute osmotic virial equation combining rules. That is, assuming negligible solute–solute interactions, the mixed-term coefficients of the polynomial could be written as the arithmetic average of suitable pure-term coefficients. These rules are practically useful for estimating the chemical potential of a solvent in the presence of multiple solutes when no multicomponent solution data are available. The model with the combining rules, in some sense, provides the best estimate of the solvent chemical potential within the context of the present theory when no multicomponent data are available. Of course, a better estimate would be possible if one has information about the interaction exchange energies of the solutes with one another and/or any multicomponent solution data. It is important to note that the non-interacting solutes assumption [i.e., assumption (vi)] and the use of the random-mixing entropy term were sufficient to arrive at the proposed combining rules, but we do not have a proof of their necessity. Comparison of the predictions of the combining-rules approach to experimental data from different kinds of multicomponent mixtures (e.g., electrolyte solutions, polymer solutions, hydrocarbon solutions, their combinations, etc.) might shed light on the extent of the applicability of the combining rules and help researchers decide whether it is worth looking for their generalizations.

SUPPLEMENTARY MATERIAL

See the supplementary material for the details of the omitted steps in the derivation.

ACKNOWLEDGMENTS

This research was funded by the Natural Sciences and Engineering Research Council of Canada (Grant No. RGPIN-2021-02773). H.B. acknowledges funding from Alberta Innovates Graduate Student Scholarship. J.A.W.E. holds a Canada Research Chair in Thermodynamics.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

All authors have contributed equally to this work.

Hikmat Binyaminov: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Janet A. W. Elliott**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

NOMENCLATURE

List of symbols

$\mathbf{A}$	set of all multisets each corresponding to a unique n-tuple composition
a_1	relative activity of the solvent
$\mathbf{A}_\ell$	element of $\mathbf{A}$ uniquely identified by its index ℓ; a multiset
C_i	ith order pure osmotic virial coefficient in the exact treatment
G	Gibbs free energy of the solution
G^{c}	configurational Gibbs free energy of the solution
G^{nc}	nonconfigurational Gibbs free energy of the solution
$g_{\mathbf{A}_\ell}$	nonconfigurational Gibbs free energy of a particle group denoted by $\mathbf{A}_\ell$ averaged over all spatial configurations
i, j, m, k	dummy indices
i_j	element of the set $\{i_1, i_2, \ldots, i_n\}$
$\{i_1, i_2, \ldots, i_n\}$	set of dummy indices used in repeated sums
k_{B}	Boltzmann constant
$(k_{\mathrm{d}})_j$	"dissociation constant" of solute j
ℓ	index used to uniquely identify elements of $\mathbf{A}$
M_1	molar mass of the solvent
n	number of particles in groups/n-tuples
N	total number of particles in a two-component solution
$N_{\mathbf{A}_\ell}$	number of n-tuples of composition $\mathbf{A}_\ell$
N_j	number of particles of type j
P	pressure of the solution
$\mathcal{P}_{\mathbf{A}_\ell}$	probability of formation of an n-tuple of composition $\mathbf{A}_\ell$
$Q_j^{n,m}$	mth order pure coefficient in the excess chemical potential expression of solute j in a degree-n expansion
$\hat{Q}_j^{n,m}$	mth order pure osmotic virial coefficient of component j in a degree-n expansion
r	number of components in the mixture
T	absolute temperature of the solution
$w_{\mathbf{A}_\ell}$	nonconfigurational Gibbs free energy change of formation of an n-tuple of composition $\mathbf{A}_\ell$
x_j	mole fraction of particles of type j
z_n	number of n-tuples sharing each particle
δ_{ij}	Kronecker delta
μ_j	chemical potential per particle of type j in the mixture
μ_j°	chemical potential per particle of pure substance j
μ_j^{ex}	excess chemical potential per particle of type j in the mixture
$\xi_{j\|\mathbf{A}_\ell}$	multiplicity of element j in the multiset $\mathbf{A}_\ell$
π	osmolality of the solvent
π^+	osmole fraction of the solvent
ω_i	Gibbs free energy change of formation of a particle group in the exact treatment
ω_i^{nc}	nonconfigurational part of ω_i

REFERENCES

[1] E. A. Guggenheim, *Mixtures: The Theory of the Equilibrium Properties of Some Simple Classes of Mixtures Solutions and Alloys* (Clarendon Press, Oxford, UK, 1952).

