THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Molecular Aggregation Equilibria. Comparison of Finite Lattice and Weighted Random Mixing Predictions

Dor Ben-Amotz* and Blake M. Rankin

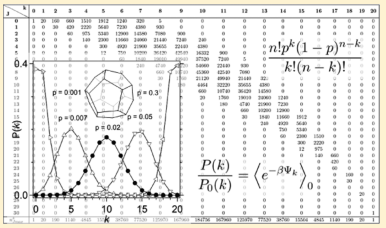
Purdue University, Department of Chemistry, West Lafayette, Indiana 47907, United States

B. Widom

Cornell University, Department of Chemistry, Ithaca, New York 14853, United States

Supporting Information

ABSTRACT: Molecular aggregation equilibria are described using finite lattice and mean field theoretical modeling strategies, both built upon a random mixture reference system. The resulting predictions are compared with each other for systems in which each aggregate consists of a central solute molecule whose first coordination shell can accommodate multiple bound ligands. Solute−ligand (direct) and ligand−ligand (cooperative) interactions are found to influence aggregate size distributions in qualitatively different ways, as direct interactions produce a shape-invariant transformation of the aggregate size distribution, whereas cooperative interactions can lead to a vapor−liquid-like transformation. When half the ligand binding sites are filled, the corresponding aggregate size distributions are invariably unimodal in the absence of cooperative interactions, but when the latter interactions are attractive, the distributions are predicted to be bimodal below and unimodal above a critical temperature. Mean field and finite lattice predictions are found to be in globally good agreement with each other, except under near-critical conditions, and even there, the predicted average aggregate sizes and equilibrium constants are remarkably similar. Potential applications of these theoretical predictions to the analysis of experimental and molecular dynamics aggregation results are discussed.



1. INTRODUCTION

Molecular aggregation plays a central role in processes ranging from antibody binding to mesoscopic self-assembly and crystallization. In general, both entropic and energetic interactions influence the concentration at which significant aggregation takes place and the shapes of the associated aggregate size distributions. Entropic contributions arise from configurational statistics, whereas energetic contributions result from intermolecular interactions. The latter interactions may include both direct interactions between each monomer and the aggregate and cooperative (allosteric) interactions between bound monomers. Here we present a general theoretical framework for describing aggregation equilibria, built upon a random mixture reference system and incorporating both direct and cooperative interactions.

Our results, which include predictions obtained using exact finite lattice (FL) statistics and a weighted random mixing (WRM) mean field approximation strategy, reveal that although both direct and cooperative interactions influence the concentration at which significant aggregation takes place, cooperative interactions may further give rise to two classes of aggregation behavior, resembling macroscopic vapor−liquid phase coexistence, terminating in a critical point. Although the importance of such transformations in aggregation equilibria have not, to the best of our knowledge, been explicitly described in previous studies, they are undoubtedly linked to micelle formation[1−3] and other aggregation processes that have been found to give rise to bimodal aggregate size distributions.[4−6] Our results are also related to previous theoretical descriptions of adsorption equilibria[7] and allosteric regulation of protein function,[8−11] and they may prove to be of more general relevance to biological and mesoscopic aggregation processes[12−16] as well as to crystallization (as recent studies point to a link between equilibrium aggregation and nucleation kinetics in explaining enormous discrepancies between experimental observations and classical nucleation theory predictions).[17−19]

Although our theoretical modeling strategy is quite general, here we present results pertaining to a class of aggregation processes that resemble the binding of one or more ligands to a central molecule. Our primary aim is to quantitatively describe the *aggregate size distributions* that dictate the relative abundance of aggregates containing different numbers of ligands bound to the central molecule. We further find it useful to consider aggregate size distributions obtained when exactly half of the available ligand binding sites are filled, as

Special Issue: James L. Skinner Festschrift

Received: January 11, 2014
Revised: March 19, 2014
Published: March 22, 2014

these not only dictate the characteristic ligand concentration above which substantial aggregation takes place but also most clearly reveal the influence of cooperative interactions in driving a transformation between unimodal and bimodal aggregate size distributions, either with increasing ligand−ligand attraction or decreasing temperature. Cooperative interactions are here defined as any changes in the binding energy of a given ligand induced by the presence of other bound ligands.

A central molecule with n equivalent ligand binding sites on its surface and c nearest neighbors per binding site may be treated as a finite two-dimensional lattice with periodic boundary conditions, resembling sites on the surface of a three-dimensional object. For example, a molecule with four equivalent binding sites, each of which are nearest neighbors, may be represented by a tetrahedral lattice with $n = 4$ and $c = 3$, whereas an aggregate with $n = 20$ and $c = 3$ is equivalent to a dodecahedral lattice. Here we utilize such lattice mappings to obtain exact FL predictions for the corresponding aggregate size distributions, and we compare the results with those obtained using the WRM mean field approximation that replaces the energy of an aggregate containing k bound ligands by its average value, evaluated in the random mixture reference system.

The remainder of this paper is organized as follows. In Section 2, we describe a general theoretical strategy linking idealized random mixing and energetically driven aggregation processes. In Section 3, the latter general results are used to obtain exact FL aggregation predictions. In Section 4, we describe the WRM approximation strategy, which may be viewed as a mean field approximation to the exact FL results. In Section 5, we compare FL and WRM predictions to both highlight the conditions under which their predictions become identical and elucidate the role of cooperative interactions in driving a transformation from unimodal to bimodal aggregation behavior. Our conclusions are summarized and discussed in Section 6, including potential applications to the analysis of both experimental and molecular dynamics simulation results.

