JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

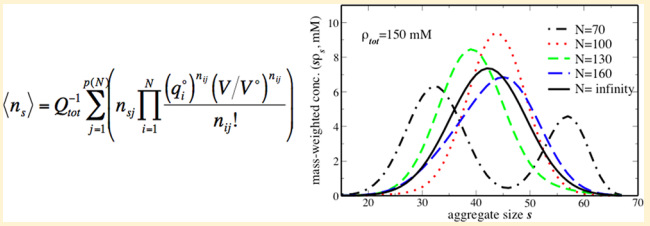
Accounting for Finite-Number Effects on Cluster Size Distributions in Simulations of Equilibrium Aggregation

James T. Kindt*

Department of Chemistry, Emory University, Atlanta, Georgia 30322, United States

Supporting Information

ABSTRACT: An approach is given to analyze aggregate size distributions obtained from simulations of a fixed number N of monomers undergoing reversible self-assembly. Equilibrium distributions are derived from size-dependent equilibrium association constants by appropriately weighted sums over all partitions of N monomers into aggregates. Conversely, equilibrium association constants can be obtained from an iterative fit to a finite-N equilibrium distribution. Model data for a micelle-forming system are used to show how results from simulations containing few micelles can yield infinite-N limiting distributions. A strategy is also suggested to exploit small-N effects on aggregate size distributions to enhance sampling of critical clusters in determination of nucleation free energy functions.



INTRODUCTION

Many phenomena of interest in the physical sciences involve the reversible aggregation of monomers (molecules, macromolecules, or colloids) into discrete aggregates. A range of computer simulation methods have been applied to study systems featuring reversible cluster formation in contexts such as the micellization of surfactants[1−5] and block copolymers,[6,7] oligomerization of amyloid peptides,[8−13] and vapor−liquid nucleation of atmospheric aerosols.[14,15] A common goal is to determine the equilibrium aggregate size distribution at a given total monomer concentration and temperature, for comparison with experiment.

Several successful strategies to generate aggregate equilibrium size distributions have been developed based on modeling stepwise growth of isolated aggregates, whether by free energy perturbation[16] or kinetics of monomer addition and dissociation.[17,18] In some cases, grand canonical ensemble simulations can be used to allow aggregate size distributions to be determined through equilibrium with a virtual reservoir of constant chemical potential.[14,19] Still, many simulations of self-assembly employ ensembles with a fixed number N of monomers free to aggregate into aggregates of any size up to N. These may included constant NVT ensembles; constant NpT ensembles, which, for approximately incompressible solvents, maintain nearly constant volume; or extended ensembles with multiple replicas but fixed N in each replica.[2] It is generally appreciated that the finite number of monomers used in the simulation may influence the aggregate size distribution; obviously, s-mers with s greater than N cannot be observed. Typically, if N is significantly greater than the typical aggregate size, it is assumed that the shape of the distribution will be maintained in the thermodynamic limit of large N at the same total monomer concentration. The purpose of this note is to examine this assumption and to provide methods to quantify and correct for associated artifacts. (Artifacts that may arise when the spatial dimensions of a cluster approach the simulation box size are a separate issue.) These tools should help researchers extrapolate from small-system simulations of equilibrium self-assembly to the large N limit, improving computational efficiency in particular for methods where scaling with system size is unfavorable.

THEORY

Following Ben-Shaul and Gelbart,[20] the statistical thermodynamics of an equilibrium aggregating system (with finite aggregates) can be treated within a framework in which an aggregate containing s monomers (s-mer) is treated as a distinct molecular species with partition function q_s, which reflects the internal free energy G_s of a single s-mer as $G_s = -k_B T \ln q_s$. The equilibrium constant for the formation of an s-mer can be associated with the relative values of the standard-state partition functions of the s-mer and the free monomer (q_s° and q_1°). To treat the case of small N, it is useful to review the assumptions (commonly derived in statistical mechanics textbooks[21,22]) that lead to the familiar equilibrium relationships in the limit of large N.