[2] T. L. Hill, *An Introduction to Statistical Thermodynamics* (Dover Publications, New York, 1986).

[3] W. G. McMillan and J. E. Mayer, "The statistical thermodynamics of multicomponent systems," J. Chem. Phys. **13**, 276–305 (1945).

[4] T. L. Hill, "Theory of solutions. I[1]," J. Am. Chem. Soc. **79**, 4885–4890 (1957).

[5] J. Reščič, V. Vlachy, A. Jamnik, and O. Glatter, "Osmotic pressure, small-angle x-ray, and dynamic light scattering studies of human serum albumin in aqueous solutions," J. Colloid Interface Sci. **239**, 49–57 (2001).

[6] T. C. Laurent, B. N. Preston, and L.-O. Sundelöf, "Transport of molecules in concentrated systems," Nature **279**, 60–62 (1979).

[7] R. S. King, H. W. Blanch, and J. M. Prausnitz, "Molecular thermodynamics of aqueous two-phase systems for bioseparations," AIChE J. **34**, 1585–1594 (1988).

[8] C. E. Neuzil, "Osmotic generation of 'anomalous' fluid pressures in geological environments," Nature **403**, 182–184 (2000).

[9] A. D. Ryabtsev, N. P. Kotsupalo, A. A. Kurakov, N. M. Nemkov, and A. G. Vakhromeev, "Rational use of multicomponent brines extracted together with oil," Theor. Found. Chem. Eng. **54**, 756–761 (2020).

[10] E. McCarthy, E. S. Fraga, and J. W. Ponton, "An automated procedure for multicomponent product separation synthesis," Comput. Chem. Eng. **22**, S77–S84 (1998).

[11] U. Hecht, L. Gránásy, T. Pusztai, B. Böttger, M. Apel, V. Witusiewicz, L. Ratke, J. De Wilde, L. Froyen, D. Camel, B. Drevet, G. Faivre, S. G. Fries, B. Legendre, and S. Rex, "Multiphase solidification in multicomponent alloys," Mater. Sci. Eng.: R: Rep. **46**, 1–49 (2004).

[12] J. Eiken, B. Böttger, and I. Steinbach, "Multiphase-field approach for multicomponent alloys with extrapolation scheme for numerical application," Phys. Rev. E **73**, 066122 (2006).

[13] S. N. Khan, J. B. Staunton, and G. M. Stocks, "Statistical physics of multicomponent alloys using KKR-CPA," Phys. Rev. B **93**, 054206 (2016).

[14] Y. Ikeda, B. Grabowski, and F. Körmann, "*Ab initio* phase stabilities and mechanical properties of multicomponent alloys: A comprehensive review for high entropy alloys and compositionally complex alloys," Mater. Charact. **147**, 464–511 (2019).

[15] Y. P. Kim, J. H. Seinfeld, and P. Saxena, "Atmospheric gas-aerosol equilibrium I. Thermodynamic model," Aerosol Sci. Technol. **19**, 157–181 (1993).

[16] A. S. Ansari and S. N. Pandis, "Prediction of multicomponent inorganic atmospheric aerosol behavior," Atmos. Environ. **33**, 745–757 (1999).

[17] T. Koop, B. Luo, A. Tsias, and T. Peter, "Water activity as the determinant for homogeneous ice nucleation in aqueous solutions," Nature **406**, 611–614 (2000).

[18] S. L. Clegg, J. H. Seinfeld, and P. Brimblecombe, "Thermodynamic modelling of aqueous aerosols containing electrolytes and dissolved organic compounds," J. Aerosol Sci. **32**, 713–738 (2001).

