Distribution Agreement

Signature:

________________________________ _______________

Xiaokun Zhang Date

Exploring Methods of Extraction of Cluster Free Energy from Small-N Molecular Dynamics Simulations

By

Xiaokun Zhang
Doctor of Philosophy
Chemistry

Dr. James T. Kindt, Ph.D.
Advisor

Dr. Joel M. Bowman, Ph.D.
Committee Member

Dr. Francesco Evangelista, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.

Dean of the James T. Laney School of Graduate Studies

Date

Exploring Methods of Extraction of Cluster Free Energy from Small-N Molecular Dynamics Simulations

By

Xiaokun Zhang
B.S., Xiamen University, 2015

Advisor: James T. Kindt, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in Chemistry
2021

Abstract

Exploring Methods of Extraction of Cluster Free Energy from Small-N Molecular Dynamics Simulations

By Xiaokun Zhang

Molecular dynamics simulations provide us a powerful tool to study the microscopic properties of reversible aggregates. However, the simulations often do not represent the behavior of a macroscopic system of aggregation due to the limit of system size. This dissertation will show (1) the establishment of fast and reliable analysis tools to extract thermodynamic properties from canonical ensemble simulations of small systems undergoing reversible aggregation and (2) example applications of these tools for micellizations of amphiphilic surfactants, including size-dependent cluster free energy, critical micelle concentration, micelle size, and predictions of X-ray scattering profiles. The tools are based on the "Partition-Enabled Analysis of Cluster Histograms" (PEACH) approach. A dramatic reduction in computational time for analysis is achieved through a strategy similar to the "selector variable method". With PEACH method and enthalpy change calculations, we explored the temperature-related cluster free energy and derived enthalpy change for micellization of sodium octanoate. The enthalpograms generated from simulation data were consistent with results from isothermal titration calorimetry experiments[1]. Molecular dynamics (MD) simulations of the zwitterionic surfactant octyl phosphocholine (OPC) in water have been performed with two force fields. Micelle size distributions from a number of trajectories were analyzed using the PEACH method to yield free energies of aggregation for premicelles and micelles over the full range from 2 to over 40 molecules. The dependence of free energy on aggregation number was consistent with the functional form derived from the “quasi-droplet” model of micellization. PEACH- BAR method was proposed to extract the free energy of aggregation vs. aggregate size systems with slow dissociation rate and low critical micelle concentration (CMC of approximately 1-2 mM) and applied to a united atom model of the surfactant dodecyl phosphocholine (DPC) in water. The new approach applies PEACH to a model with weakened attractions between aggregants, which allows sampling of a continuous range of cluster sizes, then recovers the free energy of aggregation under the original fully-attractive force field using the BAR free energy difference method.

Exploring Methods of Extraction of Cluster Free Energy from Small-N Molecular Dynamics Simulations

By

Xiaokun Zhang
B.S., Xiamen University, 2015

Advisor: James T. Kindt, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in Chemistry
2021

Acknowledgements

I would like to express my special thanks to my advisor, Prof. James T. Kindt, for his guidance, support and patience through all the years I spent at Emory University. It was a great honor for me to have an opportunity to work with him.

I would also like to thank my committee members, Prof. Joel. M. Bowman and Prof. Francesco Evangelista, for the time they set aside to be on my committee and their instructions they gave me.

For the years I spent in the Kindt research group, I appreciate the help from other group members. I would like to thank Dr. Lewen Yang, Dr. Lara Patel, Keon Reid, Dr. Ziwei Guo, Peiyao Wu, Mingning Zhu, Stephanie Hong for discussions, help and friendship.

I would like to express my gratitude to our collaborators Dr. Olivia Beckwith and Dr. Robert Schneider in Department of Mathematics at Emory University. The collaborations between Kindt group and Ken Ono group inspired the original idea of applying generating functions in PEACH method.

Lastly, I would like to thank my parents, Qing Zhang and Hui Duan, my grandparents and my fiancé. For all the years I study abroad, they always support me emotionally and encourage me to be a better person.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

The reversible aggregation of amphiphilic surfactants is an important phenomenon with a wide spectrum of applications in studies of detergents, membrane proteins[3] and nanomaterial synthesis[4]. Complete understanding of the thermodynamics of aggregation is difficult to achieve due in part to size polydispersity and sensitivity to conditions, and uncertainties in experimental techniques. Molecular dynamics simulations (MD) have long been a powerful tool to model the structures of disordered assemblies and extract thermodynamic properties for the dynamic equilibrium of aggregation. This work focuses on establishment of tools for extraction of the size-dependent free energy and related micelle statistics from molecular dynamics simulations. In particular, we are interested in the micelle formation process.

One challenge of conventional MD simulations on amphiphile assemblies is that to obtain size-dependent cluster free energy (CFE) through direct assessment of equilibrium association constants requires simulations with enough surfactants to allow several clusters to form and equilibrate with each other through exchange of monomers, enough solvent to provide a reservoir of free monomers (unimers), and a duration significantly greater than the time required for a monomer to dissociate from a micelle and diffuse to another one.[5] Statistical thermodynamics shows that applying the law of mass action directly to simulations of small systems results in distortions to the size distribution and unimer concentration.[6] Some advanced sampling methods[7, 8] were developed to obtain the cluster free energy, however, these methods involve the complications of applying a bias or sampling unphysical states. A method which provides free energy with sampling of a small number of surfactants in unbiased canonical ensemble will facilitate the investigation of micelle properties and applications.

We established “PEACH” method to address the problem with analysis of reversible aggregation simulations in canonical systems, and in particular, a core algorithm was brought up to handle the two-component system. In Chapter 2, we will show how the “PEACH” method was established and applied to the simulation of micelle formation of ionic surfactants with a coarse-grained model. The PEACH method makes rigorous estimations of bulk equilibrium association constants from unbiased, small-N simulations and eliminates the need for multiple clusters in the simulation, as it automatically corrects for the distortions in cluster size distribution that arise from a finite number of monomers. It is therefore effective so long as the need for a long enough trajectory and a large enough solvent reservoir to allow for significant fluctuations in cluster size to be observable. PEACH method facilitates the analysis of simulation results of reversible aggregates such as nucleation of salts, micelle formation and aggregation of amyloid peptides. In this work, PEACH method was applied to simulations of micellization of sodium octyl sulfate and has revealed consistent thermodynamic properties for large systems. This test case shows the efficiency and reliability of PEACH method. With PEACH method, we were able to explore the thermodynamics of micellization of various ionic surfactant with manageable computational cost.

The simple, one-component system of octyl phosphocholine (OPC) micellization was studied with direct application of PEACH method. OPC is a zwitterionic surfactant widely used for membrane protein solubilization. Another reason for the interest in OPC is that they share their headgroup structures with the well-studied bilayer-forming phosphatidyl glycero-phospholipids. In this work

(Chapter 3), we parameterized the force field so that it can work with SPC water (which is used in simulations of peptides and lipids). The force fields used for surfactant tails are based on the TraPPE model,[9] and so have been parameterized to fit alkane-alkane thermodynamic properties, but with modifications to interactions with water made to yield accurate free energies of hydration.

PEACH-BAR method was developed to handle micellization systems with rare monomer exchange events. For surfactants with low critical micelle concentrations (CMC) i.e. less than 10 mM, PEACH method is compromised as a result of the slow dissociation rate (relative to 100's of ns) and rare occurrences of unimers and premicelles in a moderately sized solvent bath (with 1000's of waters). To obtain good sampling of clusters of varying size, simulations need to be performed at low concentrations around the CMC; however, the large amount of solvents consumes excessive computational costs. Higher concentration simulations with small box sizes produce trajectories in which a micelle grows to encompass all monomers and remains in that state throughout the trajectory. In Chapter 4, an adapted version of PEACH method called PEACH-BAR method was developed to generate the cluster free energy and the micellization of dodecyl phosphocholine was analyzed with this method. The internal consistency of PEACH-BAR method was also examined by the umbrella sampling method.

The heat effect of micellizations was studied thoroughly with our techniques of PEACH method and enthalpy change calculations. One system we explored is the long-chain carboxylates. We explored the temperature-related cluster free energy and derived enthalpy change for micellization of sodium octanoate. We are interested in how the microscopic properties affect the overall heat effect. This study of enthalpogram of micellization based on molecular dynamics simulations is presented in Chapter 5.

The systems investigated as we have developed these tools have been fairly conventional zwitterionic and anionic surfactants, where we do not expect to find very new phenomena. This was intentional, as some qualitatively well-understood cases are useful tests when developing new methods. With the tools we have now to study micellization properties in this dissertation, we are interested in two directions in the future. The first is the aggregation of mixed peptide and surfactants can be studied with two-component PEACH method. Based on previous study, we hope to show how the micellization of OPC is influenced by adding peptides, and how the folding of the peptides is influenced by the presence of the surfactant. The second is the study of micellization of surfactants with wide applications and unconventional structure and properties. A bile salt, CHAPS, is a zwitterionic surfactant widely used to solubilize membrane proteins and revealed unique and controversial micellization properties of two CMC values. We attempted to study the micellization of this surfactant but observed quite slow dissociation in the simulations, which is the problem addressed by the PEACH-BAR method developed more recently.

Chapter 2 Extracting aggregation free energies of mixed clusters from simulations of small systems: application to ionic surfactant micelles

2.1 Introduction

A variety of physical and biophysical phenomena involve the reversible association of particles into aggregates of varying size. Under conditions where this association is effectively unbounded, these aggregates represent the pathway to a new phase, and information about their size-dependent free energy is essential to understanding the dynamics of nucleation of the new phase. [10-13] In other cases (as in micelle formation from some amphiphiles) the extent of cluster growth is limited by the structure of the aggregants. In those cases, the free energy of aggregation of clusters as a function of size and (for mixtures of particles) of composition is useful information to determine the cluster distribution at equilibrium, which will be sensitive to composition.[14, 15] Molecular simulations of micelles are commonly performed[16-31] and yield useful insights into their structure and their interactions with the proteins and small molecules they can solubilize. These simulations often face the challenge either of making assumptions about the number of amphiphiles contained within the micelle or of including enough molecules to form multiple aggregates and simulating for long enough that their size distribution can be assumed to have reached equilibrium. A recent overview[5] suggests that enough surfactant to form least five micelles should be included in the simulation for reliable estimation of sizes. Such large systems will be computationally expensive in atomistic simulations with explicit water.

Recently[32] we have reported a global fitting strategy that accounts for the breakdown of the law of mass action in determining cluster free energies from simulations of systems containing few clusters. Results from several trajectories with limited numbers of particles were combined to give a globally consistent free energy surface describing the growth of methyl t-butyl ether (MTBE) clusters along the pathway to nucleation of a bulk liquid phase from either the vapor or aqueous solution. The analysis of equilibrium cluster statistics in small-N systems relies on the consideration of all discrete ways in which the monomers of the system can be arranged into clusters,[6] which can be mapped onto the integer partitions of N. The number of these partitions scales as $\frac{1}{4N\sqrt{3}}\, e^{\pi\sqrt{2N/3}}$ for simple integers[33] and as $\frac{a}{N^{55/36}}\, \mathrm{e}^{bN^{2/3}+cN^{1/3}}$ (with a,b,c being explicit constants) for partitions of bipartite integers,[34, 35] which represent the combinations of clusters that can be formed from a 2-component mixture of aggregants. Explicit generation of these partitions, in the latter case in particular, becomes computationally impractical even for modest N. A new approach, derived using techniques commonly used in number theory, enables the evaluation of the appropriate sums over sets of partitions without explicitly generating those partitions, resulting in a more efficient algorithm whose computational cost scales better than N^3 for single component systems and better than N^6 for bipartite mixtures.

Here we will report on this updated "Partition-Enabled Analysis of Cluster Histograms" (PEACH) strategy, and demonstrate its application to finding the free energy surface for a coarse-grained model of the anionic surfactant sodium octyl sulfate (SOS). The stability of micellar clusters formed by ionic surfactants depends strongly both on the hydrophobic interactions of the surfactant tails and on the effects of counterion binding in reducing headgroup repulsions. The latter effect is influenced by the concentration of counterions, and to a lesser extent by the specific counterions used. These factors complicate efforts to model ionic micelles through molecular simulation for the prediction of important characteristics like the critical micelle concentration, micelle size, and degree of counterion condensation. Simulation studies[36, 37] have shown that the onset concentration for micellization (the critical micelle concentration, or CMC) can be estimated from an analysis of the free surfactant concentration, the average number of surfactants per micelle N, and the fractional neutralization α, which represents the ratio of the number of bound counterions to the number of surfactants in a micelle. Earlier, analysis of experimental results by Quina *et al.*[38] was developed on the basis that the formation of an ionic micelle from N surfactants and $\alpha \times N$ counterions can be treated using an equilibrium expression using the law of mass action. A similar assumption was made by Burov and Shchekin[39] who analyzed micelle size and charge distributions in a coarse-grained model simulations to find a two-dimensional free energy surface in the $N_{\mathrm{surfactant}}$ / $N_{\mathrm{counterion}}$ plane from simulations of ~500 coarse-grained ionic surfactants. Here we show how a similar surface can be derived from simulations of small systems containing 1 or 2 micelles. The solvent-free "Dry Martini" coarse-grained forcefield[40] was used to allow efficient sampling of these systems at low computational cost, and also to allow long enough simulations for equilibrium distributions to converge in larger simulations. This provides confirmation that statistics generated in small-N simulations can be used to make accurate predictions for a large-N system. The form of the free energy surface itself is compared with a simple phenomenological model and used to predict the concentration dependence of aggregate properties. Finally, we explore the effects of varying the criteria for defining a cluster, both on the quality of the fit and on the predicted concentration-dependent micellization behavior.

In contrast to studies of micellization free energy that use external biasing potentials[41-44] or alchemical slow-growth schemes,[45] the PEACH approach relies on the ability of unbiased simulations to produce equilibrated distributions over the course of a trajectory. An advantage is that it does not require any special algorithms in the generation of the input data. A downside is that the trajectories must produce an equilibrated distribution within an accessible amount of simulation time, which will only be possible when the rate of monomer exchange is not too slow relative to the trajectory time. We expect that this approach will be applicable to the analysis of a wide range of micelle systems represented by computationally cheap CG models. Additionally, PEACH should be useful to analyze simulations performed using atomistic models of ionic surfactant solutions for systems with CMC in the 30 mM or higher range. In these cases, the analysis will offer detailed predictions of not only the CMC but also the dependence of micelle size and charge distributions on surfactant and counterion concentrations.