At a fixed total number of monomers N, many compositions $\{n_s\} = \{n_1, n_2, \ldots\}$ of s-mers can be defined such that

$$\sum_{s=1}^{N} s n_s = N \tag{1}$$

Received: August 3, 2012
Published: November 6, 2012

dx.doi.org/10.1021/ct300686u | J. Chem. Theory Comput. 2013, 9, 147−152

In number theory, these distinct compositions are termed "partitions" of the integer N, while the term "compositions" has a different mathematical definition. To avoid confusion with the term "partition function", as used in statistical thermodynamics, we will continue to use the term composition in the chemical sense. For instance, for $N = 3$ the possible compositions are $\{n_s\}_1 = \{0,0,1\}$, $\{n_s\}_2 = \{1,1,0\}$, and $\{n_s\}_3 = \{3,0,0\}$. The number of compositions is denoted $p(N)$, for example, for $p(3) = 3$. For large N, the number of compositions can be approximated[23,24] by

$$p(N) \approx \frac{1}{4N\sqrt{3}} e^{\pi\sqrt{2N/3}} \tag{2}$$

Neglecting interactions between aggregates (and, for solution-phase systems, factoring out contributions from pure solvent), the partition function of a system of s-mers with a defined number distribution $\{n_s\} = \{n_1, n_2, ...\}$ is given by

$$Q(\{n_s\}, V, T) = \prod_{s=1}^{N} \frac{q_s^{n_s}}{n_s!} \tag{3}$$

In the limit of large N, it is appropriate to maximize Q (or minimize Helmholtz free energy $A = -k_B T \ln Q$) with respect to $\{n_s\}$, leading to the most probable distribution, which can be easily shown to satisfy

$$\frac{n_s}{n_1^s} = \frac{q_s}{q_1^s} \tag{4}$$

Fluctuations away from the most probable partition are of the order $n_s^{1/2}$, and so, can be neglected in the limit of large N. The partition function q_s is linearly dependent on volume V under ideal-dilute conditions, allowing it to be related to the standard state as

$$q_s = \frac{V}{V^\circ} q_s^\circ \tag{5}$$

with V° as the volume per molecule under standard state concentration conditions. The equilibrium constant K_s for assembly of s monomers into one s-mer can be related to q_s° by multiplying both sides of eq 4 by $(V/V^\circ)^{s-1}$:

$$\frac{\rho_s/\rho^\circ}{(\rho_1/\rho^\circ)^s} = \frac{q_s^\circ}{(q_1^\circ)^s} = K_s \tag{6}$$

where the s-mer concentration $\rho_s \equiv n_s/V$ and ρ° is a standard state concentration. Where N is finite and small, a single composition cannot satisfy eq 4 for all s, and compositions with very different distributions may contribute to the ensemble. The precise ensemble average number of s-mers $\langle n_s \rangle$ requires consideration of all possible compositions of the system, each with a probability proportional to its partition function, as defined in eq 3:

$$\langle n_s \rangle = Q_{tot}^{-1} \sum_{j=1}^{p(N)} \left(n_{sj} \prod_{i=1}^{N} \frac{(q_i^\circ)^{n_{ij}} (V/V^\circ)^{n_{ij}}}{n_{ij}!} \right) \tag{7}$$

where n_{sj} is the number of s-mers present in the j^{th} composition. The total ensemble partition function Q_{tot} is

$$Q_{tot}(N, V, T) = \sum_{j=1}^{p(N)} Q(\{n_s\}_j, V, T) = \sum_{j=1}^{p(N)} \prod_{s=1}^{N} \frac{(q_s^\circ)^{n_{sj}} (V/V^\circ)^{n_{sj}}}{n_{sj}!} \tag{8}$$

Given a set of equilibrium association constants $\{K_s\}$ at some temperature, and one of the numerous algorithms[25] developed to enumerate all compositions $\{n_s\}_j$, one can calculate $\langle n_s \rangle$ as a function of N and V by applying eq 7.

RESULTS OF MODEL CALCULATIONS

In the present report, we have selected three representative modes of reversible association and investigated finite N effects on cluster statistics on each. In each case, we start by assuming a set of equilibrium cluster association constants K_s and proceed to determine the finite-N cluster size distributions under various conditions. In the first case, we furthermore demonstrate how the finite-N distribution can be used to infer K_s through an iterative procedure.