[19] M. D. Petters and S. M. Kreidenweis, "A single parameter representation of hygroscopic growth and cloud condensation nucleus activity," Atmos. Chem. Phys. **7**, 1961–1971 (2007).

[20] K. Semeniuk and A. Dastoor, "Current state of atmospheric aerosol thermodynamics and mass transfer modeling: A review," Atmosphere **11**, 156 (2020).

[21] D. Saulov, "On the multicomponent polynomial solution models," Calphad **30**, 405–414 (2006).

[22] Y. Kakuda, E. Uchida, and N. Imai, "A new model of the excess Gibbs energy of mixing for a regular solution," Proc. Jpn. Acad., Ser. B **70**, 163–168 (1994).

[23] J. A. W. Elliott, R. C. Prickett, H. Y. Elmoazzen, K. R. Porter, and L. E. McGann, "A multisolute osmotic virial equation for solutions of interest in biology," J. Phys. Chem. B **111**, 1775–1785 (2007).

[24] R. C. Prickett, J. A. W. Elliott, and L. E. McGann, "Application of the osmotic virial equation in cryobiology," Cryobiology **60**, 30–42 (2010).

[25] M. W. Zielinski, L. E. McGann, J. A. Nychka, and J. A. W. Elliott, "Comparison of non-ideal solution theories for multi-solute solutions in cryobiology and tabulation of required coefficients," Cryobiology **69**, 305–317 (2014).

[26] M. W. Zielinski, L. E. McGann, J. A. Nychka, and J. A. W. Elliott, "Comment on 'Determination of the quaternary phase diagram of the water–ethylene glycol–sucrose–NaCl system and a comparison between two theoretical methods for synthetic phase diagrams' Cryobiology 61, 52–57 (2010)," Cryobiology **70**, 287–292 (2015).

[27] R. C. Prickett, J. A. W. Elliott, and L. E. McGann, "Application of the multisolute osmotic virial equation to solutions containing electrolytes," J. Phys. Chem. B **115**, 14531–14543 (2011).

[28] H. Binyaminov, H. Sun, and J. A. W. Elliott, "Predicting freezing points of ternary salt solutions with the multisolute osmotic virial equation," (unpublished).

[29] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 1 (John Wiley & Sons, New York, 1971).

[30] J. A. W. Elliott, "Gibbsian surface thermodynamics," J. Phys. Chem. B **124**, 10859–10878 (2020).

Supplementary Material for

“Multicomponent solutions: Combining rules for multisolute osmotic virial coefficients”

Hikmat Binyaminov, Janet A. W. Elliott*

Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada, T6G 1H9

* Corresponding author: janet.elliott@ualberta.ca

This PDF file includes:

A. Detailed steps between equations (21) and (22) and equations (22) and (23)
B. Detailed steps between equations (28) and (29)
C. Detailed steps to arrive at equation (31) from equations (24) and (29)
D. Detailed steps between equations (48) and (50) and equations (49) and (51)

A. Detailed steps between equations (21) and (22) and equations (22) and (23)

From the main text, equation (21) is

$$\mu_1^{\mathrm{ex}} = \frac{z_n}{n} \sum_{i_2, i_3 \ldots i_n = 1}^{r} w_{1 i_2 \ldots i_n} x_{i_2} x_{i_3} \ldots x_{i_n} - \frac{z_n (n-1)}{n} \sum_{\substack{i_2, i_3 \ldots i_n = 1 \\ i_1 = 2}}^{r} \left(w_{i_1 i_2 \ldots i_n} - w_{1 i_2 \ldots i_n} \right) x_{i_1} x_{i_2} \ldots x_{i_n} \tag{S1}$$

To get equation (22), the omitted steps are:

<u>Step 1)</u> Separate the term containing x_1 in the sum over i_2 in equation (S1) from the rest and substitute $1 - \sum_{j=2}^{r} x_j$ for x_1 to get

$$
\begin{aligned}
\mu_1^{\mathrm{ex}} = {} & \frac{z_n}{n} \sum_{i_3,i_4 \dots i_n=1}^{r} w_{11i_3 \dots i_n} \left(1 - \sum_{j=2}^{r} x_j\right) x_{i_3} \dots x_{i_n} \\
& + \frac{z_n}{n} \sum_{\substack{i_3,i_4 \dots i_n=1 \\ i_2=2}}^{r} w_{1i_2 \dots i_n} x_{i_2} x_{i_3} \dots x_{i_n} \\
& - \frac{z_n(n-1)}{n} \sum_{\substack{i_3,i_4 \dots i_n=1 \\ i_1=2}}^{r} \left(w_{i_1 1 i_3 \dots i_n} - w_{11i_3 \dots i_n}\right) x_{i_1} \Bigg(1 \\
& - \sum_{j=2}^{r} x_j \Bigg) x_{i_3} \dots x_{i_n} \\
& - \frac{z_n(n-1)}{n} \sum_{\substack{i_3,i_4 \dots i_n=1 \\ i_1,i_2=2}}^{r} \left(w_{i_1 i_2 \dots i_n} - w_{1i_2 \dots i_n}\right) x_{i_1} x_{i_2} \dots x_{i_n}
\end{aligned}
\qquad \text{(S2)}
$$

<u>Step 2)</u> Where present in equation (S2), split the expression $(1 - \sum_{j=2}^{r} x_j)$ into two terms and recover the consecutive labels by reindexing the sum over $i_1 = 2$ and $i_3, i_4 \dots i_n = 1$ as the sum over $i_2 = 2$ and $i_3, i_4 \dots i_n = 1$ to get

$$
\begin{aligned}
\mu_1^{\mathrm{ex}} = \frac{z_n}{n} \sum_{i_3, i_4 \dots i_n = 1}^{r} w_{11 i_3 \dots i_n} x_{i_3} x_{i_4} \dots x_{i_n} \\
- \frac{z_n}{n} \sum_{\substack{i_3, i_4 \dots i_n = 1 \\ i_2 = 2}}^{r} w_{11 i_3 \dots i_n} x_{i_2} x_{i_3} \dots x_{i_n} \\
+ \frac{z_n}{n} \sum_{\substack{i_3, i_4 \dots i_n = 1 \\ i_2 = 2}}^{r} w_{1 i_2 \dots i_n} x_{i_2} x_{i_3} \dots x_{i_n} \\
- \frac{z_n(n-1)}{n} \sum_{\substack{i_3, i_4 \dots i_n = 1 \\ i_2 = 2}}^{r} \left(w_{1 i_2 \dots i_n} - w_{11 i_3 \dots i_n} \right) x_{i_2} x_{i_3} \dots x_{i_n} \\
+ \frac{z_n(n-1)}{n} \sum_{\substack{i_3, i_4 \dots i_n = 1 \\ i_1, i_2 = 2}}^{r} \left(w_{1 i_2 \dots i_n} - w_{11 i_3 \dots i_n} \right) x_{i_1} x_{i_2} \dots x_{i_n} \\
- \frac{z_n(n-1)}{n} \sum_{\substack{i_3, i_4 \dots i_n = 1 \\ i_1, i_2 = 2}}^{r} \left(w_{i_1 i_2 \dots i_n} - w_{1 i_2 \dots i_n} \right) x_{i_1} x_{i_2} \dots x_{i_n}
\end{aligned}
\tag{S3}
$$

<u>Step 3)</u> Collect the terms over the same sums in equation (S3) to get

$$\mu_1^{\text{ex}} = \frac{z_n}{n} \sum_{i_3, i_4 \ldots i_n = 1}^{r} w_{11 i_3 \ldots i_n} x_{i_3} x_{i_4} \ldots x_{i_n}$$

$$- \frac{z_n(n-2)}{n} \sum_{\substack{i_3 \ldots i_n = 1 \\ i_2 = 2}}^{r} \left(w_{1 i_2 \ldots i_n} - w_{11 i_3 \ldots i_n}\right) x_{i_2} x_{i_3} \ldots x_{i_n} \tag{S4}$$

$$- \frac{z_n(n-1)}{n} \sum_{\substack{i_3, i_4 \ldots i_n = 1 \\ i_1, i_2 = 2}}^{r} \left(w_{i_1 i_2 \ldots i_n} - 2 w_{1 i_2 \ldots i_n}\right.$$

$$\left. + w_{11 i_3 \ldots i_n}\right) x_{i_1} x_{i_2} \ldots x_{i_n}$$

which is the same as equation (22).