2. LINKING RANDOM AND DRIVEN AGGREGATION

A uniform random mixture reference system is here defined as one in which the concentration of each species is everywhere identical to its bulk concentration.[20] In other words, such a system is one in which the probability that a particular molecule will be found in any region (of molecular dimension) is equal to the product of the molecule's bulk concentration and the volume of the region. More generally, we may consider subdividing a volume V into n subregions or *binding sites*, each of which has a volume $v_0 = (1/n)V$. Thus, the random mixing probability that a ligand will occupy each such binding site is $p = v_0[L]_0$, where $[L]_0$ is the total ligand concentration in the random mixture of interest. For any such system, the binomial distribution may be used to obtain the following probability $P_0(k)$ that exactly k of the n binding sites will be occupied by a ligand.[16,20,21]

$$P_0(k) = \frac{n!p^k(1-p)^{n-k}}{k!(n-k)!} \tag{1}$$

If we further associate the region of volume V with the coordination shell of a solute, which may be occupied by up to a maximum of n ligand molecules, then we may identify $P_0(k)$ as the corresponding aggregate size distribution in the random mixture reference system. Thus, a random mixture reference system is here defined as a system in which the aggregate size distribution is described by eq 1. Moreover, $e^{\beta\mu}/\Lambda^3 = [p/(1-p)](1/v_0)$ plays the role of the ligand activity in the random mixture reference system, where $\beta = 1/k_BT$, Λ is the ligand thermal wavelength and μ is the ligand chemical potential (excluding contributions arising from the internal partition function of the ligand).[7]

Note that $P_0(k)$ is normalized, such that $\sum_{k=0}^{n} P_0(k) = 1$, and the average number of occupied ligand binding sites is $\langle k\rangle_0 = \sum_{k=0}^{n} kP_0(k) = np$. More generally, the average $\langle ...\rangle_0$ performed over equilibrium configurations of a random mixture reference system is defined as follows.

$$\langle ...\rangle_0 = \sum_{k=0}^{n} ... P_0(k) \tag{2}$$

We may now consider how such an aggregate size distribution would change if there was an additional interaction energy change Ψ_k associated with transferring k ligands from the bulk solution (outside of the solute's first coordination shell) into the solute's coordination shell. It is important to note that each of the distinct arrangements of k ligands within an aggregate may in general have a different value of Ψ_k. The potential distribution theorem[22] may be used to obtain the following exact expression for the ratio of the equilibrium aggregate size distribution $P(k)$ in the interacting system, relative to that in the noninteracting random mixture system $P_0(k)$.

$$\frac{P(k)}{P_0(k)} = \langle e^{-\beta\Psi_k}\rangle_0 \tag{3}$$

If we wish to normalize $P(k)$, such that $\sum_{k=0}^{n} P_0(k) = 1$, then we must divide $P(k)$ by the appropriate partition function $\sum_{k=0}^{n} P_0(k)e^{-\beta\mu^x(k)}$, where $\mu^x(k) = -k_BT \ln\langle e^{-\beta\Psi_k}\rangle_0$ is the excess chemical potential associated with introducing the intermolecular interaction energies Ψ_k.

The aggregate size distributions $P(k)$ may be used to quantify various aggregate formation, interconversion, and partitioning equilibria. For example, we may consider an aggregation process such as $S + kL \rightleftharpoons SL_k$, where S and L represent free solute and ligand molecules, respectively, and SL_k is an aggregate containing k bound ligands. The equilibrium constant K_k for such an aggregation process may be expressed as follows, where $[S]$, $[L]$, and $[SL_k]$ are the corresponding equilibrium concentrations, and $P(k)$ is the normalized aggregate size distribution.

$$K_k = \frac{[SL_k]}{[S][L]^k} = \frac{P(k)}{P(0)[L]^k} \tag{4}$$

We may also consider equilibria pertaining to the interconversion between aggregates of different sizes. For example, the interconversion of an aggregate of size k to an aggregate of size k', may be expressed either as $SL_k + k'L \rightleftharpoons SL_{k'} + kL$ or equivalently (if $k' > k$) as $SL_k + (k'-k)L \rightleftharpoons SL_{k'}$. The equilibrium constant for any such interconversion process is equal to $K_{k'}/K_k$, and thus may be obtained from eq 4.

We may further use $P(k)$ to predict the equilibrium constant pertaining to the partitioning of ligands between the free and bound states $L_{free} \rightleftharpoons L_{bound}$, where $[L_{free}]$ and $[L_{bound}]$ correspond to the local concentrations of the ligand outside and inside of the coordination shell of the solute, respectively. Thus, in a random mixture, the latter two concentrations are

dx.doi.org/10.1021/jp5003337 | J. Phys. Chem. B 2014, 118, 7878−7885

necessarily equal to each other, so $K_p = 1$, whereas in a nonrandom mixture, K_p is equivalent to the ratio of the average number of bound ligands in the system of interest to that in the corresponding random mixture.

$$K_p = \frac{[L_{\text{bound}}]}{[L_{\text{free}}]} = \frac{\langle k \rangle}{\langle k \rangle_0} \tag{5}$$

Because all of the above equilibrium constants are expressed as ratios of actual concentrations (rather than activities), they are only expected to be concentration independent under sufficiently dilute conditions. However, eqs 4 and 5 are valid even when K_k or K_p are not concentration independent, as long as the corresponding aggregate size distribution $P(k)$ and free ligand concentration $[L]$ are accurately known.

We have explicitly determined $P(k)$ pertaining to solute−ligand aggregation processes in two types of model systems. One is a FL system in which the aggregate is treated as a finite two-dimensional lattice with n sites, c nearest neighbors per site, a direct solute-ligand site binding energy of ϵ_1, and a nearest neighbor ligand−ligand contact energy of ϵ_2 (as further described in Section 3 and Supporting Information). The second is the WRM model in which all aggregates of size k are assumed to have the same energy, rather than a distribution of energies (as further described in Section 4).

3. FINITE LATTICE (FL) PREDICTIONS

A lattice with n sites, c nearest neighbors, and k filled sites may have various numbers of nearest neighbor pairs $0 \le J \le J_{\max}$, where $J_{\max}$ is the maximum possible number of pairs (which cannot exceed the total number of links between neighboring lattice sites). We wish to determine the total number of possible aggregate configurations n_J^k that have k filled sites and J contacts between filled sites.

Note that there are a total of $\sum_{k=0}^{n} n!/[(n-k)!k!] = 2^n$ unique configurations that have any value of $0 \le k \le n$. To determine n_J^k (for any $0 \le k \le n$ and $0 \le J \le J_{\max}$), one must count the number of pairs in all aggregate configurations. This may be done by mapping each unique aggregate configuration onto each of the binary numbers between 0 and $2^n - 1$. Each of the latter binary numbers contains n digits (each of which may be either 0 or 1) such that no two binary numbers have the same configuration of 0's and 1's, and all possible configurations are included among the entire set of 2^n numbers.