2.2 Theory and algorithm development

2.2.1 Statement of the problem and statistical thermodynamic model

The canonical partition function for a system containing a given number of chemical entities translating freely in a volume *V*, with interactions between the entities neglected, is given by:

$$Q(N_a, N_b, N_c \ldots, V, T) = \frac{q_a(V,T)^{N_a}}{N_a!} \times \frac{q_b(V,T)^{N_b}}{N_b!} \times \frac{q_c(V,T)^{N_c}}{N_c!} \ldots \tag{1}$$

with q_a the single-molecule partition function, which can be related to some standard state partition function $q_a^{\ominus}(T)$ by:

$$q_a(V,T) = Vc^{\ominus} q_a^{\ominus}(T) \tag{2}$$

with $c^{\ominus}$ the standard-state concentration. When the entities *a,b,c…* can interchange chemically, the probability of finding a chemical compositions N_a, N_b, N_c … is proportional to the corresponding value of *Q*. The maximization of *Q* with respect to the amounts of these components (subject to conservation of matter, which introduces constraints derived from stoichiometric coefficients) allows determination of the most probable chemical composition. In the limit of large N_a, N_b… this constrained maximization produces the Law of Mass Action as taught in elementary chemistry, which relates an equilibrium quotient of product concentrations over reactant concentrations to a constant related to the standard partition functions $q^{\ominus}$.

Here we are concerned with the case where these entities are clusters formed by combining integer numbers of one or more components. Conservation of mass then dictates that, for a single component system containing N identical monomers, the possible chemical compositions of the system are given by sets $\{N_i\}$ of numbers of clusters containing i monomers such that

$$\sum_{i=1}^{N} iN_i = N \tag{3}$$

In number theory, such a set of integers is called an *integer partition*, and is commonly represented as

$$\lambda = (1^{m_1}2^{m_2}3^{m_3} \dots k^{m_k}) \tag{4}$$

with m_k taking the place of N_i, representing the multiplicity of the integer k. The number being partitioned is represented as $|\lambda|$ and termed the *size* of λ.

$$|\lambda| = \sum_{k=1} km_k \tag{5}$$

If the values of partition functions q_k are known, then for a given total number of monomers $|\lambda| = N$ the probability that these monomers will assemble into a set of clusters λ' that satisfies eq. 3, (a condition that is denoted $\lambda' \vdash N$) will then be given by:

$$P(\lambda') = \left(\prod_{k=1}^{N} \frac{q_k^{m'_k}}{m'_k!}\right) \Bigg/ \sum_{\lambda \vdash N} \left(\prod_{k=1}^{N} \frac{q_k^{m_k}}{m_k!}\right)$$

(6)

where the denominator is the overall partition function $Q = Q(N,V,T)$ obtained by summing the weights of all possible chemical compositions. The equilibrium ensemble average number $<m_j>$ of clusters of size j is then obtained by the sum over chemical compositions, each weighted by m_j and its probability:

$$\langle m_j \rangle = Q^{-1} \sum_{\lambda \vdash N} \left(m_j \prod_{k=1}^{N} \frac{q_k^{m_k}}{m_k!} \right).$$

(7)

If all partitions of N can be generated, then it is straightforward to calculate this average for each cluster size k and obtain an equilibrium cluster size distribution. For small enough N this will deviate considerably from the predictions of the Law of Mass Action.

For a system that forms clusters from two distinguishable components A and B, we can identify the number of clusters formed from j monomers of type A and k monomers of type B as $m_{j,k}$. The possible values of $m_{j,k}$ such that

$$\sum_{j=1}^{N_A} \sum_{k=0}^{N_B} j \times m_{j,k} = N_A \text{ and } \sum_{j=0}^{N_A} \sum_{k=1}^{N_B} k \times m_{j,k} = N_B$$

(8)

correspond to the set of partitions of the bipartite number (N_A,N_B). The number of bipartite partitions as a function of N_A and N_B has been studied[34, 35] and found to be significantly larger than the number of regular integer partitions of N_A+N_B. While algorithms to generate integer partitions have been described,[46] to our knowledge no such algorithm for bipartite partitions has been published; an algorithm that we have developed for this purpose is presented in the Supporting Information. After using this algorithm for preliminary calculations employing explicit enumeration of bipartite partitions to solve eq. 7 to obtain $<m_{j,k}>$ from $q_{j,k}$ for a two-component system containing more than 40 total monomers, we found it necessary to seek a more efficient strategy than "brute-force" generation of all partitions.

2.2.2 Application of Faà di Bruno's formula to canonical ensemble averages

An alternative route to evaluating $<m_j>$ in eq. 7 that does not rely on enumeration of partitions can be obtained via a useful formula attributed to the mid-nineteenth century Italian priest and mathematician, Francesco Faà di Bruno,[47] which arises as a generalization of the chain rule in calculus.[48] A slightly simplified version of Faà di Bruno's identity as a sum over all partitions λ is achieved for some set of cluster partition functions $q_1, q_2, q_3 \ldots$ as

$$\exp\left(\sum_{n=1}^{\infty} q_n z^n\right) = \sum_{\lambda \in \mathcal{P}} z^{|\lambda|} \frac{{q_1}^{m_1} {q_2}^{m_2} {q_3}^{m_3} \ldots\ {q_k}^{m_k}}{m_1!\, m_2!\, m_3! \ldots m_k!}, \tag{9}$$

assuming the series inside the exponential function converges (in this case, the right-hand side converges too). If we take z to represent the monomer activity $e^{\beta\mu}$ then the right-hand side of eq. 9 represents the grand canonical partition function, $\Xi(\mu,V,T)$,

$$\Xi(\mu, V, T) = \sum_{N} z^N Q(N, V, T), \tag{10}$$

through substitution of the expression for Q that appears in eq. 6. For a simple ideal gas, (i.e. when q is nonzero only for k=1) eq. 9 reduces to a known[49] relationship from statistical thermodynamics: $\Xi = e^{zq}$. The proof of eq. 9 follows from the well-known *multinomial theorem*, re-written as a sum over partitions λ in the set $\mathcal{P}_{[k]}$ of partitions whose parts are all $\leq k$, having length $\ell(\lambda) = n$:

$$(q_1 + q_2 + q_3 + \cdots + q_k)^n = n! \sum_{\substack{\lambda \in \mathcal{P}_{[k]} \\ \ell(\lambda)=n}} \frac{{q_1}^{m_1} {q_2}^{m_2} {q_3}^{m_3} \ldots\ {q_k}^{m_k}}{m_1!\, m_2!\, m_3! \ldots m_k!} \tag{11}$$

If we let k tend to infinity, assuming the infinite sum $q_1 + q_2 + q_3 + \cdots$ converges, the series on the right becomes a sum over all partitions of length n. Then dividing both sides of eq. 11 by $n!$ and summing over $n \neq 0$, the left-hand side yields the Maclaurin series expansion for $\exp(x)$, and the right side can be rewritten as a sum over all partitions:

$$\exp(q_1 + q_2 + q_3 \ldots + q_k) = \sum_{\lambda \in \mathcal{P}} \frac{{q_1}^{m_1} {q_2}^{m_2} {q_3}^{m_3} \ldots\ {q_k}^{m_{|\lambda|}}}{m_1!\, m_2!\, m_3! \ldots m_{|\lambda|}!} \tag{12}$$

Now, the formula in eq. 9 follows from substituting $q_k z^k$ for q_k in eq.11. So we can view Faà di Bruno's formula as a generating function for coefficients of certain partition-theoretic sums involving ${q_1}^{m_1} {q_2}^{m_2} {q_3}^{m_3} \ldots\ {q_k}^{m_{|\lambda|}} \,/\, m_1!\, m_2!\, m_3! \ldots m_{|\lambda|}!$. Choosing special substitutions for z and q_n, one obtains surprising partition-theoretic formulas involving functions arising in number theory, such as the Riemann zeta function.[50]

A concise expression for the expected number of j-mers follows from eq. 9; combining the observation

$$\sum_{\lambda\vdash N} m_j \prod_{k=1}^{N} \frac{q_k^{m_k}}{m_k!} = q_j \sum_{\lambda\vdash N} \frac{{q_1}^{m_1}{q_2}^{m_2}{q_3}^{m_3} \ldots {q_j}^{m_j-1} \ldots {q_k}^{m_k}}{m_1!\, m_2!\, m_3! \ldots (m_j-1)! \ldots m_k!} = q_j Q(N-j)$$

(13)

with eq.7, we find

$$\langle m_j \rangle_N = \frac{q_j Q(N-j)}{Q(N)}.$$

(14)

Note that the canonical partition functions Q in the numerator and denominator of eq.14 correspond to the same volume V.

To find the cluster size distribution by evaluating eq.14, the canonical partition function values Q for various N need to be extracted from eq. 9, which provides a sum over all N. Examining the representation of the grand partition function as a power series in z, we see that the coefficient of the ith term is $Q(i)$:

$$\exp\left(\sum_{k=1}^{\infty} q_k z^k\right) = \sum_{i=0}^{\infty} z^i Q(i)$$

(15)

The desired coefficient, $Q(N)$, can be isolated by taking the Nth derivative with respect to z of both sides, dividing by $N!$, and setting $z=0$:

$$Q(N) = \frac{1}{N!}\left(\frac{\partial^N}{\partial z^N}\exp\left(\sum_{k=1}^{N} q_k z^k\right)\right)_{z=0}$$

(16)

Note that the summation on the right-hand side has been truncated at N, which is permissible since any contributions from $k > N$ in the exponential cannot contribute to the Nth term in the power series. The overall strategy of the algorithm is similar to that of the Darwin-Fowler "selector variable" method[51] for determination of the mean occupation of particular quantized states in the canonical ensemble. The selector variable used here is the activity z, which allows us to pick out the contribution of terms arising from exactly N particles to the expression in eq. 15.

Evaluation of the derivative in eq. 16 is straightforward. Defining

$$\exp\left(\sum_{k=1}^{N} q_k z^k\right) = f_0(z)\exp[f_{0'}(z)]$$

(17)

with $f_0(z) = 1$, the nth derivative of $f_0(z)\exp[f_{0'}(z)]$ can be expressed recursively as a function $f_n(z)\exp[f_{0'}(z)]$ (where f_n and $f_{0'}$ are polynomials) such that

$$f_n(z) = \left(\frac{\partial f_{n-1}(z)}{\partial z} + f_{n-1}(z)\frac{\partial f_{0'}(z)}{\partial z}\right).$$

(18)

Repeated application of eq. 18 allows the evaluation of f_N, providing $Q(N)$ as $f_N(0)/N!$ by eq. 16.

Because the functions $f_i(z)$ are polynomials, the computational expense for each step will scale with the (at most) N^2 products of coefficients of f_{n-1} with coefficients of the derivative of $f_{0'}$. (Coefficients for terms z^{N-n} of order greater than N-n need not be calculated as they will not influence the final answer.) The computational cost for the full calculation, starting with n=1 and proceeding to n=N-1 derivatives to obtain $Q(N)$ will be of order N^3 (or, lower by the proportion of q_i that are zero-valued). The values of $Q(N\text{-}s,V,T)$ for s from 1 to N can be saved and stored for the final calculation of $<m_j>$ using eq. 14 at negligible additional computational cost.

Using a multivariable generalization of Faà di Bruno's formula[52] and extending the derivations shown above to a two-component system containing N_A particles of type A and N_B particles of type B, we obtain:

$$\langle m_{j,k}\rangle_{N_A,N_B} = \frac{q_{j,k}Q(N_A - j, N_B - k)}{Q(N_A, N_B)}$$

(19)

$$Q(N_A, N_B) = \frac{1}{N_A!\,N_B!}\left(\frac{\partial^{N_A}}{\partial z_A}\frac{\partial^{N_B}}{\partial z_B}\exp\left(\sum_{j=1}^{N_A}\sum_{k=1}^{N_B} q_{j,k}{z_A}^j{z_B}^k\right)\right)_{z_A,\,z_B=0}$$

(20)

The approach to evaluating the right-hand side of eq. 20 is similar to the 1-dimensional case. Defining,

$$\exp\left(\sum_{j=1}^{N_A}\sum_{k=1}^{N_B} q_{j,k}{z_A}^j{z_B}^k\right) = f_{0,0}(z_A, z_B)\exp[f_{0'}(z_A, z_B)]$$

(21)

we see that taking partial derivatives of the right-hand side with respect to z_A involves combining derivatives of the polynomials of z_A and z_B:

$$f_{j,k}(z_A, z_B) = \frac{\partial f_{j-1,k}(z_A, z_B)}{\partial z_A} + f_{j-1,k}(z_A, z_B) \frac{\partial f_{0,}(z_A, z_B)}{\partial z_A} \tag{22}$$

As there will be at most (N_A+1) × (N_B+1) coefficients to calculate in one of these polynomial expressions, the computational cost of each partial derivative step will be of order $\sim(N_A \times N_B)^2$. To reach each combination of partial derivatives (and find all $m_{j,k}$ cluster sizes will then require computational cost of order $\sim(N_A \times N_B)^3$ or less. The computational savings relative to the explicit enumeration method are very significant. A calculation of a single set of $<m_{j,k}>$ for a system with $N_A = N_B = 20$ required over 2 hours by explicit enumeration, using the algorithm described in Supporting Information; calculation of $<m_{j,k}>$ over *all* 12 sets ranging up to $N_A = N_B = 60$ required less than one minute.

2.3 Simulation and analysis details

2.3.1 Simulation parameters.

Constant-NVT molecular dynamics simulations of sodium octyl sulfate (SOS) were performed with version 4.6.5 of the Gromacs package[53] using the solvent-free coarse-grained "Dry Martini" model.[40] The parameters and settings used by Wang and Larson[54] to model sodium dodecyl sulfate aggregation were adapted by removing one coarse-grained bead from the tail, leaving a three-site octyl sulfate chain (OS) consisting of an anionic sulfate bead and two neutral hydrophobic alkyl tail beads, C1 and C2. Forces of van der Waals were smoothly shifted to zero between 0.9 and 1.2 nm. Following Wang and Larson,[54] long-range electrostatics were treated by the Particle-Mesh Ewald (PME)[55] approach using a dielectric constant of 150 instead of the experimental value near 80. The integration of the equations of motion were performed with a time step of 40 fs and using the velocity rescaling algorithm[56] to maintain a temperature of 310 K with a time constant of 4 ps. Two series of simulations, one at a concentration of 67.8 mM and the other at 114.9 mM, were performed using increasing numbers of SOS from 10 to 60. A single large simulation containing 560 SOS was also performed to test the ability of the free energies extracted from the smaller simulations to predict behavior in a large system. Table I shows compositions and durations for all simulations performed.

Table 2.1 SOS cluster aggregation simulation parameters

Trial	N_{SOS}	Box Size/ nm^3	[SOS]/mM	Duration/ns
1	10	254.9	67.8	500
2	20	509.9	67.8	500
3	30	764.8	67.8	500
4	40	1019.8	67.8	500
5	50	1274.7	67.8	500
6	60	1529.7	67.8	500
7	10	159.3	114.9	500
8	20	318.7	114.9	500
9	30	478.0	114.9	500
10	40	637.4	114.9	500

11	50	796.7	114.9	500
12	60	956.0	114.9	500
13	560	8923.1	104.3	2000

2.3.2 Cluster definition. The "dressed" ionic aggregate model[39] is used in analyzing the trajectories to determine the distribution of clusters. The model considers a micelle containing a specific number of surfactants and a specific number of associated counterions to be a distinct chemical species, *e.g.* OS_jNa_k. Two OS chains are considered to be neighbors if the distances between either of their tail beads are lower than a cut-off (0.65 nm if not otherwise specified). All OS chains that can be related by a series of neighbor interactions are considered to be part of the same micelle. Association of sodium counterions to the micelles is treated somewhat differently. Na^+ beads that are within a cut-off distance (1.5 nm if not otherwise specified) from a sulfate site on an OS chain are considered to be associated with that OS chain's micelle. In cases where a counterion bridges two distinct micelles (*i.e.*, micelles that have no OS chains in common), that counterion is treated as dividing its time evenly among all the OS. For instance, if at a certain time point a single Na^+ is within binding range of three sulfate sites, two of which are associated with one micelle and the third sulfate associated with a second micelle, it will be treated as though it spends 2/3 of its time associated with the first micelle and 1/3 of its time with the second for purposes of calculating the distribution of micelle sizes. Thus, by construction, a set of micelles that are connected only by counterion bridges will be treated as independent micelles that dynamically exchange their shared counterions.