Micelle Formation. Aggregate distributions over a range in number of monomers N and at two fixed total monomer concentrations ρ_{tot} have been generated with eq 7 from a set of equilibrium constants $\{K_s\}$ characteristic of a micelle-forming system with behavior similar to that observed in a recent coarse-grained simulation study of surfactant aggregation.[4] The $N = \infty$ limiting size distribution generated from this model is peaked near $s = 40$, and its critical micelle concentration (CMC) is slightly below 20 mM. Specific values used for $\{K_s\}$ are given in the Supporting Information; K_s was set to zero for aggregate sizes above $s = 67$ and for $10 < s < 16$. On a single core of a 2.8 GHz Opteron node, the calculation of $\langle n_s \rangle$ could be completed in under 1 min for N up to 120, whereas the calculation for $N = 170$ (a sum over $\sim 3.2 \times 10^{10}$ compositions) took 3 h. Mass-weighted concentrations of aggregates as a function of s are shown in Figure 1 at several values of the total

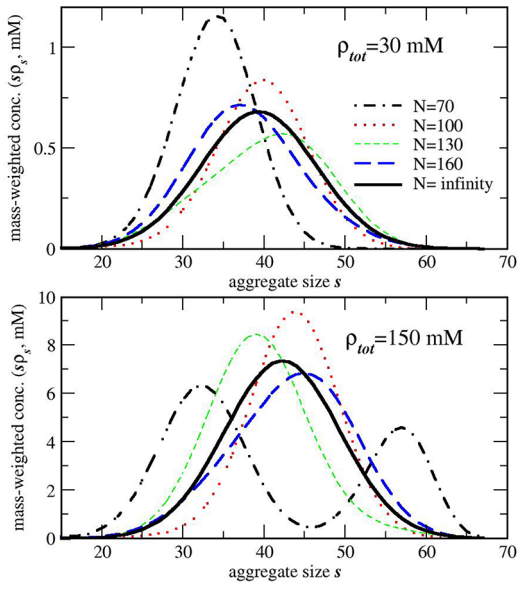


Figure 1. Equilibrium aggregate size distributions generated at two total monomer concentrations ρ_{tot} from a common set of association constants $\{K_s\}$ representing a micelle-forming system. Equation 7 was used for finite N and eq 6 for $N = \infty$.

dx.doi.org/10.1021/ct300686u | J. Chem. Theory Comput. 2013, 9, 147−152

system size N. The shape and peak value of the aggregate distribution depend nonmonotonically on N, and display a bimodal character at the higher concentration for $N = 70$.

To better understand these changes in peak maximum and peak shape, the equilibrium mean number of micelles $\langle n_{mic}\rangle = \sum_{s>15}\langle n_s\rangle$ and mean micelle size $\langle s\rangle = \sum_{s>15} s\langle n_s\rangle/\langle n_{mic}\rangle$ were calculated as a function of N and displayed in Figure 2A, B.

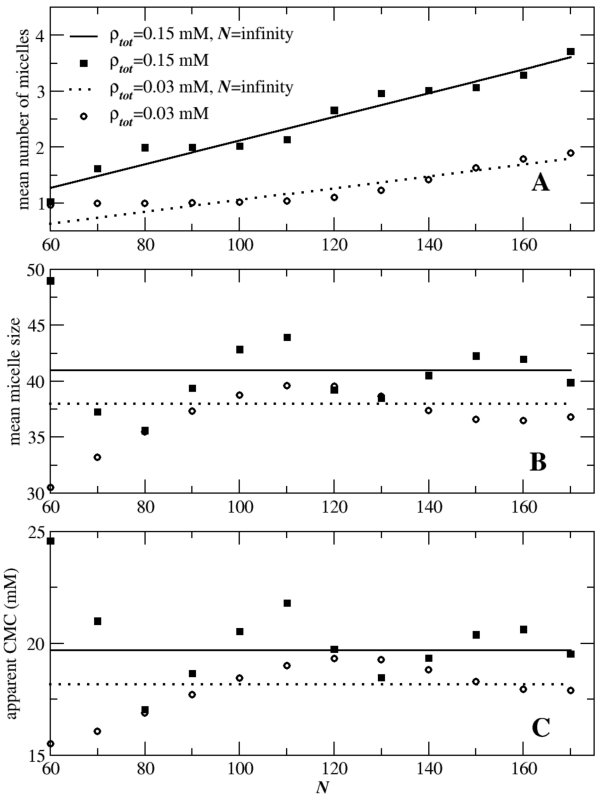


Figure 2. Averages over equilibrium distributions for systems represented in Figure 1, from finite N calculations (symbols) and limiting $N = \infty$ distributions (lines). Mean number of micelles for $N = \infty$ corresponds to mean number in a volume equal to the corresponding finite-N system.