One may map each of the 2^n binary numbers onto lattice configurations by associating each lattice site with a number between 1 and n. Each of the numbered lattice sites are then filled if the corresponding binary digit is 1 and empty if it is 0. In order to determine the value of J for a given configuration of filled lattice sites, it is necessary to make use of a list of each of the c nearest neighbors to each lattice site and thus count the number of pairs of filled sites that are in contact with each other. We have used this procedure to construct tables of n_J^k values for lattices with n = 4, 6, 8, 12, and 20 and c = 3, as well as for several other lattices with different values of c (see Supporting Information).

Once the table of n_J^k values for a given lattice is determined, the following expression may be used to obtain normalized aggregate size distributions for systems with various values of ϵ_1 and ϵ_2.

$$P(k) = \frac{1}{\chi}\left\{\left[\left(\frac{p}{1-p}\right)e^{-\beta\epsilon_1}\right]^k\left[\sum_{J=0}^{J_{\max}} n_J^k e^{-J\beta\epsilon_2}\right]\right\} \tag{6}$$

The normalization constant χ is obtained as follows.

$$\chi = \sum_{k=0}^{n}\left\{\left[\left(\frac{p}{1-p}\right)e^{-\beta\epsilon_1}\right]^k\left[\sum_{J=0}^{J_{\max}} n_J^k e^{-J\beta\epsilon_2}\right]\right\} \tag{7}$$

Note that eq 6 reduces to eq 1 when $\epsilon_1 = \epsilon_2 = 0$, because $\sum_{J=0}^{J_{\max}} n_J^k = n!/[k!(n-k)!]$ and χ becomes equivalent to $[1/(1-p)]^n$. Moreover, when $\epsilon_1 = \epsilon_2 = 0$, eq 6 implies that $(1/2)k(k-1)[c/(n-1)]$ is the average number of contacts between bound ligands, as further discussed in Section 4 (and in the Supporting Information).

4. WEIGHTED RANDOM MIXING (WRM) PREDICTIONS

The following WRM expression for $P(k)$ is obtained by replacing the distribution of energies Ψ_k in eq 3 by its average over all reference system configurations $\epsilon(k) = \langle \Psi_k \rangle_0$ (obtained using eq 2).

$$\frac{P(k)}{P_0(k)} = \langle e^{-\beta\langle\Psi_k\rangle_0}\rangle_0 = e^{-\beta\epsilon(k)} \tag{8}$$

If we wish to normalize $P(k)$, then we must divide the above expression for $P(k)$ by $\sum_{k=0}^{n} P_0(k)e^{-\beta\epsilon(k)}$.

We may use eq 8 to define the chemical potential of an aggregate containing k bound ligands as $\mu_k = -k_B T \ln P(k) = -k_B T \ln P_0(k) + \epsilon(k)$, from which the corresponding entropy s_k and energy u_k may be obtained by differentiation; $s_k = -(\partial\mu_k/\partial T)_p = k_B \ln P_0(k)$ and $u_k = [\partial(\mu_k/T)/\partial(1/T)]_p = \epsilon(k)$. Thus, the WRM approximation implies that s_k arises entirely from the random mixture reference system while u_k is equivalent to the energy change associated with transferring k ligands from the bulk to the coordination shell of the solute. The fact that the above derivatives were performed at constant p (where p is the probability defined in Section 2) is equivalent to fixing the free ligand concentration (or fixing both the total number of free ligands and system volume).

Let us first consider a class of aggregation processes that have a constant (k-independent) *direct* binding energy per ligand ϵ_1. In other words, such aggregation processes are ones for which cooperative ligand−ligand interactions are neglibible ($|\beta\epsilon_2| \ll 1$). Thus, the total direct solute−ligand interaction energy $\epsilon(k)$ is proportional to the number k of bound ligands.

$$\Psi_k = \epsilon(k) = k\epsilon_1 \tag{9}$$

Note that for any such system, WRM and FL predictions are necessarily identical, because every value of k is associated with exactly one value of $\Psi_k = \epsilon(k)$. Moreover, when $\epsilon(k)$ is given by eq 9, $P(k)$ has the following dependence on p and ϵ_1 (to within a constant of proportionality whose value depends on k and n).

$$P(k) \propto \left[\left(\frac{p}{1-p}\right)e^{-\beta\epsilon_1}\right]^k \tag{10}$$

This expression makes it clear that the functional form of $P(k)$ is invariant to any changes in both ϵ_1 and p that leave the quantity $[p/(1-p)]e^{-\beta\epsilon_1}$ unchanged. For example, this implies that the aggregate size distribution obtained with some particular pair of ϵ_1 and p values is necessarily identical to that obtained in an idealized nonaggregating random mixture

dx.doi.org/10.1021/jp5003337 | J. Phys. Chem. B 2014, 118, 7878−7885

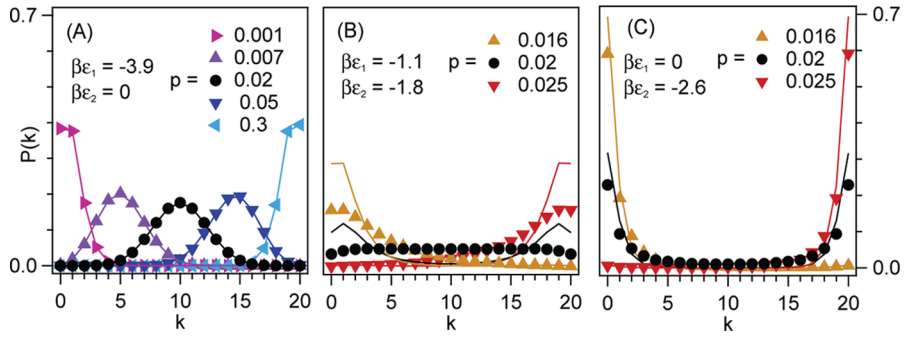


Figure 1. The effect of p (the ligand concentration) on the aggregate size distributions is illustrated when either $\epsilon_2 = 0$ (A) or $\epsilon_1 = 0$ (C). These two extremes are separated by a critical temperature $T_c \approx |\epsilon_2|/(1.8\ k_c)$ at which $P(k)$ transforms from a unimodal to bimodal distribution (B). Note that increasing p is identical to increasing the local ligand concentration, and thus the distributions in (A) may be obtained by fixing p and varying ϵ_1.