2.3.3 Global fitting of cluster histograms using PEACH

As in previous work[32] an iterative global fitting procedure was used to find a set of equilibrium cluster association constants that best reproduces the distribution of micelle sizes and counterion compositions observed in the several simulations. The procedure requires both the cluster size frequencies and their standard deviations. The SOS simulations with the numbers of monomers ranging from 10 to 60 took less than 2 ns to reach equilibrium, thus the cluster size distributions were calculated after excluding the first 5 ns period. Each trajectory was split into segments of 30 ns and the standard deviation for the frequency of the each cluster size among those blocks was calculated. To account approximately for the effect of free volume on cluster statistics,[57] and following procedure previously used in MTBE cluster simulations,[32] an effective free volume obtained by deducting the volume of the estimated SOS volume ($N_{\text{tot}}*V_{\text{monomer}}$) from the total volume of the simulation box, was used in eq. 2 to relate $q_{j,k}^{\ominus}$ to $q_{j,k}$. Accordingly, concentrations reported in moles/L should be considered as moles per liter of solvent volume (*i.e.* , corresponding more closely to molality than molarity). For this purpose, the volume of one SOS monomer used was 0.474 nm^3, estimated from the sum of volumes calculated for the LJ parameters of its four sites; changing this volume by a factor of 2 yielded only slight quantitative changes to the results given below.

Starting with an initial guess for the standard cluster partition functions $q_{j,k}^{\ominus}$ derived from the law of mass action, the algorithm based on Faà di Bruno's formula described in section IIB was used to generate the average number of micelles of each type in each simulation, then iteratively adjusted to improve the fit to the simulation data. Equilibrium association constants, which should be applicable in the limit of large *N*, can then be related to the optimized standard cluster partition functions $q_{j,k}^{\ominus}$:

$$K_{j,k} = \frac{q_{j,k}^{\ominus}}{\left({q_{1,0}^{\ominus}}^{j} {q_{0,1}^{\ominus}}^{k}\right)} \tag{23}$$

We adopt the acronym PEACH, for "Partition-Enabled Analysis of Cluster Histograms" for this approach to obtaining cluster formation equilibrium constants from a set of cluster distribution histograms obtained at finite *N*, using Faà di Bruno's formula to generate histograms at each step of the fitting process.

2.3.4 Use of free energies to predict cluster distributions in large-*N* limit.

To obtain the bulk equilibrium distribution of cluster concentrations for all j and k given the free OS^- and free Na^+ concentrations requires a straightforward application of the law of mass action:

$$c_{j,k} = K_{j,k} {c_{1,0}}^{j} {c_{0,1}}^{k} \tag{24}$$

Finding pairs of free OS^- and Na^+ concentrations $c_{1,0}$ and $c_{0,1}$ that correspond to a charge-neutral system at a given total concentration is less straightforward. The overall composition (for a charge-neutral system without added salt) is not a single-valued function of the concentration of free OS, but it is a single-valued function of the concentration of free Na^+. At a given free Na^+ concentration, the total OS and Na^+ concentrations are monotonically increasing functions of free OS, so the concentration of free OS that produces a charge-neutral solution can be obtained iteratively to any desired precision using a bisection algorithm.

2.4 Results and discussion

2.4.1 Qualitative observations of simulations

Visualizations of simulation trajectories using VMD[58] confirmed that the surfactant molecules form clusters rapidly and appear to equilibrate over the initial 5 ns after starting from an evenly-spaced array of molecule. In some cases, the aggregation state of the system fluctuated between a single large micelle and two smaller micelles, as shown in Figure 2.1 Counterions could be observed both in close contact with the aggregate surfaces and freely moving in the simulation box, and varying numbers of free OS were also present.

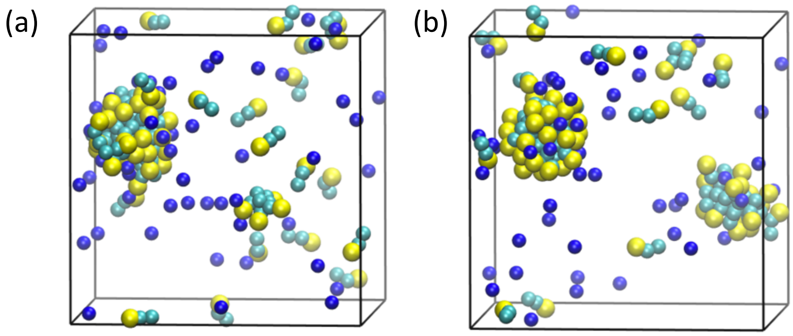


Figure 2.1 Snapshots (a) and (b) are taken from the same trajectory for the equilibrium state of 60 SOS molecules in a simulation box with a concentration of 67.8 mM at times (a) 166.4 ns (b) 167.2 ns. Navy blue beads represent Na^+, yellow beads represent the sulfate groups, and light blue beads represent C1 and C2 tail beads.

2.4.2 Fitting free energy surface from cluster histograms

The PEACH global fitting procedure produces a set of cluster association constants $K_{j,k}$ for 639 combinations of number of surfactants j and number of associated counterions k. The fit is quite successful in reproducing the distribution of surfactant and counterion contents of micelles across simulation sizes at two concentrations, as evident from colored contour maps of the fitted and the simulated distributions for several trajectories shown in Figure 2.2 (with the remainder provided in Appendix II, Figures S1 and S2). Distributions of micelle size irrespective of counterion content (obtained by summing $<m_{j,k}>$ over all k) for all trajectories 1-12 and their fits are shown in Figure 2.3. Poorest agreement is found between the fit and the simulation data in the range of 5-15 OS, where the fewest micelles are found.

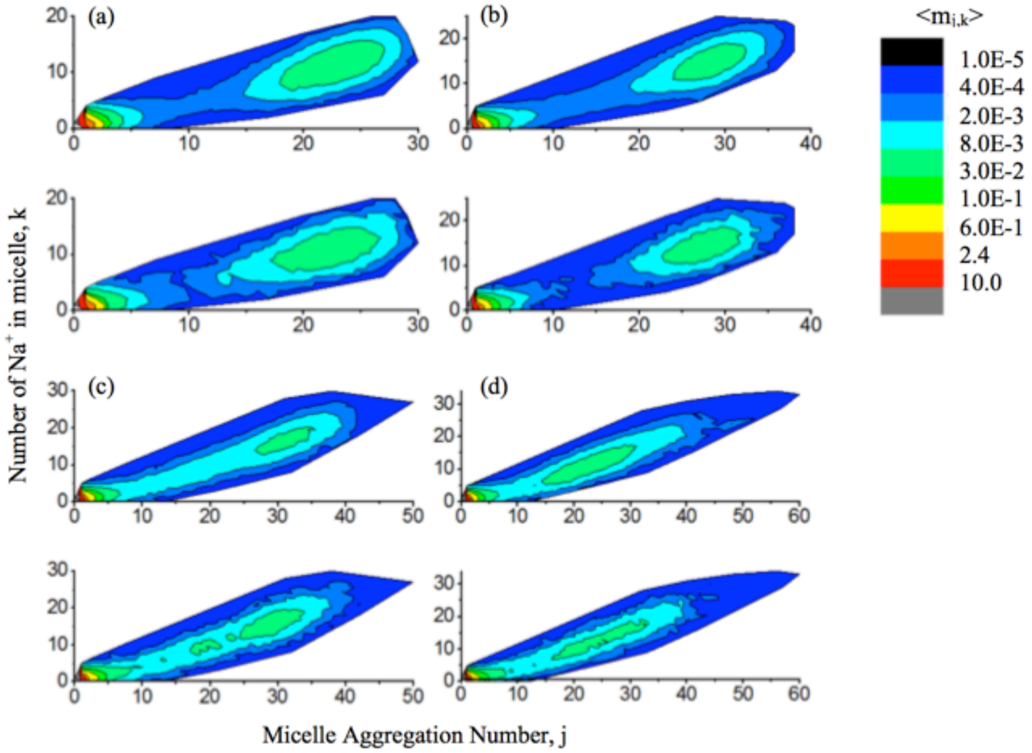


Figure 2.2 Contour plots for the equilibrium micelle size distribution with respect to surfactant number and sodium ion number for the concentration c= 67.8 mM. The four plots are for different numbers of SOS molecules in the simulation box, with (a) SOS =20, (b) SOS =30, (c) SOS=40, (d) SOS=50. The upper panel of each graph is for the fitted distribution while the lower panel is for the simulated distribution.

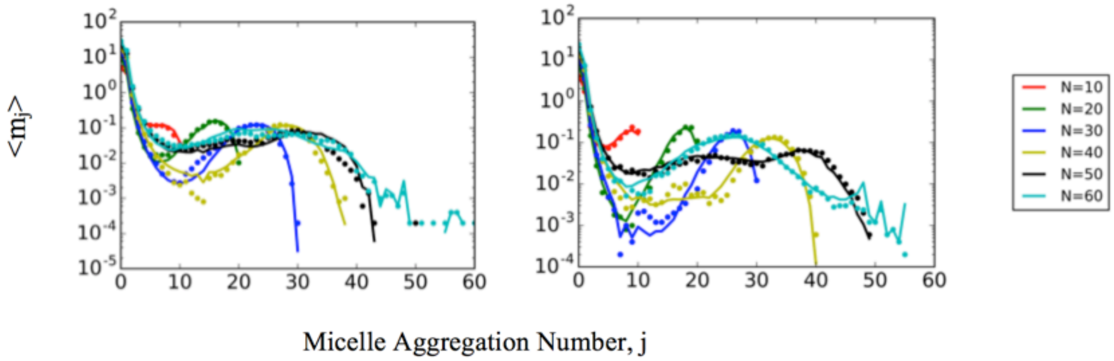


Figure 2.3 Cluster size distributions showing the average number $<m_j>$ of micelles containing a given number *j* of octyl sulfate chains with any number of associated counterions ($\sum_{k=0}^{N} < m_{j,k} >$). Left panel: 67.8 mM; right panel: 114.9 mM. Symbols show simulation data and curves show the results of the global fit. Different colors represent different system sizes N as shown in the legend.

The trends in cluster size distribution with increasing *N* at fixed concentration can be interpreted by considering the discrete number of clusters that can be formed in a small-*N* system, and are in qualitative agreement with a simple model prediction.[6] Focusing on the higher concentration, we see that increasing the number of surfactants up to *N*=40 produces first a shift towards larger micelles (where the distribution is peaked above *j*=30). Above this number we see a bifurcation (as at *N*=50) where the system fluctuations between states with two small (*j* ~ 20) micelles and states with one large (*j*~40) micelle, followed by a reversion to smaller micelles at *N*=60 as the two-micelle state predominates.

2.4.3 Validation of PEACH-derived equilibrium constants: prediction of cluster distributions at large *N*

The primary value of evaluating equilibrium association constants $K_{j,k}$ is to enable predictions at large *N* where the law of mass action will apply. To show that the global fit to small-*N* simulation data gives accurate predictions at large N, we present a comparison of predictions made by applying PEACH-derived $K_{j,k}$ values, using the law of mass action, against actual simulation data for a system of 560 SOS. The large system, which contains an average of approximately 10 micelles and 80 free OS monomers, is not very computationally demanding to equilibrate using the present solvent-free model. With an atomistic, explicit-solvent model, however, such a simulation would be rather costly, as it would require approximately 300,000 waters. Figure 2.4 shows the predicted and observed 2-d distribution of clusters according to number of OS and counterions, while Figure 5 shows the distributions only in terms of number of surfactants. The predictions made using the best-fit surface derived from small-system simulations are generally excellent, with some deviations between the predicted and observed values in regions of poor sampling (evident in Figure 2.5 for j between 5-15 and $j > 40$), and confirm that the PEACH method provides a valid path to extrapolate from small-N system distributions to large-N systems. Whether it is more efficient to perform several small-N simulations or a single large-N simulation to obtain equilibrium distributions will depend on system details and has not yet been rigorously explored. However, if computational cost per unit of simulation time or the simulation time needed to equilibrate and sample cluster size distributions scale worse than linearly with system size, small-N simulations will tend to be more efficient.

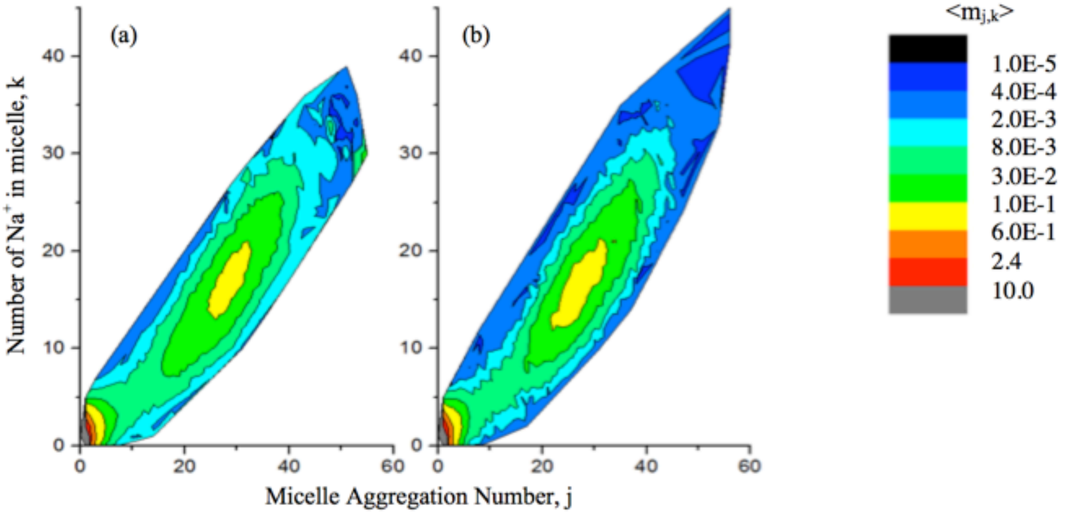


Figure 2.4 Contour plots for the equilibrium micelle size distribution with respect to surfactant number *j* and sodium ion number *k* for the concentration c= 104.25 mM with 560 molecules in the simulation box. (a) Mean <$m_{j,k}$> obtained from simulation, Trial 13 from Table 2.1. (b) Mean <$m_{j,k}$> predicted from association constants obtained from PEACH global fit to distributions observed in Trials 1-12.