Along with the micelle size distribution, an important value to be gleaned from simulations of self-assembled systems is the concentration of monomers that are either free or belong to small aggregates ("premicelles").[1] This concentration is often identified with the critical micelle concentration (CMC), the concentration below which large aggregates do not appear in solution (although variations on this definition are also in common use[18,26]), and is plotted in Figure 2C. The plot of $\langle n_{mic}\rangle$ vs N shows plateaus at integer n_{mic}. At the centers of these plateaus and at the midpoints of steps between plateaus, finite-N properties all coincide well with the limiting $N = \infty$ values. In the case of $N = 70$ at 150 mM total monomer concentration, where $<n_{mic}>$ is between 1 and 2, the bimodal distribution of micelle sizes in Figure 1 arises because the system can either accommodate one large micelle or two smaller micelles.

The results of Figure 2 are perhaps most remarkable in that, for this set of K_s, the fluctuations are not terribly large. At $N = 80$ and above, both mean micelle size and apparent CMC match their limiting $N = \infty$ values to within 10%, even though the system may contain as few as 1 or 2 micelles. Two factors contribute to the relative insensitivity of aggregation statistics to the choice of N: the moderately large number of free monomers in the system ($\langle n_1\rangle = 7.2$ at $N = 80$, $\rho_{tot} = 150$ mM) and the broad size range of stable aggregates. In a case where the intrinsic aggregate size distribution is more strongly peaked, the variation in the number of "leftover" monomers after the formation of a discrete number of micelles would more strongly influence the apparent CMC.

Whether the inaccuracy arising from finite N is large or not, it remains desirable to extract bulk equilibrium association constants K_s from observed aggregate number distributions $\{\langle n_s\rangle\}$ obtained at small N, so that the aggregate size distribution in the $N = \infty$ limit (corresponding to experiment) may be obtained. This process requires fitting a set of q_s° (which are related to equilibrium association constants via eq 6) to a set of $\langle n_s\rangle_{sim}$, where the average is taken over the production period of a simulation at constant volume and total monomer concentration. We assume that there is some consistent definition for the presence of a cluster and that all monomers can be defined to be a part of an s-mer with $s \geq 1$. To perform the fit, a first guess for $\{q_s^\circ\}$ is generated using eq 6 and substituted into eq 7 to calculate an aggregate size distribution $\{\langle n_s\rangle_{fit}\}$. The guess is refined by rescaling each q_s° by the following operation:

$$q_s^\circ \leftarrow q_s^\circ \times \left(\frac{\langle n_s\rangle_{sim}}{\langle n_s\rangle_{fit}} \frac{\langle n_1\rangle_{fit}^s}{\langle n_1\rangle_{sim}^s} \right)^a \tag{9}$$

to bring the $\langle n_s\rangle_{fit}$ closer to the actual value from simulation. The q_s° from each iteration can be used to generate both $\langle n_s\rangle_{fit}$ (using eq 6) and the limiting $N = \infty$ distribution of concentrations ρ_s (applying eq 6). The process is iterated until the mean square deviation between the fit and the input values is satisfactorily small, and the ρ_s distribution is unchanging. A choice of $a = 0.75$ was found to balance efficiency with stability of convergence.

Given a size distribution ρ_s produced at finite N from a known set of association constants $\{K_s\}$, the iterative procedure described above was successful in reproducing the original set $\{K_s\}$ and therefore in allowing extrapolation to the bulk $N = \infty$ distribution, at all values of N for the present model. An example of a successful fit after 20 iterations is given for $N = 110$ in Figure 3A. In general, accurate extrapolation to $N = \infty$ will obviously require that N be large enough such that aggregates of size $s > N$ can be neglected. Just as importantly, the distribution $\{\langle n_s\rangle_{sim}\}$ must be fully equilibrated. In simulations of self-assembled clusters, events that change the number of large clusters (nucleation, fusion, fission, complete dissociation) are typically rare, even if exchanges of monomers between clusters and their surroundings are frequent. The result of a simulation where such events are not observed can be generated from eq 6 and 7 using an appropriately restricted set of compositions $\{n_s\}$ with fixed number of micelles n_{mic}. Subsets of compositions with $n_{mic} = 2$ and $n_{mic} = 3$ were used to generate distributions for a simulation with 110 monomers in which two or three micelles persisted for the entire trajectory, growing and shrinking but never merging or splitting. These distributions are shown as solid black curves in Figure 3B and C. The peaks in these distributions correspond, as expected, to the main peak and shoulder present in the full distribution (Figure 3A). Best-fit $\{K_s\}$ distributions generated using the iterative procedure described above do not match the original equilibrium constants, and so, they produce incorrect