To get equation (23), the omitted steps are:

<u>Step 1)</u> Separate the term containing x_1 in the sum over i_3 in equation (S4) from the rest and substitute $1 - \sum_{j=2}^{r} x_j$ for x_1 to get

$$
\begin{aligned}
\mu_1^{\text{ex}} = {} & \frac{z_n}{n} \sum_{i_4,i_5\dots i_n=1}^{r} w_{111i_4\dots i_n}\left(1-\sum_{j=2}^{r} x_j\right) x_{i_4}x_{i_5}\dots x_{i_n} \\
& + \frac{z_n}{n} \sum_{\substack{i_4,i_5\dots i_n=1\\ i_3=2}}^{r} w_{11i_3\dots i_n} x_{i_3}x_{i_4}\dots x_{i_n} \\
& - \frac{z_n(n-2)}{n} \sum_{\substack{i_4,i_5\dots i_n=1\\ i_2=2}}^{r} \left(w_{1i_21i_4\dots i_n} - w_{111i_4\dots i_n}\right) x_{i_2} \left(1 \right. \\
& \left. - \sum_{j=2}^{r} x_j\right) x_{i_4}\dots x_{i_n} \\
& - \frac{z_n(n-2)}{n} \sum_{\substack{i_4,i_5\dots i_n=1\\ i_2,i_3=2}}^{r} \left(w_{1i_2\dots i_n} - w_{11i_3\dots i_n}\right) x_{i_2}x_{i_3}\dots x_{i_n} \\
& - \frac{z_n(n-1)}{n} \sum_{\substack{i_4,i_5\dots i_n=1\\ i_1,i_2=2}}^{r} \left(w_{i_1i_21i_4\dots i_n} - 2w_{1i_21i_4\dots i_n} \right. \\
& \left. + w_{111i_4\dots i_n}\right) x_{i_1}x_{i_2}\left(1-\sum_{j=2}^{r} x_j\right) x_{i_4}\dots x_{i_n} \\
& - \frac{z_n(n-1)}{n} \sum_{\substack{i_4,i_5\dots i_n=1\\ i_1,i_2,i_3=2}}^{r} \left(w_{i_1i_2\dots i_n} - 2w_{1i_2\dots i_n} \right. \\
& \left. + w_{11i_3\dots i_n}\right) x_{i_1}x_{i_2}\dots x_{i_n}
\end{aligned}
\tag{S5}
$$

Step 2) Where present in equation (S5), split the expression $(1-\sum_{j=2}^{r} x_j)$ into two terms and recover the consecutive labels where needed by reindexing to get