(for which $\epsilon_1 = 0$) with a site occupancy probability p_0 that is related to ϵ_1 and p as follows.

$$p_0 = \frac{1}{\left(\frac{1}{p} - 1\right)e^{\beta\epsilon_1} + 1} \tag{11}$$

The value of p_0 is equivalent to the average ligand binding probability (per site) in an aggregate characterized by ϵ_1 and p. In other words, the above invariant property implies that p_0/v_0 is the local monomer concentration in the aggregate when the monomer binding energy is ϵ_1, and the monomer concentration in the surrounding bulk solution (outside of the solute's coordination shell) is p/v_0. Note that this implies that the average number of bound ligands for such an aggregation process is $\langle k\rangle = \sum_{k=0}^{n} kP(k) = np_0$. Moreover, $p/p_0 = \langle k\rangle/\langle k\rangle_0 = K_p$ is equivalent to the partition coefficient of ligand molecules between the bulk and the aggregate (as given by eq 5).

More generally, we may consider the influence of *cooperative* ligand–ligand interactions on aggregation equilibria. Specifically, as in the FL model (Section 3), we identify ϵ_2 as the average interaction energy between any one of the k bound ligands and any other ligands that occupy any one of the c nearest neighbor sites around that bound ligand. For such systems, the WRM approximation implies that $\epsilon(k) = \langle\Psi_k\rangle_0$ has the following nonlinear dependence on k.

$$\epsilon(k) = k\epsilon_1 + \frac{1}{2}k(k-1)\left(\frac{c}{n-1}\right)\epsilon_2 \tag{12}$$

Note that $(1/2)k(k-1)$ is equivalent to the total number of unique pairs among k objects, and $c/(n-1)$ is the ratio of the number of nearest neighbors to the total number of sites around a given bound ligand (in a system with a total of n binding sites). Thus, $(1/2)k(k-1)[c/(n-1)]$ is exactly equivalent to the average number of occupied nearest neighbor contacts in a random mixing aggregation process (as further explained in the Supporting Information). More generally, we may define $\epsilon_2' = [c/(n-1)]\epsilon_2$, where ϵ_2' is the effective interaction energy between any pair of bound ligands, and thus replace eq 12 by $\epsilon(k) = k\epsilon_1 + (1/2)k(k-1)\epsilon_2'$.

The above WRM aggregation predictions lead to the following additional symmetry condition pertaining to aggregates in which half of the binding sites are filled, and thus $\langle k\rangle/n = 0.5$. Note that any aggregate size distribution that is symmetric, in the sense that $P(k) = P(n-k)$, is necessarily exactly half-filled. Our WRM results indicate that the converse is also true, as half-filled distributions are symmetric. For any system with such a symmetric half-filled aggregate size distribution, the parameters p, ϵ_1, and ϵ_2 must be related as follows.

$$\left(\frac{p}{1-p}\right)e^{-\beta\epsilon_1} = e^{(1/2)\beta c\epsilon_2} \tag{13}$$

Equivalently, eq 13 may be rearranged to obtain the following expression for the value of $p = p_{1/2}$ at which half of the ligand binding sites are filled.

$$p_{1/2} = \frac{1}{e^{-\beta(\epsilon_1 + \frac{c}{2}\epsilon_2)} + 1} \tag{14}$$

We will see that the shapes of the aggregate size distributions $P(k)$ for any such symmetric half-filled systems may either be unimodal or bimodal. Moreover, for any system with a bimodal aggregate size distribution, there exists a critical temperature $T^* = T_c(k_B/|\epsilon_2|) = (\beta_c|\epsilon_2|)^{-1}$, where $\epsilon_2 < 0$ and T_c is the temperature above which the distribution becomes unimodal. Because the left-hand-side of eq 13 is invariant to simultaneous changes in ϵ_1 and p (as explained above), the value of T^* is necessarily independent of the direct solute–ligand interaction energy ϵ_1 and thus depends only on the cooperative ligand–ligand interaction energy ϵ_2 (as well as the values of c and n). Equation 14 pertains not only to the WRM approximation but also to exact FL aggregation processes with c nearest neighbors per site, in the macroscopic limit.[23,24] Moreover, the explicit FL results we have obtained (as described in Section 3 and the Supporting Information) imply that this is also the case when n is finite.

5. COMPARISON OF FL AND WRM PREDICTIONS

In this section, we compare FL and WRM predictions under various conditions, including those under which the two predictions become identical, as well as in the near critical region, where FL and WRM predictions deviate most greatly from each other. However, FL and WRM predictions for experimentally measurable parameters such as the average aggregate size and ligand partitioning equilibrium constant are found to remain in remarkably good global agreement with each other, even in the near critical region.

Figure 1 illustrates the influence of p, ϵ_1, and ϵ_2 on the aggregate size distributions for a system with $n = 20$ binding sites and $c = 3$ nearest neighbors per site. More specifically, all of the ϵ_1 and ϵ_2 values in Figure 1 were chosen such that a symmetric (half-filled) distribution is attained when $p = p_{1/2} = 0.02$. The exact FL (points) and mean field WRM (solid curve)

dx.doi.org/10.1021/jp5003337 | J. Phys. Chem. B 2014, 118, 7878–7885

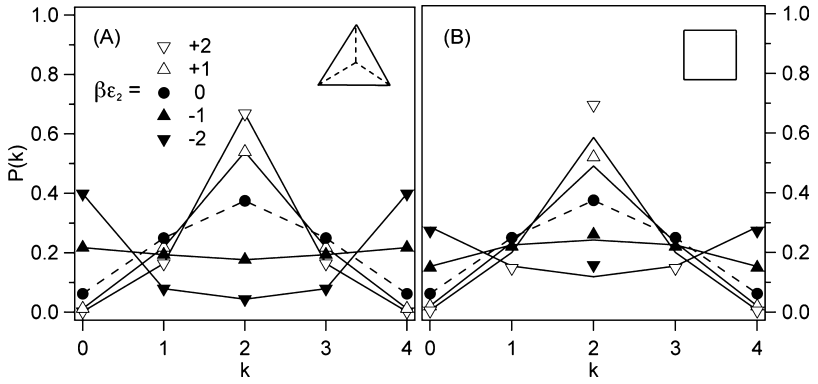


Figure 2. Comparison of aggregate size distribution predictions obtained when $p = p_{1/2}$ (so half the ligand binding sites are filled) for systems with either (A) tetrahedral ($n = 4$ and $c = 3$) or (B) square ($n = 4$ and $c = 2$) binding site geometries (as illustrated by the inset figures). The points and lines correspond to FL and WRM predictions, respectively, which agree exactly when $c = n - 1$ (A) as well as when the ligand−ligand interaction energy is zero (dashed curves with closed circular points).

predictions shown in Figure 1 were obtained using eqs 3 and 8 (as described in Sections 3 and 4 and the Supporting Information).