dx.doi.org/10.1021/ct300686u | J. Chem. Theory Comput. 2013, 9, 147−152

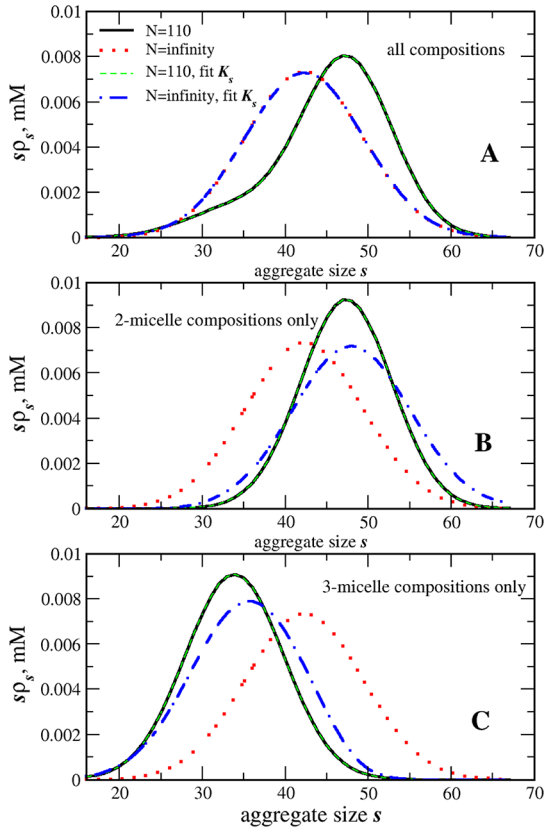


Figure 3. Solid curves: micelle size distributions generated using eq 6 from $\{K_s\}$ as in Figure 1 for N = 110 with a full set of compositions (panel A) or subsets with exactly 2 or 3 aggregates of size greater than s = 15 (panels B and C.) Dotted curves: size distributions at N = ∞ from $\{K_s\}$. Dashed curves: size distributions at N = 110 from $\{K_s\}_{fit}$ after 20 iterations of eq 6 and eq 8. Dot-dash curves: size distributions at N = ∞ from $\{K_s\}_{fit}$. Total monomer concentration ρ_{tot} = 150 mM for all cases.

extrapolations to the N = ∞ concentration distributions, shown as dot-dash curves in Figure 3B and C. This failure is not surprising; a distribution generated from a restricted set of compositions, where the number of large micelles is fixed, can only be sensitive to the relative values of K_s for large s but not to their absolute values.

Small Clusters. Some modes of association will lead to strongly peaked cluster size distributions at small s due to specificity and cooperativity in monomer interactions. One case that has been studied through NVT simulation[27] is the association of 1-hexanol in *n*-hexane under dilute conditions, where simulations containing N = 10, 30, or 50 hexanols (out of 1000 total solute + solvent, for mole fractions of 1%, 3%, and 5%) yielded a cluster size distribution peaked at s = 4. To address the question of the extent to which the cluster size distributions were influenced by finite N effects, a set of K_s for association into clusters of size s = 2 through s = 8 were generated to qualitatively reproduce the cluster distribution statistics from molecular simulations published in Figure 9 of ref 27. Equations 6 and 7 were used to generate cluster size distributions at the three compositions simulated, where both N and ρ_{tot} are increasing. Normalized cluster size distributions were calculated for each case and are plotted, along with the N = ∞ limiting distributions, in Figure 4.

Relative errors are greatest for N = 10 at the largest cluster sizes, with finite N effects reducing the concentration of

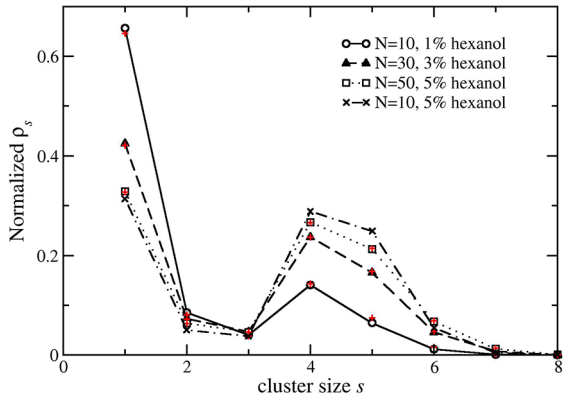


Figure 4. Cluster concentrations (normalized to the sum of concentrations of all clusters) vs cluster size modeled under four sets of finite-N conditions (lines with symbols) using $\{K_s\}$ based on simulation results[27] for 1-hexanol in *n*-hexane. N = ∞ distributions are shown with + symbols.