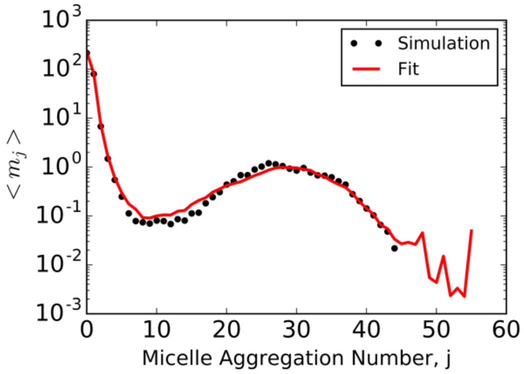


Figure 2.5 Cluster size distributions showing the average number <m_j> of micelles containing a given number j of octyl sulfate chains with any number of associated counterions $\langle m_j \rangle = \sum_{k=0}^{N} \langle m_{j,k} \rangle$) from simulation of 560 SOS system, Trial 13 (symbols) and predicted from association constants fit using PEACH algorithm to Trials 1-12 (curve).

2.4.4 Modeling and applying the PEACH-derived free energy surface

As discussed in section C above, the PEACH-derived equilibrium constants can be used to make predictions of cluster size distributions as a function of total system concentration. To better understand the factors that determine these equilibrium constants, it is convenient to convert them to cluster free energies (CFE, ΔG) for ease of visualization and to make connections with simple models. CFE for clusters of all sizes at a given set of free monomer concentrations ($c_{1,0}$ for OS and $c_{0,1}$ for counterions) are readily obtained from the fitted equilibrium association constants $K_{j,k}$:

$$\frac{\Delta G_{\mathrm{PEACH}}(j,k)}{k_B T} = -\ln K_{j,k} - (j-1)\ln\left(c_{1,0}/c^{\circ}\right) - k\ln\left(c_{0,1}/c^{\circ}\right) \tag{25}$$

Figure 2.6(a) shows a representation of the surface obtained under conditions that correspond to total SOS concentration of 97 mM. It is interesting from a fundamental perspective to compare the free energy surface with a simple phenomenological model free energy to test whether simple principles can rationalize the form of the surface. To construct a model free energy to fit and compare with we start with the suggestion by Maibaum et al.[59] that the free energy profile for micelle-forming systems can be described by the following dependence on number of surfactants:

$$\Delta G_{\mathrm{micelle}}(j) = -j\Delta\mu + gj^{2/3} + hj^{5/3} \tag{26}$$

where the terms containing linear and 2/3 powers of j correspond to bulk and surface tension contributions (analogous to those used in classical nucleation theory). The $j^{5/3}$ term arises from the limitations of packing amphiphiles. Comparison with the free energies obtained using the PEACH analysis suggested that substitution of j^2 for $j^{5/3}$ gives a more successful fit to the free energy in the current system; whether this is true in other systems may be an interesting topic for future study.

Although various models have been applied to the electrostatic energy of micelle-counterion interactions,[15, 37] the simplest model to incorporate the number of bound counterions is to assume binding of counterions to headgroups with binding energy ε, independent of the degree of association of the micelle or the presence of other counterions. Including the appropriate combinatorial factor for the number of ways to distribute k counterions among j surfactant headgroups, and following Girshick and Chiu[60] in constructing the free energy as a difference between the cluster and the free monomer, this gives

$$\Delta G_{\mathrm{model}}(j,k) = -(j-1)\Delta\mu + g\left(j^{\frac{2}{3}}-1\right) + h(j^2-1) - \varepsilon k - k_B T \ln\left(\frac{j!}{(j-k)!k!}\right) \qquad (27)$$

We found approximate values for parameters g,h, ε, and $\Delta\mu$ by first performing a 1-dimensional fit along the cut k=j/2 (where sampling is relatively good) to estimate $g = 10.9\ k_\mathrm{B}T$ and $h = 0.028\ k_\mathrm{B}T$. Subtracting out the non-linear terms from the best-fit surface yielded an approximate plane:

$$\Delta G_{\mathrm{linear}}(j,k) = \Delta G_{\mathrm{PEACH}}(j,k) - g\left(j^{\frac{2}{3}}-1\right) - h(j^2-1) + k_B T \ln\left(\frac{j!}{(j-k)!k!}\right) \qquad (28)$$

which was then subject to a linear regression fit to find $\Delta\mu$=7.92 $k_\mathrm{B}T$ and ε=3.58 $k_\mathrm{B}T$. The resulting surface, which is depicted alongside the surface $\Delta G_{\mathrm{PEACH}}$ obtained directly from simulation data in Figure 2.6(b), does a good job of reproducing the dependence of cluster free energy on number of surfactant and counterions, except for the premicellar range (j<10) and the poorly sampled large cluster region (j>50).

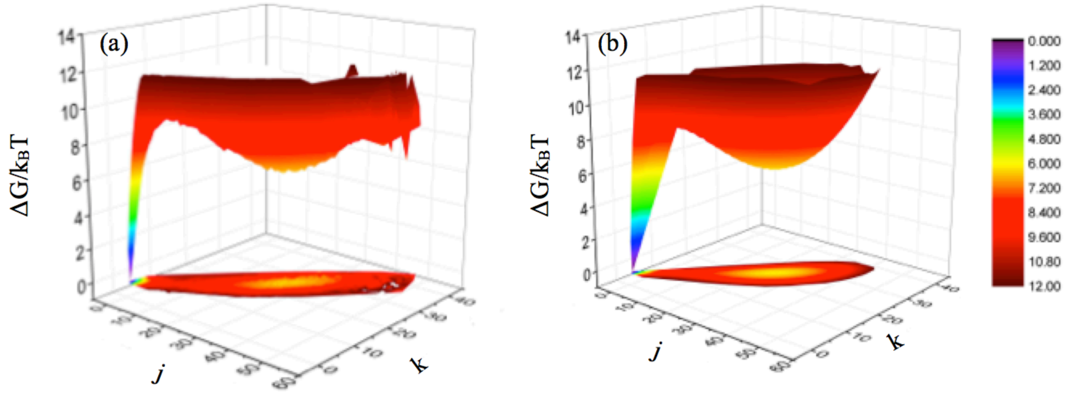


Figure 2.6 Cluster free energy (CFE) surfaces for free surfactant concentration $c_{1,0}$ =6.85 mM and free sodium ion concentration $c_{0,1}$ =51.8 mM. (a) ΔG_{PEACH}, values obtained from fit to trajectories 1-12 and (b) ΔG_{model}.

Having an approximate functional form for $\Delta G(j,k)$ that is grounded in physical principles is also helpful for evaluating predictions obtained from the PEACH fit. Concentrations of clusters with compositions that were not well-sampled (or not sampled at all) may be inaccurately represented (or inaccurately excluded) when this free energy function is used to make predictions. Since the model agrees with the PEACH-derived values where they are most reliable and moves smoothly and in a physically reasonable form to cover cluster compositions that are sampled poorly (or not at all) in the simulations, it will be useful in checking for the influence of sampling-related artifacts on predictions obtained with the PEACH-derived ΔG.

In the following, predictions of concentration-dependent equilibrium properties of the system will be made using three free energy functions: the original ΔG_{PEACH} shown in Figure 2.6(a), obtained directly from simulation data; ΔG_{model} from eq. 27 using parameters g, h, $\Delta\mu$, and ε obtained by fitting to ΔG_{sim}, and an "extended" version ΔG_{ext}, of the PEACH model, which uses values from ΔG_{PEACH} where available but fills in values from ΔG_{model} for any cluster not represented in ΔG_{PEACH}.

(Both ΔG_{model} and ΔG_{ext} were truncated above j=k=75.) The form of ΔG_{model} does not allow overcharged clusters, those with k>j, which were unexpectedly observed at low j; whether these clusters are in fact stable is not clear, but an examination of their impact on the overall cluster size distributions (not shown) suggests that they are of negligible importance except at very high SOS concentrations (>2 mol/L, beyond the concentration range shown here).

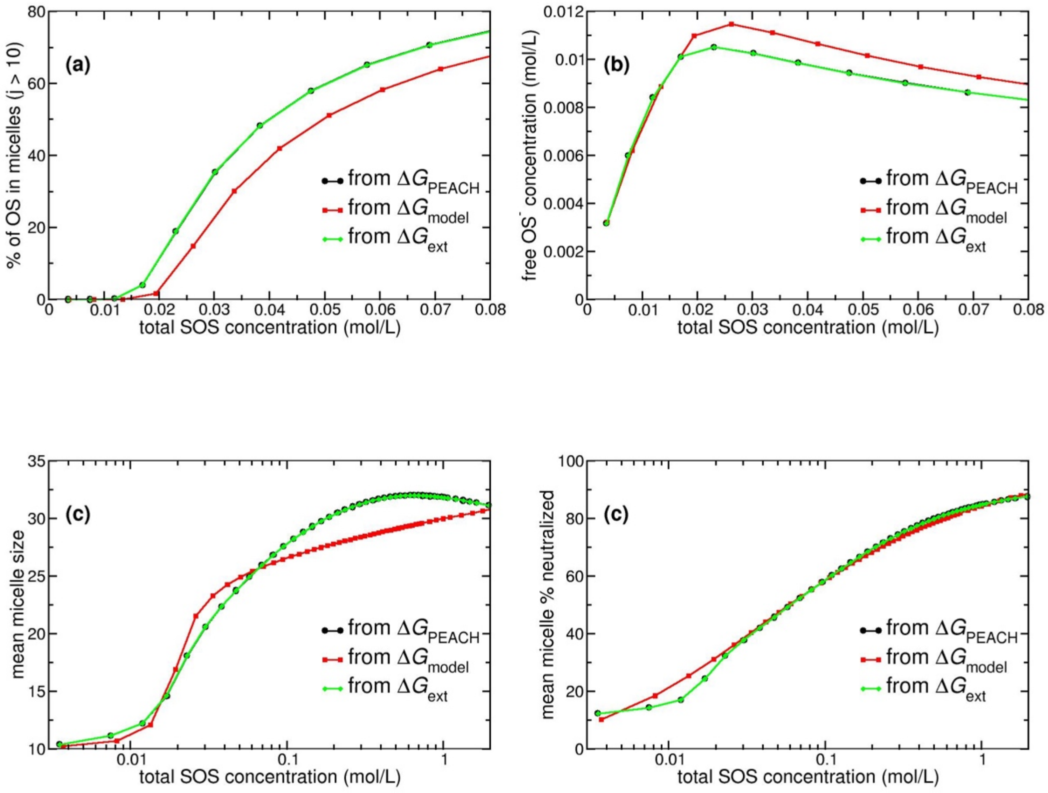


Figure 2.7 Average quantities obtained from cluster size distributions using free energy derived from PEACH analysis (ΔG_{PEACH}), from a fit to ΔG_{PEACH} using eq. 27 (ΔG_{model}), and a combination of the two (ΔG_{ext}). a) Percentage of OS associated with micelles with 10 or more OS molecules (j>9). b) Free OS concentration (only counting free monomers not associated with counterions). c) Mean micelle size, defined as average number of OS monomers that belong to clusters of size j>9 divided by the average number of those clusters. d) Mean percentage neutralization of micelles, defined as the percent ratio of the

number of counterions associated with clusters of size j>9 to the number of OS associated with clusters in that size range.

Measures of the onset of micellization (the fraction of surfactants contained in micelles shown in Figure 7(a) and the peak in free monomer concentration shown in Figure 7(b)) are approximately consistent among the different models. Using one definition for the CMC, the total concentration at which 50% of surfactants will be found in micelles, the concentration prediction using ΔG_{model} exceeds the prediction obtained from ΔG_{PEACH} by about 12%. The apparent CMC for the simulation model, near 40 mM, is well below the value of 110-130 mM at 298K reported from experiments.[61, 62] A significant increase in the micelle fraction neutralized (mean number of counterions bound per OS in micelles) was predicted with increasing system concentration within all models, from about 50% near the CMC to over 70% at 200 mM.

It is probably not a coincidence that the models agree best for total SOS concentrations near 100 mM, where the small-N simulations were conducted. The greatest qualitative discrepancy among the models is seen in the variation of mean micelle size with concentration, where a peak is predicted when the PEACH-derived ΔG_{PEACH} values are used alone in contrast to a steady climb in average micelle size when the model free energy is used. For neutral or zwitterionic surfactant micelles (or any other single-component system) the law of mass action would always yield an increase in average micelle size with increasing concentration. The present non-monotonic behavior is related to the two-component nature of the self-assembly. As total SOS concentration increases, so does total free Na^+ concentration, because micelles are not fully neutralized; the average charge state of the micelles shifts to favor clusters with more counterions. If the binding of counterions to small micelles (on a per-molecule basis) is more favorable than to large ones, then this shift can produce a shift towards smaller average micelle size. Whether such an effect is real (that is, reflective of the physical behavior of the simulation model), or is an artifact of the sampling and/or the PEACH free energy fitting procedure is difficult to tell. There are no significant differences between ΔG_{PEACH} and ΔG_{ext}, which may indicate that the domain of clusters sampled in the small-N simulations is sufficient to describe the clusters that are populated in the concentration range of the predictions. All we can conclude at this point is that care is needed in interpreting subtle trends inferred from PEACH-derived free energies, especially for predictions about concentration conditions that are very different from the small-N simulations.

2.4.5 Sensitivity to cluster definition

A further reason for caution in applying PEACH-derived free energies at concentration conditions far from the fitted systems is that the model does not use a sophisticated treatment of non-ideal effects, only accounting roughly for excluded volume. In previous work,[32] we have shown that an optimal choice of the distance cutoff used in cluster definitions can minimize the influence of non-ideality on predicted cluster size distributions. Here again we have investigated the dependence of the quality of the global fit to evaluate the quality of distance cutoffs used in cluster definitions, but do not arrive at such a clear-cut result. The fit quality is defined using a convergence criterion C_{tot}, which is a sum over all cluster of the average (taken over the individual trajectories, weighted by the estimated statistical certainty) of the mean-squared deviation between fitted and simulated cluster levels:.

$$C_{tot} = N_{\text{cluster types}}^{-1} * \sum\nolimits_{j,k} \exp\left(\sum_{i=1}^{J_i} w_{i,j,k} \log\left(\frac{\langle m_{j,k}\rangle_{\text{fit},i} - \langle m_{j,k}\rangle_{\text{sim,i}}}{\langle m_{j,k}\rangle_{\text{sim,i}}}\right)^2\right) \tag{29}$$

Table 2.2 The convergence criterion C_{tot} (over a range of inter-atomic cutoff distances r_c used to define member of a cluster) for different cutoffs from the global fitting procedure.