$$
\begin{aligned}
\mu_1^{\text{ex}} = {} & \frac{z_n}{n}\sum_{i_4,i_5\ldots i_n=1}^{r} w_{111i_4\ldots i_n}x_{i_4}x_{i_5}\ldots x_{i_n} - \frac{z_n}{n}\sum_{\substack{i_4,i_5\ldots i_n=1\\ i_3=2}}^{r} w_{111i_4\ldots i_n}x_{i_3}x_{i_4}\ldots x_{i_n} \\
& + \frac{z_n}{n}\sum_{\substack{i_4,i_5\ldots i_n=1\\ i_3=2}}^{r} w_{11i_3\ldots i_n}x_{i_3}x_{i_4}\ldots x_{i_n} \\
& - \frac{z_n(n-2)}{n}\sum_{\substack{i_4,i_5\ldots i_n=1\\ i_3=2}}^{r} \left(w_{11i_3i_4\ldots i_n} - w_{111i_4\ldots i_n}\right)x_{i_3}x_{i_4}\ldots x_{i_n} \\
& + \frac{z_n(n-2)}{n}\sum_{\substack{i_4,i_5\ldots i_n=1\\ i_2,i_3=2}}^{r} \left(w_{11i_3i_4\ldots i_n} - w_{111i_4\ldots i_n}\right)x_{i_2}x_{i_3}\ldots x_{i_n} \\
& - \frac{z_n(n-2)}{n}\sum_{\substack{i_4,i_5\ldots i_n=1\\ i_2,i_3=2}}^{r} \left(w_{1i_2\ldots i_n} - w_{11i_3\ldots i_n}\right)x_{i_2}x_{i_3}\ldots x_{i_n} \\
& - \frac{z_n(n-1)}{n}\sum_{\substack{i_4,i_5\ldots i_n=1\\ i_2,i_3=2}}^{r} \left(w_{1i_2\ldots i_n} - 2w_{11i_3\ldots i_n}\right. \\
& \left. + w_{111i_4\ldots i_n}\right)x_{i_2}x_{i_3}\ldots x_{i_n} \\
& + \frac{z_n(n-1)}{n}\sum_{\substack{i_4,i_5\ldots i_n=1\\ i_1,i_2,i_3=2}}^{r} \left(w_{1i_2\ldots i_n} - 2w_{11i_3\ldots i_n}\right. \\
& \left. + w_{111i_4\ldots i_n}\right)x_{i_1}x_{i_2}\ldots x_{i_n} \\
& - \frac{z_n(n-1)}{n}\sum_{\substack{i_4,i_5\ldots i_n=1\\ i_1,i_2,i_3=2}}^{r} \left(w_{i_1i_2\ldots i_n} - 2w_{1i_2\ldots i_n}\right. \\
& \left. + w_{11i_3\ldots i_n}\right)x_{i_1}x_{i_2}\ldots x_{i_n}
\end{aligned}
\tag{S6}
$$

<u>Step 3)</u> Collect the terms over the same sums in equation (S6) to get

$$
\begin{aligned}
\mu_1^{\text{ex}} = \frac{z_n}{n} & \sum_{i_4,i_5\dots i_n=1}^{r} w_{111i_4\dots i_n} x_{i_4} x_{i_5} \dots x_{i_n} \\
& - \frac{z_n(n-3)}{n} \sum_{\substack{i_4,i_5\dots i_n=1\\ i_3=2}}^{r} \left(w_{11i_3i_4\dots i_n} - w_{111i_4\dots i_n}\right) x_{i_3} x_{i_4} \dots x_{i_n} \\
& - \frac{z_n(2n-3)}{n} \sum_{\substack{i_4,i_5\dots i_n=1\\ i_2,i_3=2}}^{r} \left(w_{1i_2\dots i_n} - 2w_{11i_3\dots i_n} \right. \\
& \left. + w_{111i_4\dots i_n}\right) x_{i_2} x_{i_3} \dots x_{i_n} \\
& - \frac{z_n(n-1)}{n} \sum_{\substack{i_4,i_5\dots i_n=1\\ i_1,i_2,i_3=2}}^{r} \left(w_{i_1i_2\dots i_n} - 3w_{1i_2\dots i_n} + 3w_{11i_3\dots i_n} \right. \\
& \left. - w_{111i_4\dots i_n}\right) x_{i_1} x_{i_2} \dots x_{i_n}
\end{aligned}
\tag{S7}
$$

which is the same as equation (23).

B. Detailed steps between equations (28) and (29)

Using the definition of w given by equation (9), in the explicit notation, we can write

$$
w_{1\dots 1i_{k+1}\dots i_m} = g_{1\dots 1i_{k+1}\dots i_m} - \frac{1}{n}\left[(n-m+k)g_{1\dots 1} + \sum_{j=k+1}^{m} g_{i_j\dots i_j}\right] \tag{S8}
$$

noting that each variable index here can only represent a solute: $i_j \in [2, r]$. If we look at a generic binary group containing the same number of solvent particles (*i.e.,* $n-m+k$) and the same number of solute particles as above (*i.e.,* $m-k$), we have

$$w_{1\dots 1 i_j \dots i_j} = g_{1\dots 1 i_j \dots i_j} - \frac{1}{n}\left[(n-m+k)g_{1\dots 1} + (m-k)g_{i_j \dots i_j}\right] \tag{S9}$$