octamers by a factor of 2; these have a negligible contribution to the cluster size distribution, however. The greatest absolute errors in cluster size distribution values due to finite-N effects are seen for N = 10, where the fraction of clusters with s = 1 is overpredicted by 0.01 (0.65 vs 0.64) and the fraction of clusters with s = 5 is underpredicted by 0.01 (0.065 vs 0.075); in both cases, this systematic error falls within the statistical error bars of order ±0.02 reported for the normalized size distribution. Correcting for finite-N effects would therefore make little or no difference in the analysis of the results of ref 27. Hypothetically, if a 5% hexanol concentration had been achieved by reducing the system volume (and number of hexanes) while fixing the number of hexanols at N = 10, finite-N effects would have tended to inflate the fractions of tetramers and pentamers by 0.022 and 0.035 respectively, shown as "x" symbols in Figure 4.

Critical Nucleation Clusters. In the cases described, the primary goal is accurate characterization of the association constants for the most abundant clusters in the system. For study of reversible aggregate formation leading to nucleation of a new phase, determination of the association constants of the rarest clusters is essential (though not sufficient[14]) to predicting nucleation rates. By definition, "critical clusters" of size s_{crit} are found at a minimum of equilibrium ensemble cluster concentration vs s for a given total monomer concentration ρ_0 of interest. Once formed, a critical cluster will tend to either dissipate rapidly or irreversibly nucleate a droplet much larger than s_{crit}. Increasing the total monomer concentration ρ_{tot} will increase the rate of formation of clusters of the size s_{crit} of interest but will shift the equilibrium toward even larger clusters. For this reason, enhanced sampling strategies employing various forms of bias to produce a more uniform sampling of clusters of different sizes have been implemented to determine nucleation free energy (NFE) functions.[14,28,29]

In some cases, the shifts in cluster size distributions associated with small N may provide enough of a bias to make a simple NVT simulation, followed by the application of eqs 7–9, a viable method to obtain K_s for all $s \leq s_{crit}$. If N is chosen to be only slightly greater than s_{crit}, the simulated concentration ρ_{tot} can be set greater than $\rho_{tot,0}$ without producing an excess of clusters with $s \gg s_{crit}$. To obtain the best statistical sampling over all s, the optimal choice of N and ρ_{tot} will be the one with the shallowest minimum in $\langle n_s \rangle$ vs s. For illustration, a model NFE function and associated

dx.doi.org/10.1021/ct300686u | J. Chem. Theory Comput. 2013, 9, 147–152

association constants were selected to produce a maximum at s_{crit} = 19 of 32.4 k_BT:

$$\frac{\mathrm{NFE}(s)}{kT} = -\ln(K_s\rho_0^{s-1}) = 20s^{2/3} - 5s - 15 \quad (10)$$

as shown in Figure 5A. The equilibrium cluster size distribution was determined using eqs 6 and 7 for a system containing 22

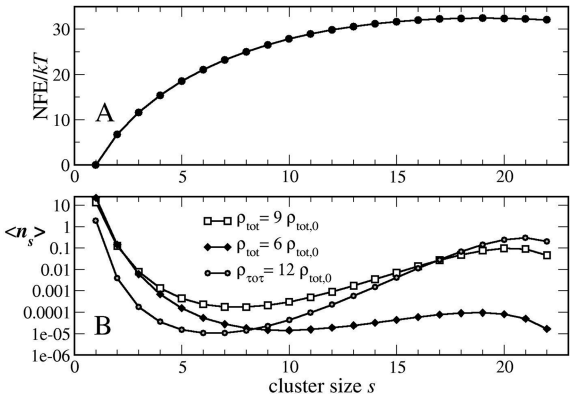


Figure 5. Cluster size distributions for example classical nucleation theory model. (A) Nucleation free energy function from which bulk association equilibrium constants were taken. (B) Ensemble mean numbers of clusters from simulations with N = 22 and total monomer concentrations of 6, 9, and 12 times the experimental concentration of interest.