Trial	Cutoff 1: the distance between tail beads (nm)	Cutoff 2: the distance between headgroup and counterion (nm)	Convergence Criterion C_{tot}
1	0.85	1.5	0.1719
2	0.45	1.5	0.1166
3	0.65	1.7	1.3695
4	0.65	1.3	0.1480
5	0.65	1.5	0.1647

Table 2.2 shows the convergence criteria for the fits when we apply different cutoff distances in the post-analysis for the simulations. The "optimal" cluster definition should yield the lowest convergence criterion in the global fitting procedure. In contrast to the sensitive dependence of fit quality on cutoff distances demonstrated for MTBE clusters forming from the vapor phase,[32] fairly large variations in cluster definition here do little to change the weighted mean-square error away from the value obtained at the original cutoff combination of 0.65 nm and 1.5 nm. In part this is reflective of the increased amount of noise in the statistics in the present case; increased statistical noise will mask effects of fundamental thermodynamic relevance in the cluster definitions. The greater statistical noise can be attributed in turn to the two-dimensional array of cluster compositions, for which there are more individual cluster identities to sample, with a greater proportion on the edges of the two-dimensional distribution in any given simulation. Whatever the cause, it makes it difficult to address or estimate non-ideal effects in the present case.

Looking at the actual distributions (some of which are shown in the Appendix II, Figure S3), we can also see that some changes to cutoffs have nearly no effect on the actual cluster size distribution: for instance, extending the OS-OS contact cutoff from 0.65 nm to 0.85 nm makes almost no difference to the fits. This is probably because only tail-tail contacts are used to define OS-OS neighbors in the cluster definition, and in the typical micelle structure the tail beads are separated from other monomers or micelles by a layer of headgroups and counterions. Expanding the range of possible neighbors is not likely to include additional surfactant tailgroups. By contrast, reducing the cutoff distance from 0.65 nm to 0.45 nm restricts the OS-OS contact definition to an unphysical extent, so that no large clusters are counted in the cluster size distributions (see Figure S3(b)), without significantly changing the convergence criterion. This example is a reason to use caution in looking at the fit quality alone in assigning cluster definitions. Changing the cutoff for headgroup-counterion interactions has qualitatively predictable effects, in that increasing the cutoff will shift the cluster distribution towards greater numbers of counterions per cluster.

We have applied sets of cluster free energies $\Delta G_{\mathrm{PEACH}}$ obtained from fitting distributions obtained from trajectories 1-12 using different cutoffs (omitting the case that did not recognize large micelles) to evaluate average system properties versus total SOS concentration, with results shown in Figure 2.8. We find that changing the OS-OS cutoff makes very little difference to the system properties, as expected given its small effect on the cluster size distributions. Changing the OS-Na cutoff yields quantitative changes to the CMC, mean micelle size, and micelle charge. The prominence of the peak in the concentration dependence of average micelle size is influenced by the cutoff, which gives more reason to suspect that it is artifactual. Variations in cutoff do not alter the predicted trend that the mean micelle size will increase from ~25 to ~30 as the total concentration is increased over an order of magnitude above the CMC. Although the specific value of the degree of charge neutralization is sensitive to the cutoff, the trend of a significant increase in degree of neutralization with increasing total SOS concentration is also consistent across the different cutoffs; the change in degree of neutralization with concentration appears to be greater than its sensitivity to cutoff definition.

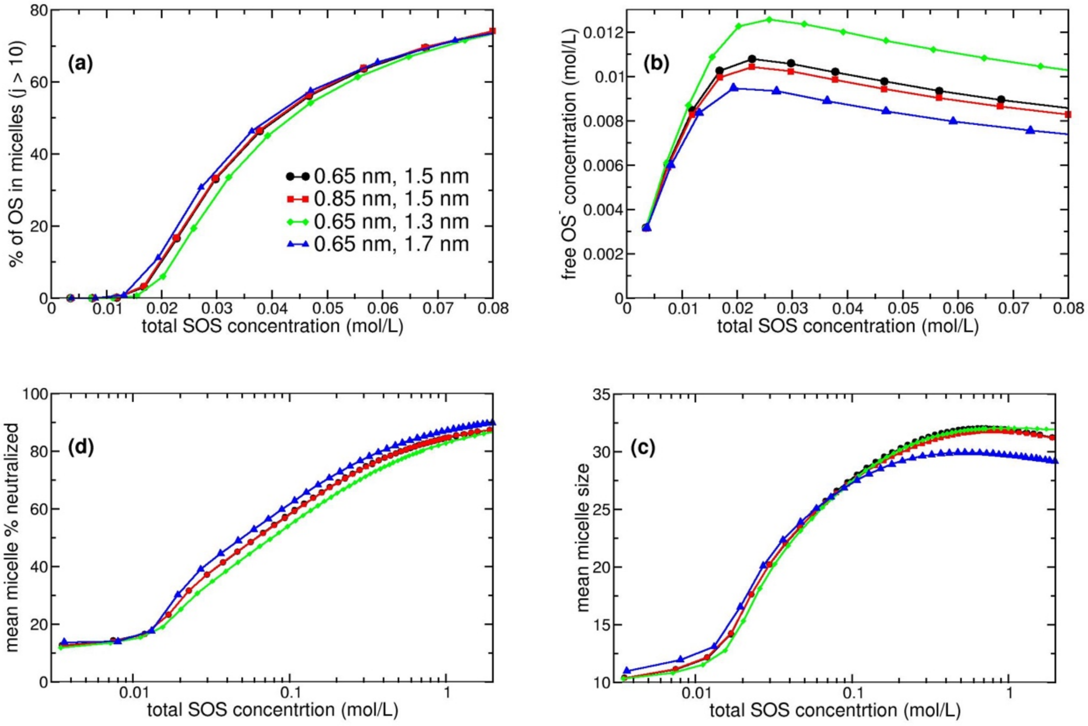


Figure 2.8 Average quantities obtained from cluster size distributions using free energies derived from PEACH analysis at different cutoffs. Legend for all panels is given in Panel (a); first value given is the cutoff for OS tail groups, second is the headgroup-sodium cutoff. Quantities plotted are as defined in Figure 2.7.

2.5 Conclusions

A very efficient approach to calculating canonical averages of cluster size distributions at finite *N* has been derived from the Faà di Bruno's formula, and has been implemented within a global fitting process to obtain cluster free energies from cluster distributions obtained in several simulations. The PEACH (Partition-Enabled Analysis of Cluster Histograms) approach has been applied to finding cluster free energies for a solvent-free coarse-grained model of sodium octyl sulfate (SOS), within the "dressed micelle" framework that treats counterions as part of a micelle's chemical identity. A single set of cluster free energies, optimized to reproduce results from twelve simulations with *N* up to 60 SOS that featured an average of one or two micelles, was able to predict micelle size and charge distributions for a system of 560 SOS. The free energy function (consisting of discrete values for 639 clusters containing different numbers of OS and counterions), and a 4-parameter model fit were used to make predictions for average properties of SOS solutions over a range of concentrations, showing modest growth in average cluster size and significant increase in the ratio of bound counterions to surfactants in micelles as concentration increases. The sensitivities of the fitting procedure and predicted concentration-dependent system behavior to the cutoff distances used in cluster definition were explored. The methods described should be applicable to the detailed analysis of the thermodynamics of self-assembly of both neutral and charged surfactants using atomistic models, so long as converged cluster size distributions can be obtained from simulations of one or two micelles, and to the statistics of reversible aggregation observed in small-*N* simulations of other self-assembled systems.

Chapter 3. Derivation of micelle size-dependent free energies of aggregation for octyl phosphocholine from molecular dynamics simulation

3.1 Introduction

Surfactant micelles, disordered aggregates of amphiphiles in aqueous solution, are typically formed above a critical micelle concentration (cmc) as the hydrophobic attraction of the surfactant tails overcomes the entropy of mixing individual surfactant monomers (unimers) and small clusters (premicelles) throughout the solution. The structural disorder, polydispersity, and responsiveness to conditions present challenges to the unambiguous determination of the degree of aggregation and structure of surfactant micelles through experiment alone. Molecular simulation has long been a source of insight into these features.[63-65] Alkyl phosphocholines have been a focus of particular attention[66-68] in part because they are zwitterionic (and so bypass complications associated with counterions) and in part because they share their headgroup structures with the well-studied bilayer-forming phosphatidylcholine glycero-phospholipids.

A number of ingredients are required to achieve a level of confidence in the structural and thermodynamic properties displayed through simulation models, which by necessity rely on certain approximations. The use of a sufficiently realistic force field (or at least one whose limitations are well understood and characterized) is one such ingredient. The ability to extrapolate from simulation data obtained on a finite size system to the limit of a macroscopic number of molecules and clusters is another, and the ability to make connections between simulated structures and experimental observables is a third. This report will describe a simulation study on micellization of a simple surfactant, octyl phosphocholine (OPC), in which steps are taken to address all three ingredients. The force fields used for surfactant tails are based on the TraPPE-UA model,[69] and so have been parameterized to fit alkane-alkane thermodynamic properties, but with modifications to interactions with water made to yield accurate free energies of hydration.[70] The PEACH statistical analysis method[32, 71] allows the size-dependent cluster free energy to be determined over a range of micelle sizes from a series of simulations with relatively few surfactants; the resulting free energy function allows the cmc to be calculated in a way that facilitates comparison with experiment. Finally, the effect of incorporating polydispersity into predictions of the x-ray scattering of OPC (where polydispersity is relatively high) will be investigated.

3.2 Methods

3.2.1 General elements of force fields.

Two united-atom (UA) force fields, in which similar models for octyl phosphocholine were matched with different water potentials, were used and compared. In both, bonded interactions of the alkyl tails, as well as all pairwise interactions of CH_x groups with other CH_x groups (including headgroup methyl and methylenes) used parameters from the TraPPE-UA force field.[72] Parameters for the OPC headgroup (bond lengths, bending and torsional potentials, partial charges, and Lennard-Jones parameters) were taken from the Gromos lipid forcefield of Poger *et al.*, [73] derived from the Gromos G53A6 force field[74] but including an adjustment to the interaction between the choline methyl group with non-ester phosphate oxygens that was made to bring lipid bilayer properties in line with experiment.

For simulations using TIP4P-2005 water, Lennard-Jones parameters for interactions between water oxygen and all CH_x groups were taken from the HHAlkane model;[70] this force field will be referred to as "HHA". For simulations using SPC water (denoted "Alk-SPC") these parameters were calculated as described below.

3.2.2 Alkane-SPC water interactions

The TraPPE-UA model was reparameterized to reproduce experimental hydration free energies for n-alkanes in SPC water, by adjusting the nonbonded alkane site-water interactions (CH_x-OW) following the concept of Ashbaugh et al.[70] The excess hydration free energy is defined as the free energy to transfer the solute molecule from an ideal gas to an ideal-dilute aqueous solution at the same concentration. This quantity was calculated from the g_bar utility in Gromacs 5.0[75] which is based on the BAR (Bennett Acceptance Ratio) method[76]. In the implementation of BAR method, the system of alkane with water was sampled for multiple coupling states with a parameter λ to adjust the strength of the Van der Waals interactions between the alkane and the solvent. For our calculations, 20 points of λ were chosen from 0 to 1 with an increment of 0.05. Free energy differences for each two adjacent points i and i+1 are calculated based on $<U_{i+1}-U_i>_i$ and $<U_i-U_{i+1}>_{i+1}$.

The simulation for each λ point was set up with one alkane in a box of ~420 SPC water using Gromacs 5.0[75]. The pressure is maintained at 1 bar and a compressibility of 4.5×10^{-5} by Parrinello-Rahman barostat[77] with $\tau_P = 2\ ps$. The stochastic dynamics integrator with a 2 fs time step and coupling time $\tau_T = 2\ ps$ was used for integration of equations of motion and to maintain a constant temperature of 300 K. The Verlet[78] cutoff-scheme was applied for short-range non-bonded interactions with a cutoff of 9 Å. The long-range dispersion correction was applied for energy and pressure. Particle mesh Ewald summation[55] was used to account for long-range electrostatics with a real space cutoff of 9 Å. After a 100 ps pre-equilibration, a 1 ns simulation was used for each λ-point in the hydration free energy calculation.

To optimize the nonbonded CH_x-OW interactions for accurate interaction free energies of alkyl tail in SPC water, the LJ well depth $\varepsilon_{CHx\text{-}OW}$ and LJ diameter $\sigma_{CHx\text{-}OW}$ are adjusted to fit the hydration free energies of n-alkanes. The initial cross interactions were determined by Lorentz-Berthelot combining rules

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_i + \sigma_j\right), \epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{1}$$

The strength of the alkane-water attraction was varied by scaling $\varepsilon_{CHx\text{-}OW}$ while adjusting $\sigma_{CHx\text{-}OW}$ to maintain constant thermal radius[79] r_{iw}^{therm}. Given as below, r_{iw}^{therm} is fixed during the fitting to maintain the solute excluded volume.

$$r_{iw}^{therm} = \sigma_{iw}\left[\frac{2}{1+(kT/\varepsilon_{iw})^{1/2}}\right]^{1/6} \tag{2}$$

3.2.3 Simulations of OPC and PEACH analysis

Simulations of OPC in SPC and TIP4P-2005 water were performed with the Gromacs 5.0[75] software package. For PEACH calculations, we set up a series of simulations of different concentrations (20-55 OPCs with an increment of 5 OPC's, each paired with two sizes of solvent box). This set is chosen to obtain sufficient samplings across the full range of cluster sizes and allowing K_i to be obtained for cluster sizes $2 < i < 40\text{-}50$. The choices of N and V used in PEACH analysis are given in Table 1.

The temperature was maintained at 300K by velocity rescaling thermostat[56] with $\tau_T = 2\ ps.$ The pressure is maintained at 1 bar and a compressibility of 4.5×10^{-5} bar^{-1} by Berendsen barostat[80] with $\tau_P = 2\ ps$. The Gromacs default (leap-frog) integrator with a 2 fs time step was used for integration of equations of motion. The Verlet[78] cutoff-scheme was applied for short-range non-bonded interactions with a cutoff of 1.4 nm. The long-range dispersion correction was applied for energy and pressure. Particle-mesh Ewald summation[55] was used to account for Coulomb interactions with a real space cutoff of 1.4 nm.