Figure 1A shows results pertaining to a system with no cooperative ligand−ligand interaction ($\epsilon_2 = 0$). These results illustrate how $P(k)$ changes shape and shifts to larger aggregate sizes with increasing ligand concentration $p = [L]v_0$. Note that when $p = 0.02$, the distribution is symmetric and unimodal, with $\langle k \rangle = n/2 = 10$. The exact agreement between the points and solid curves in Figure 1A confirms that the FL and WRM predictions are identical to each other for any such system (because, in the absence of cooperative interactions, all configurations with k bound ligands have precisely the same energy). Moreover, the distributions shown in Figure 1A are identical to those pertaining to a random mixture $P_0(k)$ (as given by eq 1) when p is replaced by p_0 (as given by eq 11).

Figure 1C shows the quite different influence of cooperative ligand−ligand interactions on the concentration dependence of $P(k)$. In this case a slight change in p produces a dramatic change from a distribution consisting primarily of free solutes, with $\langle k \rangle \sim 0$, to one consisting of essentially completely formed aggregates, with $\langle k \rangle \sim n = 20$. Notice that the aggregate size distributions may also now be bimodal but only over a narrow range of ligand concentrations $0.016 \leq p \leq 0.024$; outside of this range, the distributions are one-sided with over 60% of the population either in $P(0)$ or in $P(n)$, whereas at precisely $p = 0.02$, the bimodal aggregate size distribution is perfectly symmetric.

Figure 1B shows results obtained when both ϵ_1 and ϵ_2 are nonzero, and the value of $\beta\epsilon_2$ is chosen to correspond precisely to the critical point for the corresponding FL system. Note that the symmetric aggregate size distribution produced when $p = 0.02$ is now nearly flat; a smaller, less negative, value of ϵ_2 (or a higher temperature) would produce a unimodal distribution, and a larger, more negative, value of ϵ_2 (or a lower temperature) would produce a bimodal distribution.

The discrepancies between the FL (points) and WRM (solid curve) predictions in Figure 1 (B,C) illustrate the fact that cooperative ligand−ligand interactions may lead to deviations between FL and WRM predictions, because cooperative interactions generally lead to configurations which have the same value of k but different binding energies (as further discussed below). However, a special subclass of aggregation process for which $c = n - 1$ invariably produce identical FL and WRM predictions, because in such systems, all bound ligands are necessarily nearest neighbors (and so all configurations with a given value of k have the same binding energy). One example of such a system is an aggregate in which the ligand binding sites are arranged tetrahedrally, so $n = 4$ and $c = 3$, as shown in Figure 2A. In contrast, the results shown in Figure 2B pertain to a system in which the ligand binding sites are arranged on a square with $n = 4$ and $c = 2$. Thus, not all bound ligands are nearest neighbors, and so FL and WRM predictions are no longer identical.

Note that both the FL and WRM results shown in Figure 2 reveal the existence of a critical temperature below which the aggregate size distributions become bimodal, and above which they are unimodal. The precise value of the critical temperature depends not only on ϵ_2 but also on the values of both n and c, and the critical temperature is generally not the same for the FL and WRM models.

Figure 3 shows the half-filled aggregate size distribution predictions pertaining precisely to the critical temperature of the corresponding FL system, for aggregates ranging in size from $n = 4$ to $n = 20$, all of which have exactly $c = 3$ nearest neighbors per ligand binding site. At the FL critical temperature, there is clearly a significant difference between the FL (points) and WRM (solid curves) predictions for all but the $n = 4$ lattice. The difference between the FL and WRM increases as the difference between c and n increases, as a consequence of increasing fluctuations in the binding energy Ψ_k relative to $\epsilon(k) = \langle \Psi_k \rangle_0$.

The inset panel in Figure 3 shows how the critical temperature depends on n for both FL (points and dashed curve) and WRM (points and solid curve) models. The dashed and solid horizontal lines represent the corresponding critical temperature T^* in the macroscopic (infinite aggregate size) system, $\lim_{n\to\infty} T^*_{\mathrm{WRM}} = c/4 = 0.75$ and $\lim_{n\to\infty} T^*_{\mathrm{FL}} = 1/[2\ln(2 + \sqrt{3})] \approx 0.38$. In both cases, the critical temperatures approach the macroscopic limit with increasing n in a way that is reasonably well represented by the double exponential best fit curves shown in the inset panel of Figure 3 (but are not as well represented by a stretched exponential function, not shown). Although T^* depends on n, it is invariably the case that $T^* \sim 1$ and thus $k_B T_c \sim |\epsilon_2|$ (where $\epsilon_2 < 0$), which implies that the onset of bimodal aggregation behavior occurs when the ligand−ligand cohesive interaction energy exceeds the ambient thermal energy $k_B T$.

Although the discrepancies between the WRM and FL predictions for $P(k)$ are clearly quite large in the near critical

dx.doi.org/10.1021/jp5003337 | J. Phys. Chem. B 2014, 118, 7878−7885

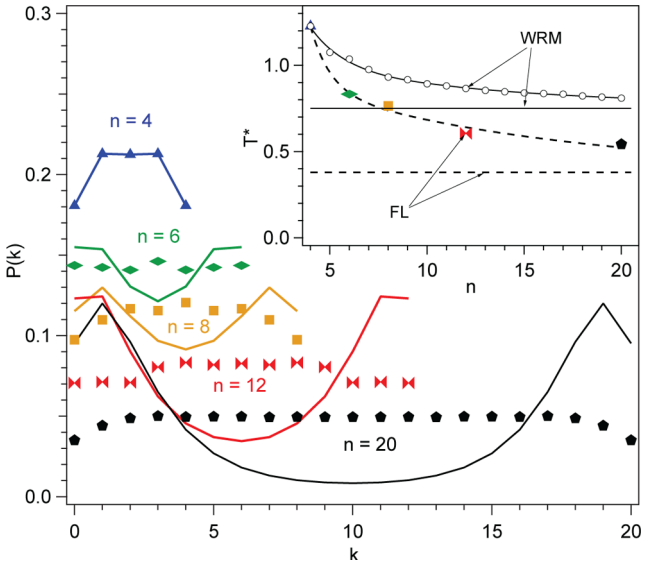


Figure 3. Comparison between FL (points) and WRM (solid curves) aggregate size distribution predictions for systems with various values of $4 \leq n \leq 20$, obtained at the FL critical temperature. The inset shows how the critical temperature $T^* = T_c(k_B/|\epsilon_2|) = (\beta_c|\epsilon_2|)^{-1}$ of the WRM (points and solid curves) and FL (points and dashed curve) aggregation models approach the corresponding macroscopic critical temperatures.