monomers over a range of concentrations. A choice of ρ_{tot}= $9\rho_{tot,0}$ was close to optimal, with minimum $\langle n_s\rangle$ of 1.7×10^{-4} at s = 8, shown in Figure 5B; increasing or decreasing concentrations yielded lower cluster representation at either high s or low s, which would produce worse statistics in either case. Under these optimal conditions, sampling the full cluster size distribution will be limited by the low equilibrium levels of the 8-mer; these levels are, however, 10 orders of magnitude greater than the concentration of the 19-mer at $\rho_{tot,0}$. Depending on rates of monomer dissociation and association, crossing this barrier may still be frequent enough for reliable sampling of equilibrium $\langle n_s\rangle$ (from which K_s may then be inferred using eq 6−9). In practice, this approach is unlikely to be more efficient than existing enhanced sampling methods but could be useful in cases where these methods become complicated to implement, for example, in solution.

SUMMARY

The complete enumeration of compositions permits the identification and correction of systematic errors in equilibrium aggregate size distribution statistics due to the use of finite fixed number of monomers N, for aggregates containing N or fewer monomers. The iterative strategy outlined here should facilitate the calculation of equilibrium association constants and, therefore, the aggregate size distribution at large N, from simulations at minimal N. It will not, however, compensate for incomplete equilibration, particularly with respect to variations in the number of large aggregates. Given that computational cost often scales greater than linearly with system size, reduced system size will generally offer improved sampling efficiency. There may be cases when the use of small N introduces kinetic bottlenecks that would tend to produce poorer performance. On the other hand, in some cases, the careful choice of N and total monomer concentration may allow efficient calculation of equilibrium association constants for critical clusters involved in rare nucleation events from a direct, unbiased NVT simulation.

ASSOCIATED CONTENT

Supporting Information

Association constants K_s and corresponding $N = \infty$ equilibrium aggregate concentrations ρ_s at 30 and 150 mM total monomer concentrations for model micelle-forming system used in the present calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jkindt@emory.edu.

Notes

The authors declare no competing financial interest.
Implementations of the methods described are available upon request from the author as Fortran 90 codes.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation through Grant No. NSF-0911285.

REFERENCES

(1) Brodskaya, E. N. Computer simulations of micellar systems. *Colloid J.* **2012**, *74*, 154−171.
(2) Sanders, S. A.; Panagiotopoulos, A. Z. Micellization behavior of coarse grained surfactant models. *J. Chem. Phys.* **2010**, *132*, 114902.
(3) Kraft, J. F.; Vestergaard, M.; Schiøtt, B.; Thøgersen, L. Modeling the self-assembly and stability of DHPC micelles using atomic resolution and coarse grained MD simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1556−1569.
(4) LeBard, D. N.; Levine, B. G.; DeVane, R.; Shinoda, W.; Klein, M. L. Premicelles and monomer exchange in aqueous surfactant solutions above and below the critical micelle concentration. *Chem. Phys. Lett.* **2012**, *522*, 38−42.
(5) Turner, D. C.; Yin, F. C.; Kindt, J. T.; Zhang, H. L. Molecular dynamics simulations of glycocholate-oleic acid mixed micelle assembly. *Langmuir* **2010**, *26*, 4687−4692.
(6) Georgiadis, C.; Moultos, O.; Gergidis, L. N.; Vlahos, C. Brownian dynamics simulations on the self-assembly behavior of AB hybrid dendritic-star copolymers. *Langmuir* **2010**, *27*, 835−842.
(7) Mondal, J.; Yethiraj, A. Effect of secondary structure on the self-assembly of amphiphilic molecules: A multiscale simulation study. *J. Chem. Phys.* **2012**, *136*, 084902.
(8) Wu, C.; Lei, H.; Duan, Y. Formation of partially ordered oligomers of amyloidogenic hexapeptide (NFGAIL) in aqueous solution observed in molecular dynamics simulations. *Biophys. J.* **2004**, *87*, 3000−3009.
(9) Melquiond, A.; Boucher, G.; Mousseau, N.; Derreumaux, P. Following the aggregation of amyloid-forming peptides by computer simulations. *J. Chem. Phys.* **2005**, *122*, 174904.
(10) Li, D.-W.; Mohanty, S.; Irbäck, A.; Huo, S. Formation and growth of oligomers: A Monte Carlo study of an amyloid Tau fragment. *PLoS Comput. Biol.* **2008**, *4*, e1000238.
(11) Straub, J. E.; Thirumalai, D. Principles governing oligomer formation in amyloidogenic peptides. *Curr. Opin. Struct. Biol.* **2010**, *20* (2), 187−195.
(12) Urbanc, B.; Betnel, M.; Cruz, L.; Bitan, G.; Teplow, D. B. Elucidation of amyloid β-protein oligomerization mechanisms: Discrete molecular dynamics study. *J. Am. Chem. Soc.* **2010**, *132*, 4266−4280.
(13) Seo, M.; Rauscher, S.; Pomès, R.; Tieleman, D. P. Improving internal peptide dynamics in the coarse-grained MARTINI model: Toward large-scale simulations of amyloid- and elastin-like peptides. *J. Chem. Theory Comput.* **2012**, *8*, 1774−1785.