Microsecond simulations (1.0 μs for HHA, 1.2 μs for Alk-SPC) with a 20 ns pre-equilibration were performed for each trial and used for PEACH analysis. (An additional large system simulation was performed with 250 OPC solvated by 30330 SPC waters and run for 200 ns.) The trajectory was analyzed by a modified version of Gromacs utility g_clustsize to determine the distribution of cluster sizes. Two OPC molecules are considered to be neighbors if the distance between any of their alkyl tail sites is lower than a cut-off distance r_{cut}. Values of r_{cut} at intervals of 0.05 nm between 0.45 and 0.65 nm were used, and results derived from these are compared as described below. The chains related by neighboring interactions are considered to be of the same cluster. Histograms of $<n_i>$, representing the mean numbers of clusters of aggregation number *i* averaged over each trajectory, were used as input for PEACH (Partition-Enabled Analysis of Cluster Histograms) to find a globally optimized set $\{K_i\}$ of equilibrium association constants for each cluster size observed. As described in previous publications,[32, 71] the method involves assigning an association constant to each possible cluster size *i*, generating a model cluster size distribution for each simulation based on its total *N* and *V*, and adjusting it iteratively to find the best set K_i to reproduce the cluster size distributions from simulations. The core algorithm uses the following equation for calculation of cluster size distribution from the current model $\{K_i\}$:[6]

$$< n_i >= Q(N,V)^{-1} \sum_{j=1}^{p(N)} n_{ij} Q(\{n_i\}_j, V) = Q(N,V)^{-1} \sum_{j=1}^{p(N)} \left(n_{ij} \prod_{i=1}^{N} \frac{(K_i V c^o)^{n_{ij}}}{n_{ij}!} \right) \quad (3)$$

in which $< n_i >$ is the ensemble average number of *i*-mer clusters present, p(*N*) represents the number of possible compositions of clusters that add up to *N* (called "partitions of *N*" in number theory), $Q(\{n_i\}_j, V)$ is the canonical partition function for a possible composition $\{n_i\}_j$, and $Q(N,V)$

is the sum of these partition functions over all compositions. This calculation can be simplified[71], avoiding the need to generate all partitions of N, using a generating function approach. Briefly, at fixed V, $Q(N)$ is obtained as the coefficient of λ^N calculated from the polynomial expansion of the grand partition function $\Xi(\lambda)$ (where λ is now the thermodynamic activity, unlike in section *2.1*):

$$\Xi(\lambda) = \exp\left(\sum_{i=1}^{\infty} q_i \lambda^i\right) = \sum_{N=0}^{\infty} \lambda^N Q(N) \tag{4}$$

Evaluation of $Q(N)$ can be achieved by differentiation of $\Xi(\lambda)$:

$$Q(N) = \frac{1}{N!}\left[\frac{\partial^N \exp\left(\sum_{i=1}^{\infty} q_i \lambda^i\right)}{\partial \lambda^N}\right]_{\lambda=0} \tag{5}$$

The single-cluster partition function q_i is related to the association constant K_i, the system volume V, the standard-state concentration $c°$, and the standard unimer partition function $q°_1$ as:

$$q_i = K_i c° V (q°_1)^i \tag{6}$$

The mean number of clusters of a given size can then be calculated as[71]:

$$\langle n_i \rangle = \frac{q_i Q(N-i)}{Q(N)} \tag{7}$$

The Gibbs free energy of formation of an i-mer cluster from i unimers, under conditions where both unimer and i-mer concentrations are given by c, can be expressed as:

$$\frac{\Delta G_i}{kT} = -lnK_i - (i-1) * ln\frac{c}{c^0} \tag{8}$$

Table 3.1 Details of simulations used in PEACH fitting

Simulation details for OPC cluster aggregation in SPC water (Alk-SPC)									
	Set One			Set Two			Set Three		
N_{OPC}	N_{water}	Box Size/nm^3	c_{tot}/mM	N_{water}	Box Size/nm^3	c_{tot}/mM	N_{water}	Box Size/nm^3	c_{tot}/mM
20	3812	123.6	282	5623	179.1	191	7000	221	153
25		125.4	352		180.9	238		223	191
30		127.7	422		182.6	286		225	230
35		129.4	493		184.5	334		227	268
40		131.4	563		186.7	381		229	306
45		133.3	633		188.6	429		231	345
50		135.1	704		190.4	477		233	383
55		136.9	774		192.3	524		234	421

Simulation details for OPC cluster aggregation in TIP4P/2005 water (HHA)						
	Set One			Set Two		
N_{OPC}	N_{water}	Box Size/nm^3	c_{tot}/mM	N_{water}	Box Size/nm^3	c_{tot}/mM
15	3812	120.3	220	5025	156.8	165
20		122.2	293	5053	159.5	219
25		124.1	366	5035	160.8	275

30		126.3	439	5030	162.6	330
35		128.2	512	5047	165.3	384
40		130.4	585	6021	196.5	367
45	4549	154.4	547	6017	198.3	414
50	-	-	-	6024	200.6	459
55	-	-	-	6768	234.7	450

3.2.4 WAXSiS analysis for polydisperse micelles

To model the small-angle X-ray scattering (SAXS) profile for micelles of a specific aggregation number, structures of single OPC micelles were submitted to the WAXSiS server.[2] The server re-solvates the structure and calculates averages over a short MD trajectory to average over small fluctuations of lipid and solvent configurations. Larger shape fluctuations were observed over the trajectory, thus scattering profiles are averaged across 5~10 configurations of a single micelle observed throughout a trajectory to represent the profile for that size.

PEACH-derived association constants or free energies can be used to calculate the full distribution of concentrations for micelles of each possible aggregation number. Neglecting inter-micelle interactions, the expected form factor of the mixture should mirror the number-weighted average of the individual form factors. To avoid having to generate scattering profiles for 40 distinct micelle sizes, we instead chose representative sizes at intervals of 5. The r_{cut} =0.45 nm cluster criterion was used, based on the reasoning that more compact clusters would contribute more strongly to scattering. Because the scattering intensity of a spherical object scales approximately with the square of its volume[81], clusters of size similar to these representative clusters are assumed to scatter as the nearest representative size micelle, scaled by the square of the ratio of their aggregation numbers;

$$I = I_{20}\sum_{i=14}^{22}(\frac{i}{20})^2 * c_i + I_{25}\sum_{i=23}^{27}(\frac{i}{25})^2 * c_i + I_{30}\sum_{i=28}^{32}(\frac{i}{30})^2 * c_i + I_{35}\sum_{i=33}^{37}(\frac{i}{35})^2 * c_i + I_{40}\sum_{i=38}^{42}(\frac{i}{40})^2 * c_i + I_{45}\sum_{i=43}^{47}(\frac{i}{45})^2 * c_i + I_{50}\sum_{i=48}^{52}(\frac{i}{50})^2 * c_i \quad (9)$$

3.3 Results and Discussion

3.3.1 Force field parameter selection

A key element of a force field for accurate simulations of micelle formation is to reproduce the interaction free energies of alkyl surfactant tails with each other and with water. As shown in previous simulation studies of micellization[36], a number of force fields yield cmc estimates below experiment by a factor of 2 or worse. One of these, the TraPPE-UA model[69] was designed to provide an accurate description of alkane-alkane interactions, reproducing the thermodynamics of alkane liquid-vapor equilibria. The likely weakness therefore lies in hydration free energies that are consistently too high – presumably due to unaccounted polarization effects between water and alkanes. To address this, Ashbaugh *et al.*[70] used TraPPE as a basis for their HH-Alkane model, where the CH_x-OW Lennard-Jones interactions were adjusted slightly to improve agreement of alkane hydration free energies upon solvation with TIP4P/2005 water.[82] Here we use the HH-Alkane parameters with TIP4P/2005 in one of our two force fields for OPC. For future use in studies of the important area of micelle-protein interactions,[17] we sought to develop a similar set of CH_x-OW parameters for alkane interactions with the SPC water model to aid in compatibility with the Gromos G54a7 UA protein force field,[83] which was designed for use with SPC. To do so, we again started with the TraPPE parameters and followed an abbreviated version of the steps taken for the HH-Alkane force field to optimize hydration free energies of small alkanes.

The initial parameters for the fitting were determined using Lorentz-Berthelot combining rules based on the LJ parameters for TraPPE alkane sites and SPC water. It is shown in Table 2 that the initial set of parameters used in ethane predicts higher hydration free energies than experiments i.e. stronger repulsions between SPC water and alkanes; thus we increase the LJ well depth $\varepsilon_{CHx\text{-}OW}$ iteratively in seek of a best set of parameters to fit the hydration free energies. The LJ diameter $\sigma_{CHx\text{-}OW}$ was adjusted at the same time to keep the thermal radius fixed. CH_3 group parameters were first adjusted to fit to the experimental hydration free energy of ethane at 300 K. Keeping those parameters fixed, CH_2 parameters were adjusted to give agreement with the experimental hydration free energy of *n*-propane at 300 K. To check the transferability of the parameters, the free enegy of hydration of *n*-butane was then evaluated using the CH_2 and CH_3 parameters. CH-OW parameters were also adjusted according to experimental hydration free energy of isobutane, while the neopentane hydration free energy (which presumably is most sensitive to the CH_3-OW parameters as the quaternary carbon is shielded from contact with water) was close to experiment without further adjustment of the C-OW parameters. The results are given in Table 2. As for the HH-Alkane force field, small adjustments in attractive strength (by 1.6% for CH_3, 3.5% for CH_2) were sufficient to bring simulated hydration free energies in line with experiment.

Table 3.2 LJ Parameters and test on hydration free energy of alkanes

Cross LJ parameters before and after optimization

	Initial parameters		New (Alk-SPC) parameters		
	σ_{ij} /nm	ε_{ij} /K	σ_{ij} /nm	ε_{ij} /K	r_{iw}^{therm} /nm

<table>
<tr><td>CH3-OW</td><td>0.3458</td><td>87.5451</td><td>0.3455</td><td>88.9805</td><td>0.32596</td></tr>
<tr><td>CH2-OW</td><td>0.3558</td><td>59.9790</td><td>0.3551</td><td>62.0652</td><td>0.32837</td></tr>
<tr><td>CH-OW</td><td>0.38985</td><td>30.8780</td><td>0.38981</td><td>30.9283</td><td>0.34565</td></tr>
<tr><td>C-OW</td><td>0.46175</td><td>10.2410</td><td>-</td><td>-</td><td>0.38026</td></tr>
<tr><td colspan="6">Test on hydration free energy for new parameters</td></tr>
<tr><td rowspan="2"></td><td colspan="5">Hydration free energy (kJ/mol)</td></tr>
<tr><td colspan="2">Initial Parameter</td><td colspan="2">New (Alk-SPC) Parameter</td><td>Experiment</td></tr>
<tr><td>Ethane</td><td colspan="2">8.30(0.12)</td><td colspan="2">7.75(0.17)</td><td>7.82</td></tr>
<tr><td>Propane</td><td colspan="2">8.61(0.26)</td><td colspan="2">8.43(0.26)</td><td>8.37</td></tr>
<tr><td>*n*-Butane</td><td colspan="4">8.80(0.19)</td><td>8.91</td></tr>
<tr><td>Isobutane</td><td colspan="2">9.24(0.09)</td><td colspan="2">9.47(0.24)</td><td>9.58</td></tr>
<tr><td>Neopentane</td><td colspan="2">10.69(0.33)</td><td colspan="2">-</td><td>10.71</td></tr>
</table>

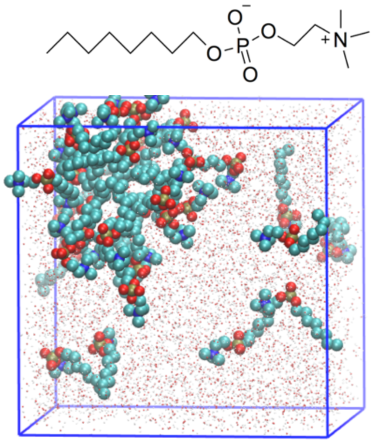


Figure 3.1 Structure of OPC and snapshot of 35 OPC in 3812 SPC water, including a 28-mer micelle. Bead colors represent corresponding elements; Cyan-carbon, red-oxygen, blue-nitrogen, yellow-phosphorus.

3.3.2. Simulation and PEACH analysis of OPC

Simulations over a range of system sizes from 20 to 55 OPC (the structure shown in Fig.3.1) were performed at two concentrations for both force fields. A snapshot of a typical system is shown in Fig.1. For systems with at least 30 OPC, one large micelle is present in the simulation box for most of the trajectory in equilibrium with several monomers or small clusters. The rare event of one large micelle splitting into two occurs a few times in the simulation timescale. Micelle size distributions evaluated with a cluster cut-off criterion $r_{\mathrm{cut}} = 0.45$ nm are shown in Fig. 3.2. For each force field, free energies of aggregation for all cluster sizes j were determined iteratively to optimize the global fit to the full set of size distributions using the PEACH algorithm. Agreement between the model fits (curves, Fig. 3.2) and simulation data (symbols, Fig. 3.2) is reasonably good. The fitting is poorer in the range from cluster size 10 to 20 due to less sampling. Distributions and fits derived using other cut-offs are given in Appendix III Fig. S1.

The HHA data sets appear noisier than the ALK-SPC data primarily because of poorer statistical sampling associated with markedly slower fluctuations in micelle size, as shown in Appendix III Fig. S2. Although a quantitative consideration of the OPC exchange kinetics is outside of the scope of this paper, slower size fluctuations under the same conditions reflect a slower rate of monomer exchange between the micelle and the solvent. One possible contribution is that the viscosity of SPC water is known to be lower than that of TIP4P-2005;[84] this is consistent with our observation that the diffusion constant of dilute OPC using Alk-SPC interactions in SPC water is about 60% higher than using HHA in TIP4P-2005 water at 300 K (results not shown). This difference in diffusion rate is not, however, sufficient to account for the large apparent difference in monomer exchange rates. Given the fairly slight difference in aggregation free energies, it appears that the two solvent models produce significant differences in the activation barrier for OPC exchange.