region, they do not have nearly as significant an influence on properties such as the average aggregate sizes $\langle k \rangle$ and ligand partitioning equilibrium constant K_p, as illustrated in Figure 4. The WRM (solid curves) and FL (points) results shown in Figure 4 pertain to an aggregate with a total of $n = 20$ ligand binding sites and $c = 3$ nearest neighbors per site. The predicted values of $\langle k \rangle$ and K_p are plotted as a function of the effective ligand concentration p, for various values of ϵ_1 or ϵ_2.

The two upper panels of Figure 4 show results obtained when varying ϵ_1 (with $\epsilon_2 = 0$), and the lower two panels in Figure 4 show results obtained when varying ϵ_2 (with $\epsilon_1 = 0$). The values of ϵ_2 in Figure 4C are chosen so as to have the same characteristic binding concentration as the corresponding results in Figure 4A. In other words, both sets of curves have half-filled aggregate size distributions, $\langle k \rangle = 10$, at the same free ligand concentrations p (whose values are given in the caption of Figure 4). The × symbol in Figure 4C marks the location of the WRM critical point in the macroscopic limit, which occurs at $\beta\epsilon_2 = -4/3$ and $p \approx 0.119$. The corresponding FL critical point in the macroscopic limit occurs at a $\beta\epsilon_2 \approx -2.63$ and $p \approx 0.019$.

The agreement between the WRM (curves) and FL (points) predictions shown in Figure 4 is clearly remarkably good. Note that the WRM and FL predictions are identical when $\epsilon_2 = 0$ (as shown in the upper two panels of Figure 4). On the other hand, when $\epsilon_1 = 0$ (as shown in the lower two panels of Figure 4), the WRM and FL predictions are no longer identical but clearly remain in quite good agreement with each other. This agreement is in part due to the fact that eq 14 holds for both the WRM and FL models. Thus, the WRM and FL predictions pertaining to the same values of ϵ_1 and/or ϵ_2 necessarily have the same characteristic binding concentrations and thus must cross $\langle k \rangle = 10$ at the same value of p.

The most significant deviations between the WRM and FL predictions shown in Figure 4 occur when $\epsilon_2 > 0$, in which case the average aggregate size $\langle k \rangle$ predictions obtained using the FL model invariably exceed those obtained using the WRM model. This discrepancy makes physical sense, as the WRM model assumes that the bound ligands are randomly distributed

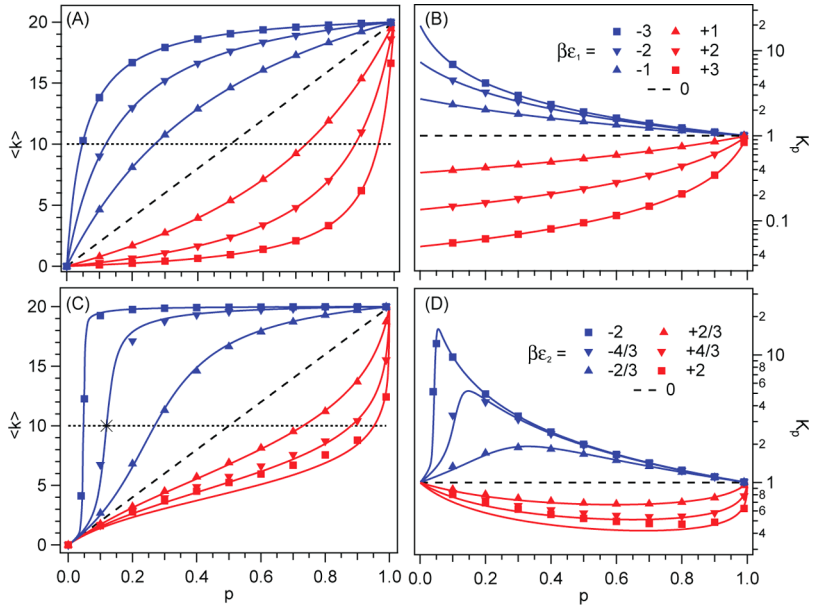


Figure 4. WRM (solid curves) and FL (points) predictions for $\langle k \rangle$ (left-hand panels) and K_p (right-hand panels), pertaining to aggregates with $n = 20$ binding sites and $c = 3$ nearest neighbors per site. Results obtained assuming various values of ϵ_1 (upper panels) and ϵ_2 (lower panels) are plotted as a function of the effective ligand concentration p. In panels (A) and (B), ϵ_1 is varied and $\epsilon_2 = 0$. In panels (C) and (D), ϵ_2 is varied and $\epsilon_1 = 0$. The dashed lines represent random mixing predictions (with $\epsilon_1 = \epsilon_2 = 0$). The horizontal dotted lines in panels (A) and (C) correspond to half-filled aggregate configurations. The corresponding characteristic aggregation concentrations, at which the WRM and FL predictions cross $\langle k \rangle = 10$, are the same in both panels (A) and (C); $p = 0.0474$, 0.119, 0.269, 0.5, 0.731, 0.881, and 0.953 when increasing either ϵ_1 or ϵ_2 from negative to positive values. The × symbol in panel (C) marks the location of the WRM critical point in the macroscopic limit.

dx.doi.org/10.1021/jp5003337 | J. Phys. Chem. B 2014, 118, 7878−7885

among the binding sites, whereas the exact FL results more realistically distribute the bound ligands so as to minimize repulsive ligand−ligand interactions. It is also interesting to note that when ϵ_2 becomes very large and positive, the WRM predictions approach $\langle k \rangle = 1$ (at all values of $0 < p < 1$), although the FL predictions typically approach a larger value of $\langle k \rangle$ in this limit. This difference between the WRM and FL predictions again stems from the fact that the exact FL results imply that there is no repulsive interaction energy between ligands as long as no ligands are in contact with each other, and thus when ϵ_2 becomes very large and positive, the aggregates will arrange so as to avoid any ligand−ligand contacts by filling approximately half of the binding sites. On the other hand, the WRM predictions must approach $\langle k \rangle = 1$ when ϵ_2 becomes arbitrarily large and positive, because a random arrangement of bound ligands will produce a large positive average energy whenever more than one ligand is bound.