(14) Kathmann, S. M.; Schenter, G. K.; Garrett, B. C.; Chen, B.; Siepmann, J. I. Thermodynamics and Kinetics of nanoclusters controlling gas-to-particle nucleation. *J. Phys. Chem. C* **2009**, *113*, 10354−10370.
(15) Matsubara, H.; Ebisuzaki, T.; Yasuoka, K. Microscopic insights into nucleation in a sulfuric acid-water vapor mixture based on molecular dynamics simulation. *J. Chem. Phys.* **2009**, *130*, 104705.
(16) Yoshii, N.; Iwahashi, K.; Okazaki, S. A molecular dynamics study of free energy of micelle formation for sodium dodecyl sulfate in water and its size distribution. *J. Chem. Phys.* **2006**, *124*, 184901.
(17) Mohan, G.; Kopelevich, D. I. A multiscale model for kinetics of formation and disintegration of spherical micelles. *J. Chem. Phys.* **2008**, *128*, 044905.
(18) Burov, S. V.; Vanin, A. A.; Brodskaya, E. N. Principal role of the stepwise aggregation mechanism in ionic surfactant solutions near the critical micelle concentration. Molecular dynamics study. *J. Phys. Chem. B* **2009**, *113*, 10715−10720.
(19) Pam, L. S.; Spell, L. L.; Kindt, J. T. Simulation and theory of flexible equilibrium polymers under poor solvent conditions. *J. Chem. Phys.* **2007**, *126*, 134906.
(20) Ben-Shaul, A.; Gelbart, W. M. In *Micelles, Membranes, Microemulsions, and Monolayers*; Gelbart, W. M., Ben-Shaul, A., Roux, D., Eds.; Springer-Verlag: New York, 1994; pp 1−104.
(21) McQuarrie, D. A. *Statistical Thermodynamics*; University Science Books: Mill Valley, 1973.
(22) Widom, B. *Statistical Mechanics: A Concise Introduction for Chemists*; Cambridge University Press: Cambridge, 2002.
(23) Hardy, G. H.; Ramanujan, S. Asymptotic formulae in combinatory analysis. *Proc. London Math. Soc.* **1918**, *17*, 75−115.
(24) Ahlgren, S.; Ono, K. Addition and counting: The arithmetic of partitions. *Notice Am. Math. Soc.* **2001**, *48*, 978−984.
(25) Zoghbi, A.; Stojmenovic, I. Fast algorithms for generating integer partitions. *Int. J. Comput. Math.* **1998**, *70*, 319−332.
(26) Nagarajan, R.; Ruckenstein, E. Theory of surfactant self-assembly: A predictive molecular thermodynamic approach. *Langmuir* **1991**, *7*, 2934−2969.
(27) Stubbs, J. M.; Siepmann, J. I. Aggregation in dilute solutions of 1-hexanol in *n*-hexane: A Monte Carlo simulation study. *J. Phys. Chem. B* **2002**, *106*, 3968−3978.
(28) Kusaka, I.; Wang, Z.-G.; Seinfeld, J. H. Direct evaluation of the equilibrium distribution of physical clusters by a grand canonical Monte Carlo simulation. *J. Chem. Phys.* **1998**, *108*, 3416−3423.
(29) Chen, B.; Siepmann, J. I.; Oh, K. J.; Klein, M. L. Simulating vapor−liquid nucleation of *n*-alkanes. *J. Chem. Phys.* **2002**, *116*, 4317−4329.