Several 3-parameter models of the form:

$$\frac{\Delta G(i)}{kT} = Ai^{a} - Bi^{b} + Ci^{c} \tag{10}$$

have been proposed to describe the dependence of micelle free energy on aggregation number.[59, 85, 86] In the droplet model of Shchekin and co-workers[85] and the model of Maibaum and Chandler[59], both derived from classical nucleation theory, the first term arises from the surface free energy and so scales as the surface area of a compact object with a=2/3. The second term is related to the free energy of transfer of the surfactant from solvent to a bulk-like hydrophobic environment and is therefore linear, b=1 and B=$\Delta\mu/kT$. For the the last term, which accounts for the structural limitations to the size of the micelle, considerations of dipole-dipole interactions[85] or entropic costs associated with arranging headgroups at the micelle surface[59] have motivated choices of c = 4/3 and c = 5/3 respectively. A third “quasi-droplet” model also proposed by Shchekin et al.[86] is derived from a model for the micelle that includes partially hydrated segments of chains protruding from a compact hydrophobic core; the evolving balance between core and hydrated segments leads to combinations of terms with a=1, b=3/2, and c=2. In testing these models for their ability to fit the PEACH-derived free energies from simulation, we modify them slightly to ensure that $\Delta G(1)=0$ as follows:

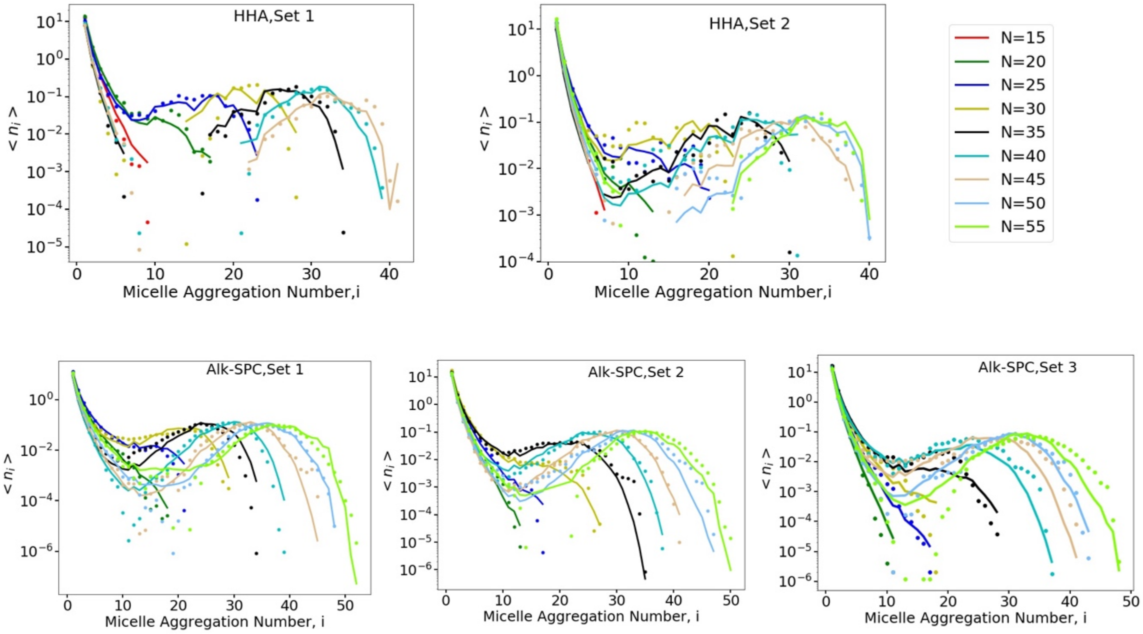


Figure 3.2 Cluster size distribution for simulations (symbols) and the PEACH fit (curves) for sets of MD simulations of OPC, with force fields and sets as labelled (set corresponding to Table 1) and number N of OPC given according to the color legend.

$$\frac{\Delta G(i)}{kT} = A(i^a - 1) - B \times (i^b - 1) + C \times (i^c - 1) \tag{11}$$

Rationalizations for this modification, in the context of nucleation free energies, have been offered in the literature.[32, 60] Attempts to fit the free energy curves using eq. 11 with both the Maibaum/Chandler model and the quasi-droplet model are shown for Alk-SPC and HHA free energies (derived using a r_{cut} = 0.45 nm) in Fig. 3.3. (Agreement with the droplet model, fit not shown, was worse than with the Maibaum/Chandler model.) The quasi-droplet model fit the free energies of association derived from Alk-SPC simulations better than the Maibaum/Chandler model; as shown in Appendix III Fig. S3, this held true for all tested values of r_{cut}. Both models performed equally well for the less complete and noisier sets of HHA-derived data. We note that a similarly good fit could also be obtained using the combination a=2/3, b=1, c=4, but we do not know of any physical justification for the choice of c=4. It will be of interest to compare results from future simulations of other micelle-forming systems to tell whether the quasi-droplet model is generally the best descriptor for free energy of aggregation of globular micelles.

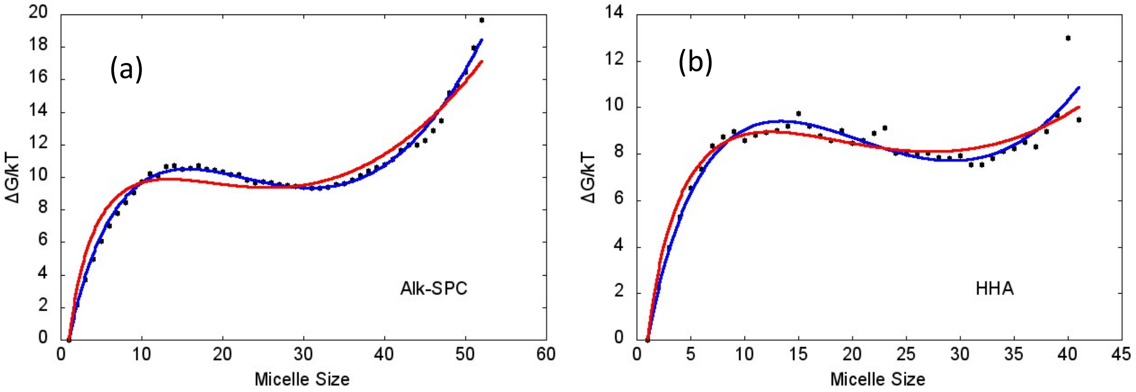


Figure 3.3 PEACH-derived cluster free energy curves (symbols) for OPC, simulated using (a)Alk-SPC and (b) HHA force fields, calculated for a total OPC concentration of 116 mM respectively. Free energy curves correspond to r_{cut} = 0.45 nm. Curves correspond to best fits by eq. 11 using quasi-droplet model (blue) and Maibaum-Chandler model (red).

Table 3.3 Best-fit parameters for the quasi-droplet model for r_{cut}=0.45 nm. (Parameters for other cutoffs given in Supplemental Information Table S1.)

	Alk-SPC	HHA
A	3.864	3.821
B	1.122	1.133

C	0.08781	0.09102

Using the free energies of aggregation $\{\Delta G_i\}$ (or the equilibrium constants $\{K_i\}$ to which they are related by eqn. 8), the distribution of cluster sizes can be easily determined for any bulk concentration, which can then be used to derive information about cmc and average micelle size. The onset of micellization can be qualitatively gauged by considering the fraction of monomers that belong to clusters greater than a given size. Here we choose i=11, near the maximum in the cluster free energy curves, as the minimum cluster size to be defined as a micelle. Figure 3.4 shows this micellar fraction calculated from equilibrium constants derived using different cluster neighbor site distance cut-offs r_{cut}. Predictions are almost identical over the range of r_{cut} from 0.45 to 0.55 nm. Since only contacts between the alkyl tail sites are counted towards neighbour definitions, the insensitivity of the predictions to r_{cut} over this range suggests that the clusters that are captured by this definition have a well-defined micellar organization with with each surfactant's tail and forming redundant site-site contacts with other tails, so that distance fluctuations affecting one pair of sites does not affect its membership in the cluster. For Alk-SPC, the onset of micellization shifts slightly between r_{cut} = 0.55 nm and 0.60 nm and then strongly between r_{cut} = 0.60 nm and r_{cut} = 0.65 nm. Observation of the structures of clusters defined using this largest cut-off shows that under this criterion, two OPC molecules may be counted as neighbors even when their tails are separated by solvent (see Fig. S4 in Appendix III), which suggested to us that this cut-off length is too permissive for description of micelles as they are conventionally understood.

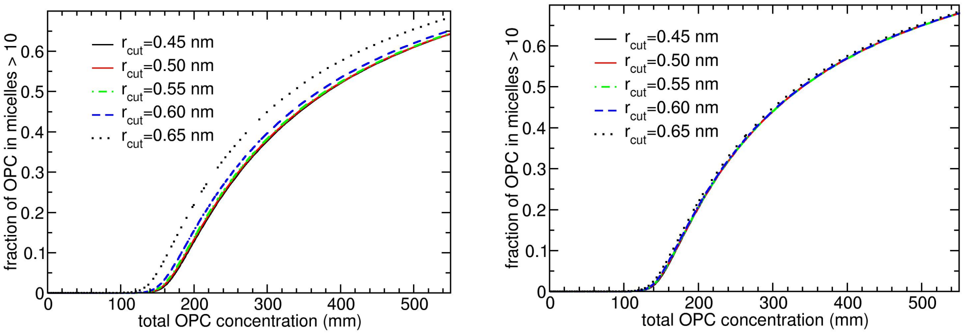


Figure 3.4 Fraction of OPC in clusters of size > 10 vs. total OPC millimolal concentration, as calculated from free energies derived from MD simulations performed with a) Alk-SPC force field and b) HHA force field. Different curves represent predictions based on free energies derived using different tail site cut-off distances r_{cut} in the neighbor criterion for defining clusters.

As has been noted[5], in comparing with experimental data for the cmc, it is important to consider what experiment was performed to determine cmc. A cmc value of 114 mM is reported by Anatrace, which markets OPC under the name FOS-choline-8. The experimental technique used in that determination was to measure the degree of capillary rise vs. surfactant concentration and find the intersection of straight lines fit to the low-concentration (below cmc) and high-concentration (above cmc) regimes.[87] Although we do not obtain interfacial tensions directly from simulations, we may presume from thermodynamic principles that surface tension is related linearly to surfactant activity, which in turn is approximately given by the concentration of unimers. The predicted dependences of unimer concentration on total OPC concentration, calculated from PEACH-derived $\{K_i\}$ using a cluster cut-off definition of r_{cut}=0.45 nm, are shown in Fig. 3.5a-b; data generated from $\{K_i\}$ derived using other cut-off values are presented in Appendix III Fig. S5. Bilinear fits give cmc estimates of 159 mM (Alk-SPC) and 143 mM (HHA), 40% and 25% higher than experiment. (The point of intersection depends on the concentration ranges used in the fits.

Here we used the concentration ranges from 0.5 to 0.75 times the cmc and from 1.5 to 2.0 times the cmc, iterating as needed to reach a self-consistent intersection point.) The dependence of apparent cmc on r_{cut} is shown in Table 3.2.

The crossover concentrations correspond rather well to the onset of formation of micelles shown in Fig. 3.4. Average micelle size from the predicted cluster size distribution increases rapidly with total concentration just above the cmc but slowly after about 200 mM, as shown in Fig. 6. Increasing r_{cut} yields modest increases in average micelle size. The Alk-SPC model, in spite of having weaker attractions as reflected by the higher cmc, favors somewhat greater average micelle sizes than the HHA model. Concentration-dependent size distributions for the Alk-SPC model will be considered in more detail in section 3.3.4.

The concentration of unimers at equilibrium with micelles, or total concentration of unimers and small "premicelle" clusters, is often used to estimate the cmc. Pitfalls associated with this definition have been identified in the literature: the unimer concentration tends to drop off with increasing concentration above the cmc for ionic surfactants,[36] and excluded volume [57] influences the effective concentration. In simulations, finite-number effects also play an important role;[6] it is therefore interesting to compare estimates results obtained from simulations of small systems to the cmc extracted from the global PEACH fits to cluster statistics. In Fig. 3.6, the mean concentration of OPC present as unimers and small clusters can be seen to oscillate as the total number of OPC is varied, increasing at low N until the first micelle is formed and then dropping sharply. In the regime after this drop, where the unimers and small clusters are at equilibrium with a micelle, their concentration ranges from 150 mM to 225 mM in the Alk-SPC system and from 125 mM to 180 mM in the HHA system. In such a small system, the unimer concentration is strongly influenced by the number of monomers left over after a micelle of optimal size is formed, and so indeed can only be used for a very rough measure of cmc.

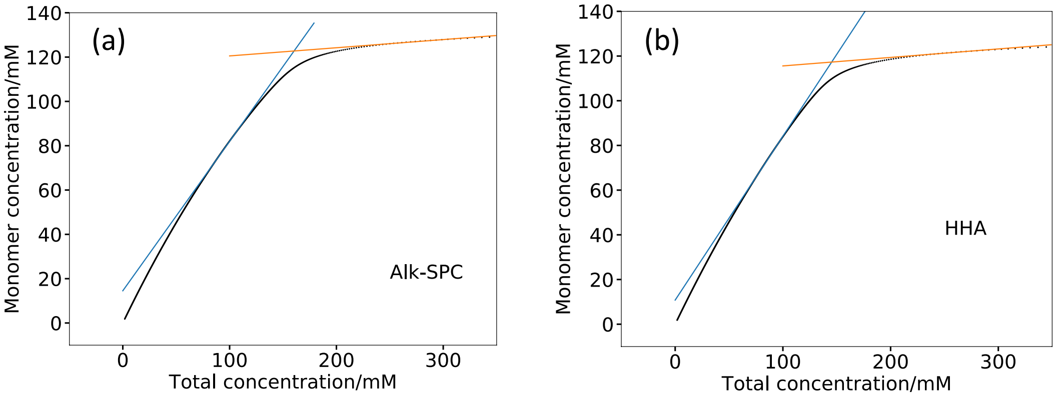


Figure 3.5 Free monomer vs. total OPC concentration calculated from PEACH-derived cluster free energy profile (a) with Alk-SPC (predicted cmc is 160 mM) (b) with HHA (predicted CMC is 145 mM)

Table 3.4: cmc values in mM predicted from PEACH free energies by force field and r_{cut}

Cutoff/nm	Alk-SPC	HHA
0.45	160	145
0.50	158	145
0.55	156	144
0.60	152	143
0.65	141	140

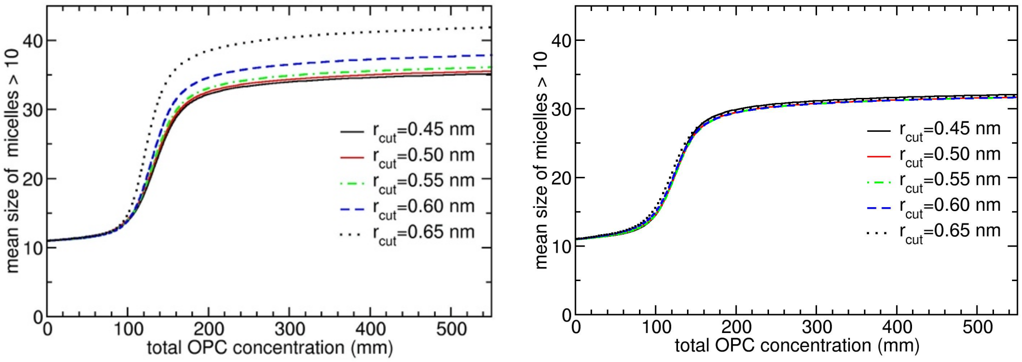


Figure 3.6 Mean micelle size vs. concentration, as calculated from free energies derived from MD simulations performed with a) Alk-SPC force field and b) HHA force field. Different curves represent predictions based on free energies derived using different tail site cut-off distances r_{cut} in the neighbor criterion for defining clusters.

Keeping the roughness of this measure in mind, we will use it now to make a rough comparison of the ability of the HHA model and the original TraPPE force field to reproduce the experimental cmc. (Having only performed two simulations using the original TraPPE CH_X/OW Lennard-Jones parameters, we do not have sufficient cluster statistics for a full PEACH analysis.) Two 200 ns trajectories of OPC using the original TraPPE parameters (with 40 and 30 OPC + 312 solvent) yielded concentrations of monomers in unimers and small clusters of 64 mM and 39 mM respectively. These values fall below the corresponding levels from HHA simulations (Fig. 7) by factors of 3.2 and 2.3 respectively. This suggests that full PEACH analysis of the TraPPE model would give a cmc roughly 3.2-2.3 times lower than the HHA model, *i.e.* in the range 44-62 mM. Such a range, about a factor of two lower than the experimentally reported cmc value of 114 mM, would be consistent with a previous report that surfactants modeled with TraPPE tails showed cmc values a factor of two lower than experiment.[36] We can conclude that tuning the force field to match hydration free energy, as was the goal in the HH-Alkane force field development[70], improves agreement with the experimental onset of micelle formation.