Now, if we take the sum over the labels from $j = k+1$ to m of both sides of equation (S9) and divide both sides by $m-k$, we have

$$\begin{aligned} &\frac{1}{m-k}\sum_{j=k+1}^{m} w_{1\dots 1 i_j \dots i_j} \\ &\qquad = \frac{1}{m-k}\sum_{j=k+1}^{m} g_{1\dots 1 i_j \dots i_j} \\ &\qquad - \frac{1}{n}\left[(n-m+k)g_{1\dots 1} + \sum_{j=k+1}^{m} g_{i_j \dots i_j}\right] \end{aligned} \tag{S10}$$

Using equation (28), we can substitute for the first term on the right-hand side of equation (S10) and obtain

$$\begin{aligned} &\frac{1}{m-k}\sum_{j=k+1}^{m} w_{1\dots 1 i_j \dots i_j} \\ &\qquad = g_{1\dots 1 i_{k+1} \dots i_m} - \frac{1}{n}\left[(n-m+k)g_{1\dots 1} + \sum_{j=k+1}^{m} g_{i_j \dots i_j}\right] \end{aligned} \tag{S11}$$

Comparing equations (S11) and (S8), we get

$$w_{1\dots 1 i_{k+1}\dots i_m} = \frac{1}{m-k} \sum_{j=k+1}^{m} w_{1\dots 1 i_j \dots i_j} \tag{S12}$$

which is the same as equation (29).

C. Detailed steps to arrive at equation (31) using equations (24) and (29)

Given equation (29), we want to show that the following holds for the sum appearing in equation (24):

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} w_{1\dots 1 i_{k+1}\dots i_m} = \frac{1}{m} \sum_{j=1}^{m} \sum_{k=0}^{m} (-1)^k \binom{m}{k} w_{1\dots 1 i_j \dots i_j} \tag{S13}$$

which would give us the sum over Q's when multiplied by the coefficient $\frac{z_n(1-m)}{n}\binom{n}{m}$ (omitted here).

Starting with the left-hand side of equation (S13), we substitute for $w_{1\dots 1 i_{k+1}\dots i_m}$ using its decomposition given by equation (29):

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} w_{1\dots 1 i_{k+1}\dots i_m} = \sum_{k=0}^{m} (-1)^k \binom{m}{k} \frac{1}{m-k} \sum_{j=k+1}^{m} w_{1\dots 1 i_j \dots i_j} \tag{S14}$$

Now, we extend the sum from $j = k + 1$ to m to $j = 1$ to m by noting

$$\sum_{j=1}^{m} w_{1\ldots 1 i_j \ldots i_j} = \sum_{j=k+1}^{m} w_{1\ldots 1 i_j \ldots i_j} + \sum_{j=1}^{k} w_{1\ldots 1 i_j \ldots i_j} \tag{S15}$$

Since we at most need $m-k$ distinct labels for the indices of w on the left-hand side of equation (S13) and the sum from $j=k+1$ to m is chosen arbitrarily, we can reindex the terms in the second sum on the right-hand side of equation (S15) by splitting each term into a sum over index labels $k+1$ to m. That is, we can write

$$\begin{aligned} \sum_{j=1}^{m} w_{1\ldots 1 i_j \ldots i_j} &= \sum_{j=k+1}^{m} w_{1\ldots 1 i_j \ldots i_j} + \sum_{j=1}^{k} \frac{1}{m-k} \sum_{t=k+1}^{m} w_{1\ldots 1 i_t \ldots i_t} = \\ &= \sum_{j=k+1}^{m} w_{1\ldots 1 i_j \ldots i_j} + \frac{1}{m-k} \sum_{t=k+1}^{m} \sum_{j=1}^{k} w_{1\ldots 1 i_t \ldots i_t} \\ &= \sum_{j=k+1}^{m} w_{1\ldots 1 i_j \ldots i_j} + \frac{k}{m-k} \sum_{t=k+1}^{m} w_{1\ldots 1 i_t \ldots i_t} \\ &= \frac{m}{m-k} \sum_{j=k+1}^{m} w_{1\ldots 1 i_j \ldots i_j} \end{aligned} \tag{S16}$$