6. SUMMARY AND CONCLUSIONS

Aggregation processes of biological and environmental importance are often driven by a delicate balance of entropic and energetic contributions. Here we have presented a theoretical strategy for incorporating such contributions by thermally averaging the configurations of a random mixture reference system. In the mean field (WRM) limit, our results reveal that entropic and energetic contributions are entirely decoupled from each other, as the aggregation entropy arises entirely from the random mixture reference system, whereas the aggregation energy arises entirely from the solute−ligand (ϵ_1) and ligand−ligand (ϵ_2) binding energies. The decoupling of entropic and energetic contributions in the WRM model is analogous to that arising in generalized van der Waals or first order thermodynamic perturbation theories of macroscopic fluids.[25−27]

More generally, an aggregation process in which entropic and energetic contributions are not entirely decoupled from each other may be described by using eq 3 to relate aggregate size distributions $P(k)$ in the system of interest to those in the corresponding random mixture reference system $P_0(k)$. We have illustrated how this more general strategy may be used to obtain exact aggregate size distribution predictions for systems in which the aggregation of ligands around a central molecule is described using finite two-dimensional lattice (FL) statistics. Our results reveal that FL and WRM predictions become identical in certain limits (when $\epsilon_2 = 0$ or $c = n - 1$) and are otherwise in reasonable agreement with each other, except near the bimodal to unimodal critical point. The WRM model provides a physically appealing approximation strategy as well as general analytical expressions for the influence of intermolecular interactions on aggregation equilibria.

We have highlighted the importance of cooperative (ligand−ligand) interactions in giving rise to bimodal aggregate size distributions and a critical point above which bimodal distributions become unimodal. This bimodal to unimodal transformation is the finite system analogue of vapor−liquid coexistence in macroscopic systems. In other words, a bimodal aggregate size distribution corresponds to the coexistence of aggregates with low and high ligand densities, analogous to macroscopic vapor and liquid phases. Bimodal aggregation is also closely related to micelle formation,[1−3] in which free monomers are in equilibrium with micelles with a high local density of bound monomers, whose aggregate size distribution evolves with increasing total monomer concentration in much the same way as the results shown in Figure 1C. More generally, our predictions are reminiscent of the behavior of various types of aggregating systems,[12−16] including some that display bimodal aggregation behavior.[4−6]

The influence of cooperative interactions on aggregation is closely related to biochemical allosteric regulation of enzyme activity and protein function, in which the binding of one ligand may either promote or inhibit the binding of one or more other ligands.[8−11] More specifically, a negative ligand−ligand interaction energy ϵ_2 leads to allosteric activation, and a positive ϵ_2 leads lead to allosteric inhibition of subsequent ligand binding. Although we have here assumed that all ligands are identical to each other, one could readily generalize our results to systems with more than one type of ligand and/or binding site.

The theoretical FL and WRM results that we have obtained may prove to be useful in the analysis and interpretation of computer simulation results, by directly comparing calculated and predicted aggregate size distributions $P(k)$, ligand partitioning K_p, and/or aggregation K_k equilibrium constants.[20] Such comparisons could be used to determine the effective direct (ϵ_1) and cooperative (ϵ_2) interaction energies for a given aggregation process and then to predict $P(k)$, K_p, and/or K_k beyond the regime in which the simulations were performed.

Comparing experimental results with FL and WRM predictions may require focusing on properties that are more readily measurable than $P(k)$ itself. For example, Raman spectroscopy has been combined with multivariate curve resolution (Raman-MCR) to obtain $\langle k \rangle$, K_1, and/or K_p values pertaining to the number of dangling OH bonds[21] or the number of iodide ions[28,29] in the first coordination shell of a hydrophobic molecule dissolved in water. As another example, thermodynamic measurements have been used to obtain estimates of the equilibrium constants pertaining to the partitioning of ions between a bulk aqueous solution and the coordination shells of various solutes.[30] Such experimental results may be compared with WRM and FL predictions to quantify direct ϵ_1 and cooperative ϵ_2 interaction energies and infer the corresponding aggregate size distributions $P(k)$. Note that the results shown in Figure 4 indicate that at low ligand concentration (small p), the slope of $\langle k \rangle$ as a function of p is entirely dictated by ϵ_1, whereas the corresponding second derivative may be used to determine ϵ_2. More generally, the experimentally measured ligand concentration dependence of $\langle k \rangle$, K_p, and/or K_k may be used to determine ϵ_1 and ϵ_2 and n (for an assumed c), although the value of c may be more difficult to independently determine, because eq 12 implies that changing c is equivalent to rescaling ϵ_2. We plan to demonstrate such experimental applications of the present theoretical results in forthcoming publications.

The key difference between the WRM and FL theoretical modeling strategies stems from the fact that the WRM model invokes a mean field approximation in assuming that all aggregates of size k have the same energy. The FL model, on the other hand, more accurately incorporates the exact distribution of energies arising from all configurations of ligands bound to an aggregate that is treated as a two-dimensional lattice with n binding sites and c nearest neighbors per site. Although there are significant discrepancies between the mean field WRM and exact FL predictions in the near-critical region, the corresponding average aggregate size and ligand partitioning equilibrium constant predictions are found to be in remarkably good global agreement with each other.

dx.doi.org/10.1021/jp5003337 | J. Phys. Chem. B 2014, 118, 7878−7885

■ ASSOCIATED CONTENT

Supporting Information

Additional FL results as noted in the text. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Email: bendor@purdue.edu. Tel.: (765) 494-5256.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

It is a pleasure to dedicate this paper to Jim Skinner, who has been an inspiration to all who know him as a person and as a scientist. Support for this work from the National Science Foundation (CHE-1213338 for D.B.A. and B.M.R.; CHE-0842022 for B.W.), as well as from the Purdue Research Foundation (for B.M.R.), is gratefully acknowledged.