Rearranging this result and substituting it back into equation (S14), we obtain equation (S13):

$$\sum_{k=0}^{m}(-1)^k\binom{m}{k}w_{1\ldots1i_{k+1}\ldots i_m}$$

$$=\sum_{k=0}^{m}(-1)^k\binom{m}{k}\frac{1}{m-k}\left(\frac{m-k}{m}\sum_{j=1}^{m}w_{1\ldots1i_j\ldots i_j}\right)\qquad\text{(S17)}$$

$$=\frac{1}{m}\sum_{j=1}^{m}\sum_{k=0}^{m}(-1)^k\binom{m}{k}w_{1\ldots1i_j\ldots i_j}$$

D. Detailed steps between equations (48) and (50) and equations (49) and (51)

According to the definition before equation (45), ω_2^{nc} in equation (48) and (49) is defined as the nonconfigurational Gibbs free energy change for the following process: 2 systems, each made of $N-1$ solvent molecules and 1 solute molecule, being rearranged into 1 system made of $N-2$ solvent molecules with 2 solute molecules, plus 1 pure solvent system with N solvent particles. Similarly, ω_3^{nc} in equation (49) is defined as the nonconfigurational Gibbs free energy change for the following process: 3 systems, each made of $N-1$ solvent molecules and 1 solute molecule, being rearranged into 1 system made of $N-3$ solvent molecules with 3 solute molecules, plus 2 pure solvent system with N solvent particles. Based on these definitions, denoting the solute by the subscript 2, we can write

$$\omega_2^{\mathrm{nc}}=g_{1\ldots122}+g_{1\ldots1}-2g_{1\ldots12}\qquad\text{(S18)}$$

and

$$\omega_3^{\mathrm{nc}} = g_{1\ldots1222} + 2g_{1\ldots1} - 3g_{1\ldots12} \tag{S19}$$

where g's denote the energies of the respective groups made of N particles each, consistent with our notation (*i.e.,* N-tuple energies). We can recast equations (S18) into the following form:

$$\begin{aligned}\omega_2^{\mathrm{nc}} &= g_{1\ldots122} - \frac{1}{N}[(N-2)g_{1\ldots1} + 2g_{2\ldots2}] \\ &\quad - 2\left\{g_{1\ldots12} - \frac{1}{N}[(N-1)g_{1\ldots1} + g_{2\ldots2}]\right\} \\ &= w_{1\ldots122} - 2w_{1\ldots12}\end{aligned} \tag{S20}$$

which is the same equivalence used in equation (50), and recast equation (S19) into the following form:

$$\begin{aligned}\omega_3^{\mathrm{nc}} &= g_{1\ldots1222} - \frac{1}{N}[(N-3)g_{1\ldots1} + 3g_{2\ldots2}] \\ &\quad - 3\left\{g_{1\ldots12} - \frac{1}{N}[(N-1)g_{1\ldots1} + g_{2\ldots2}]\right\} \\ &= w_{1\ldots1222} - 3w_{1\ldots12}\end{aligned} \tag{S21}$$

Combining equations (S20) and (S21) into the expression present in equation (49), we have

$$\begin{aligned}\omega_3^{\mathrm{nc}} - 3\omega_2^{\mathrm{nc}} &= w_{1\ldots1222} - 3w_{1\ldots12} - 3(w_{1\ldots122} - 2w_{1\ldots12}) \\ &= w_{1\ldots1222} - 3w_{1\ldots122} + 3w_{1\ldots12}\end{aligned} \tag{S22}$$

which is the same equivalence used in equation (51). The arguments above can be generalized to relate any order coefficients expressed in terms of ω^{nc} to the respective coefficients expressed in terms of w's. We will not show this here to keep the scope of the supplementary material consistent with that of the main text.