■ REFERENCES

(1) Mukerjee, P.; Mysels, K. J. Critical Micelle Concentrations of Aqueous Surfactant Systems. National Bureau of Standards No. NSRDS-NBS36; U.S. Government Printing Office: Washington, DC, 1971.

(2) Israelachvili, J. N.; Mitchell, D. J.; Ninham, B. W. Theory of Self-Assembly of Hydrocarbon Amphiphiles into Micelles and Bilayers. *J. Chem. Soc., Faraday Trans. 2* **1976**, *72*, 1525−1568.

(3) Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd ed.; Wiley-Interscience: New York, 1980.

(4) Kumar, S.; Patel, H.; Patil, S. R. Test of Hofmeister-Like Series of Anionic Headgroups: Clouding and Micellar Growth. *Colloid Polym. Sci.* **2013**, *291*, 2069−2077.

(5) Zangi, R.; Berne, B. J. Aggregation and Dispersion of Small Hydrophobic Particles in Aqueous Electrolyte Solutions. *J. Phys. Chem. B* **2006**, *110*, 22736−22741.

(6) Raschke, T. M.; Tsai, J.; Levitt, M. Quantification of the Hydrophobic Interaction by Simulations of the Aggregation of Small Hydrophobic Solutes in Water. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5965−5969.

(7) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover Books on Press: New York, 1960; pp 124−146.

(8) Bu, Z. M.; Callaway, D. J. E. Proteins Move! Protein Dynamics and Long-Range Allostery in Cell Signaling. *Adv. Protein Chem. Struct. Biol.* **2011**, *83*, 163−221.

(9) Jaffe, E. K. Morpheeins—A New Structural Paradigm for Allosteric Regulation. *Trends Biochem. Sci.* **2005**, *30*, 490−497.

(10) Koshland, D. E.; Nemethy, G.; Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **1966**, *5*, 365−385.

(11) Monod, J.; Wyman, J.; Changeux, J. P. On the Nature of Allosteric Transitions—A Plausible Model. *J. Mol. Biol.* **1965**, *12*, 88−118.

(12) Li, P. L.; Banjade, S.; Cheng, H. C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; et al. Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* **2012**, *483*, 336−341.

(13) Lohman, T. M.; Ferrari, M. E. *Escherichia coli* Single-Stranded DNA-Binding Protein—Multiple DNA-Binding Modes and Cooperativities. *Annu. Rev. Biochem.* **1994**, *63*, 527−570.

(14) Subramanian, D.; Boughter, C. T.; Klauda, J. B.; Hammouda, B.; Anisimov, M. A. Mesoscale Inhomogeneities in Aqueous Solutions of Small Amphiphilic Molecules. *Faraday Discuss.* **2013**, *167*, 217−238.

(15) Jawor-Baczynska, A.; Moore, B. D.; Lee, H. S.; McCormick, A. V.; Sefcik, J. Population and Size Distribution of Solute-Rich Mesospecies within Mesostructured Aqueous Amino Acid Solutions. *Faraday Discuss.* **2013**, *167*, 425−440.

(16) Wilcox, D. S.; Rankin, B. M.; Ben-Amotz, D. Distinguishing Aggregation from Random Mixing in Aqueous *t*-Butyl Alcohol Solutions. *Faraday Discuss.* **2013**, *167*, 177−190.

(17) Erdemir, D.; Lee, A. Y.; Myerson, A. S. Nucleation of Crystals from Solution: Classical and Two-Step Models. *Acc. Chem. Res.* **2009**, *42*, 621−629.

(18) Gebauer, D.; Colfen, H. Prenucleation Clusters and Non-Classical Nucleation. *Nano Today* **2011**, *6*, 564−584.

(19) Davey, R. J.; Schroeder, S. L. M.; ter Horst, J. H. Nucleation of Organic Crystals. A Molecular Perspective. *Angew. Chem., Int. Ed.* **2013**, *52*, 2166−2179.

(20) Rankin, B. M.; Ben-Amotz, D. Analysis of Molecular Aggregation Equilibria using Random Mixing Statistics. *J. Phys. Chem. B* **2013**, *117*, 15667−15674.

(21) Davis, J. G.; Rankin, B. M.; Gierszal, K. P.; Ben-Amotz, D. On the Cooperative Formation of Non-Hydrogen Bonded Water at Molecular Hydrophobic Interfaces. *Nat. Chem.* **2013**, *5*, 796−802.

(22) Widom, B. Potential-Distribution Theory and the Statistical-Mechanics of Fluids. *J. Phys. Chem.* **1982**, *86*, 869−872.

(23) Lee, T. D.; Yang, C. N. Statistical Theory of Equations of State and Phase Transitions. 2. Lattice Gas and Ising Model. *Phys. Rev.* **1952**, *87*, 410−419.

(24) Hill, T. L. *Statistical Mechanics*; Dover Publications, Inc.: New York, 1956; pp 302−304.

(25) Widom, B. Intermolecular Forces and Nature of Liquid State. *Science* **1967**, *157*, 375−382.

(26) Chandler, D.; Weeks, J. D.; Andersen, H. C. van der Waals Picture of Liquids, Solids, and Phase-Transformations. *Science* **1983**, *220*, 787−794.

(27) Ben-Amotz, D.; Raineri, F.; Stell, G. Solvation Thermodynamics: Theory and Applications. *J. Phys. Chem. B* **2005**, *109*, 6866−6878.

(28) Rankin, B. M.; Ben-Amotz, D. Expulsion of Ions from Hydrophobic Hydration Shells. *J. Am. Chem. Soc.* **2013**, *135*, 8818−8821.

(29) Rankin, B. M.; Hands, M. D.; Wilcox, D. S.; Fega, K. R.; Slipchenko, L. V.; Ben-Amotz, D. Interactions Between Halide Anions and a Molecular Hydrophobic Interface. *Faraday Discuss.* **2013**, *160*, 255−270.

(30) Pegram, L. M.; Record, M. T. Thermodynamic Origin of Hofmeister Ion Effects. *J. Phys. Chem. B* **2008**, *112*, 9428−9436.