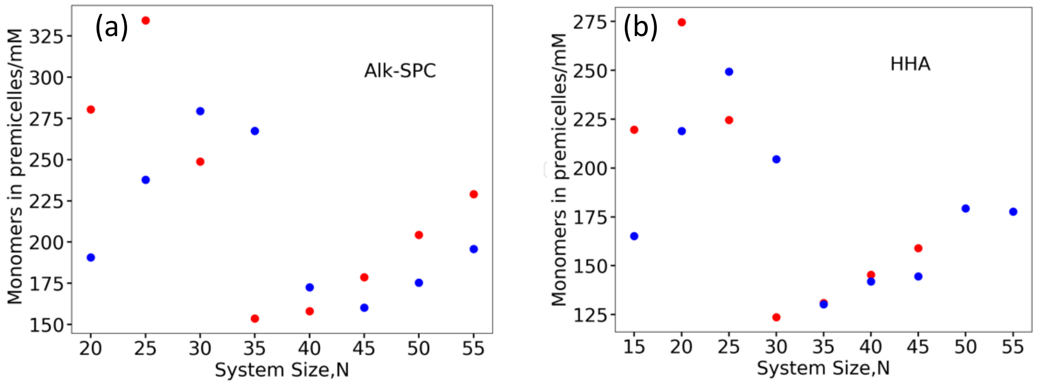


Figure 3.7. Concentrations of monomers in premicelles (*i*<=10) averaged over MD trajectories with varying total number of OPC using (a) Alk-SPC and (b) HHA force fields. (red dots for Set One, blue dots for Set Two as tabulated in Table 3.1.)

3.3.3 Effects of limited system size and results of large system simulation

The up-turn in cluster formation free energies evident at large cluster size *i* evident in Fig. 3.3 suggests that, if the phenomenological models for spherical micelle formation (like the quasi-droplet model) are applicable in this system, micelles much larger than ~50 monomers should be uncommon in solution. A large system (250 OPC, 30330 SPC; 0.46 molal) was simulated for 200 ns to test whether this is the case. Figure 8 indicates that the answer is not a simple yes or no. The cluster size histogram (Fig. 3.8a) does show a drop-off above *i* = 50, qualitatively consistent with the increase in free energy derived from small-*N* simulations; however, a long tail in the distribution extends to significantly higher cluster sizes. The nature of these larger clusters is suggested by a snapshot (Fig.3.8b) of the last frame of the trajectory, showing modest sized clusters that appear to be sticking together. The time dependence of the maximum cluster size (Fig. 3.8c) shows that the lifetimes of the large clusters are very short, consistent with their formation by transient bridges between more persistent smaller structures. The strong sensitivity to r_{cut} of the levels of large clusters is also consistent with an agglomeration of distinct hydrophobic cores whose alkyl tails only occasionally approach each other, rather than a continuous hydrophobic core connected through many pathways whose redundancy lowers sensitivity to fluctuations. The structures formed are not simple aggregates of the micelles present in the starting configuration of the simulation, but are thoroughly mixed as shown in an animation with the initial micellar components shown in different colors. We conclude that the large clusters formed are in fact aggregates of small clusters, connected by a partially solvated double layer of headgroups and transiently bridged by tail sites. Such a hierarchical structure has been invoked to explain the high flexibility of certain large micelles[88] or as a transient intermediate in the "sticky-collision" mechanism of dye exchange between micelles[89] The free energies of association of these higher-order clusters cannot be extrapolated from either the single-micelle simulations or the single-micelle theoretical models. Even considering the individual micelles that make up these higher-order clusters, their size-dependent free energies are likely to be perturbed away from what predictions based on free, fully-solvated micelles would give. Unfortunately, the computational expense required for long simulations at the large system sizes required prohibit a full quantitative investigation of these effects. Given that inter-micellar interactions should be weaker than intra-micellar interactions, it is likely that clustering would not become important until micelles start to become crowded together. Predictions related to the onset of micellization should therefore not be influenced by this phenomenon.

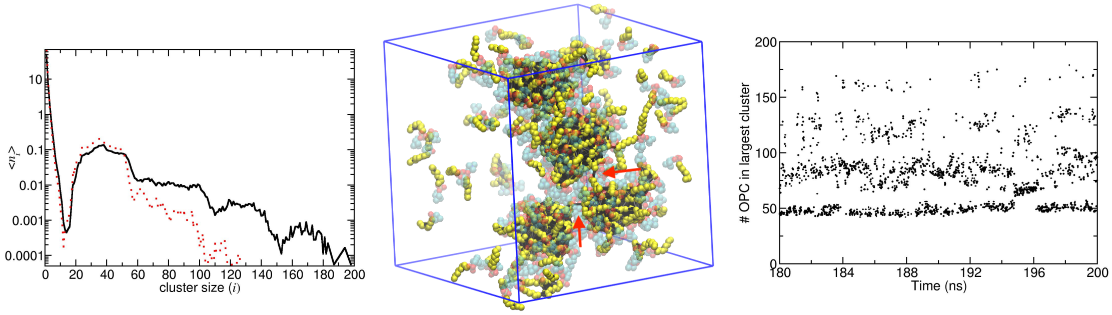


Figure 3.8 Results from 200 ns simulation of 250 OPC / 30330 SPC water. a) Cluster histograms generated from 20-200 ns using $r_{\text{cut}} = 0.45$ nm (red dashed curve) and $r_{\text{cut}} = 0.60$ nm (heavy black curve). b) Snapshot of last frame showing tail sites in yellow, headgroup sites in transparent colors, and “neighbor ties” connecting tail sites nearer than 0.60 nm in black. Red arrows point to neighbor connections bridging micelle clusters. c) Detail from 180-200 ns of the time dependence of the largest cluster size as evaluated using $r_{\text{cut}} = 0.60$ nm.

Whether inter-micelle effects of this nature are important experimentally for OPC or are an artifact of the headgroup force field parameters is not clear. On the one hand, this could explain neutron scattering results[90] that suggest dehydration and increased core-shell mixing in dodecyl phosphocholine (DPC) micelles (sharing the same headgroup as OPC but with 12-carbon tails) upon increasing concentration from 10 mM to 100 mM. On the other hand, formation of persistent supermicellar clusters would not be indicated by the reduction in apparent hydrodynamic radius obtained through dynamic light scattering in the same report.

3.3.4 Modeling SAXS data

Small and wide angle X-ray scattering (SAXS/WAXS) is a common technique to characterize the size and shape of micelles[91, 92]. The contribution of individual micelle structure to the overall scattering profiles, the "form factor," can be related to Fourier transforms of the electron density distributions, averaged over configurations and orientations of micelles. Information about cluster size and general shape can be inferred from fitting experimental scattering profiles to a simple model, for instance a two-shelled ellipsoid with the interior representing the hydrophobic interior and the outer shell representing the headgroups (which, due to the presence here of phosphorus, have significant contrast in electron density against the interior or the solvent). Known complications in interpreting the scattering profile include correlations between cluster positions (which grow in importance with concentration, as discussed in the previous section), polydispersity, and variations in mean micelle size with concentration[91]. Efforts have been made to generate scattering data from molecular simulation based on both implicit solvent[93] and explicit solvent[94] models.

Faramarzi and coworkers have used the WAXSiS server[2, 95] to generate scattering profiles generated from simulation coordinates can be used to select a micelle size that is most consistent with experiment.[94] In their report, they have suggested that imperfect agreement between the profile generated from simulation data (using the WAXSiS server) for dodecyl phosphocholine could be the result of polydispersity. We were unable to find SAXS data for OPC in the literature for direct comparison with the simulation results, but instead here we aim to address the more general question: can the combined scattering from a polydisperse equilibrium mixture of micelles be adequately represented by the scattering profile of a single typical micelle size?

Fig. 3.9(a) shows the model which the WAXSiS server used to calculate the scattering profile. The hydration layer and excluded solvent were constructed from a short trajectory of simulation with position-restraining potentials to keep the structures close to initial structures.[2] As is shown in Fig. 3.9(b), variations were observed for SAXS curves of the same micelle size. This variation is due to the slightly different configuration of a micelle, thus an average of multiple curves for each micelle size was calculated. In these averaged SAXS profiles (Fig. 3.9(c)), a shift to lower q was observed with the increase of micelle size. The peaks are in the range of 2.3-2.7 nm^{-1}. Table 3.5 gives the radius of gyration calculated on the WAXSiS server based by Guinier analysis, which shows the radius is about 1.97 nm for a cluster size of 45. These features of OPC are consistent with previous experimental and simulation results on dodecyl phosphocholine.[92, 94]

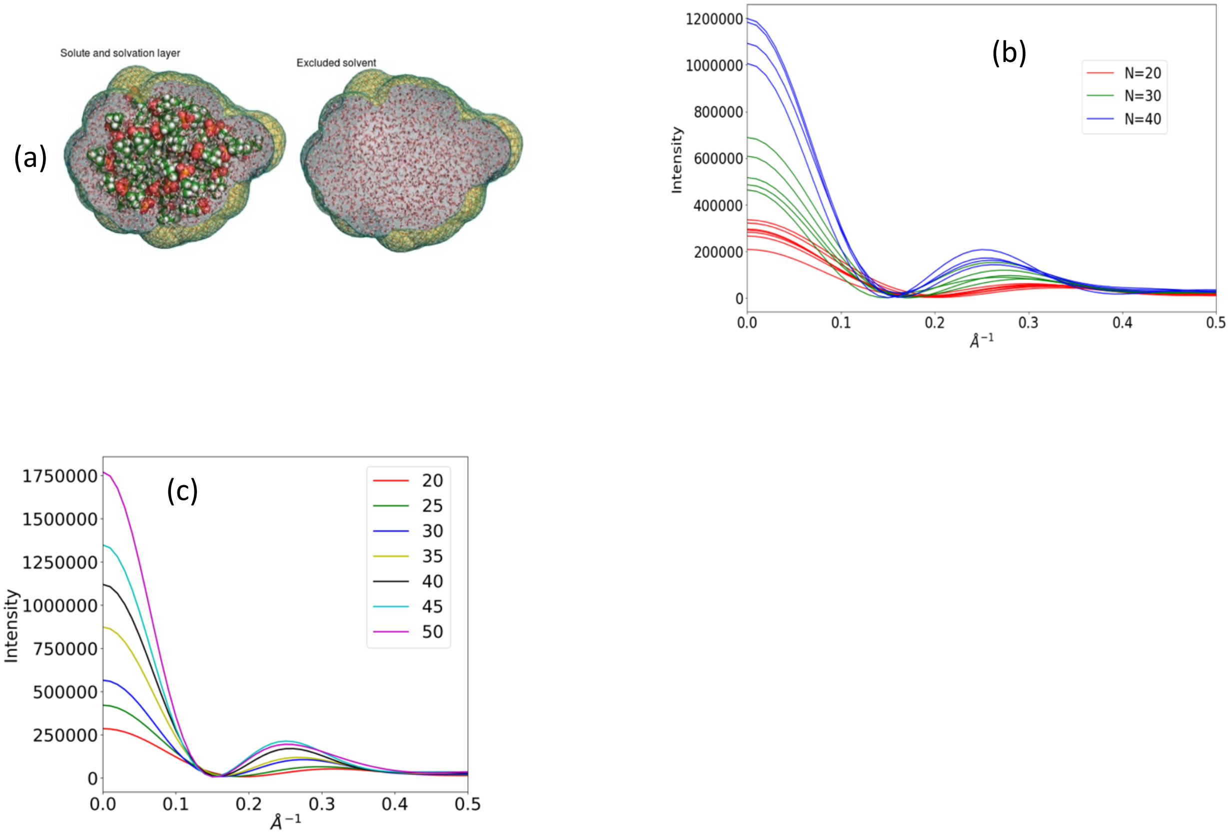


Figure 3.9 (a) The model of OPC micelle generated by WAXSiS server (b) Comparison of SAXS scattering profiles for different configurations of micelles with micelle size N=20,30,40 (c) Averaged SAXS scattering profiles for each cluster size from 20 to 50 at intervals of 5.

Based on the cluster free energy profile obtained from PEACH method, we can calculate the cluster size distribution for polydisperse systems for chosen concentrations. Fig. 3.10(a) shows the cluster size distribution at total OPC concentrations 198 and 787 mM. Composite SAXS curves were calculated using Eqn 9. Concentrations for each aggregation number *i* (curves in Fig. 3.10) are combined into weighting coefficients (bars in Fig. 10) for *i* = 20, 25, 30, etc. according to Eqn. 9. These coefficients are used in calculating a weighted sum of scattering profiles (Fig. 3.9c) to yield an approximate composite representation of the scattering from the size distribution. Premicelles for cluster sizes below 13 were not counted as they do not have SAXS features in the spectral range of interest. To compare this polydisperse scattering curve with ones for monodisperse micelle, the curves of N=30,35,40 were normalized based on I(q=0) to compensate the concentration variation. As seen in Fig. 3.11(a), the scattering profile of the mixture is quite similar to the profile of monodisperse 35-mers, with no obvious signature of the polydispersity presenting itself as a feature. From this we conclude that the deviation observed between experiment and the WAXSiS-derived results by Faramarzi *et al.*, where is unlikely to have its origin in polydispersity.

The specific deviation that was observed was overprediction by the model of the intensity of low-q scattering relative to a peak at q~1.7 nm^{-1}. We note that the ratio of intensities of the peak at q=0 to the peak at 1.7 nm^{-1} for our OPC structures is quite sensitive to even small changes in the choice of background solvent electron density used in the WAXSiS calculation. The default selection for this parameter is 334 nm^{-3} [95] whereas for the buffer used in the experiments is reported[91] to have slightly higher density at 3.4×10^2 nm^{-3}. Increasing the background solvent electron density from 334 to 340 nm^{-3} lowers the zero-q peak by 17% without affecting the height of the q=1.7 nm^{-1} peak. (not shown) This is not to say that the difference in density between solvent and buffer accounts for the discrepancy between simulated and experimental scattering profiles, rather that this element of the scattering profile is highly sensitive to details of contrast in electron density.

Table 3.5 Average radius of gyration for each cluster size

Micelle Size	Average radius of gyration(Å)
20	15.8443
25	17.4305
30	18.7757
35	19.1788
40	19.7028
45	20.2235
50	20.8528

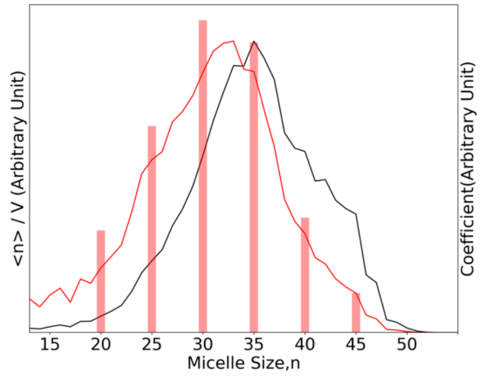


Figure 3.10 Micelle size distribution for c_{tot}= 198 mM. [$c_{monomer}$= 127 mM] and c_{tot}= 787 mM [$c_{monomer}$= 141 mM]. The bars represent the weighting coefficients for each group of cluster sizes for c_{tot}= 198 mM.

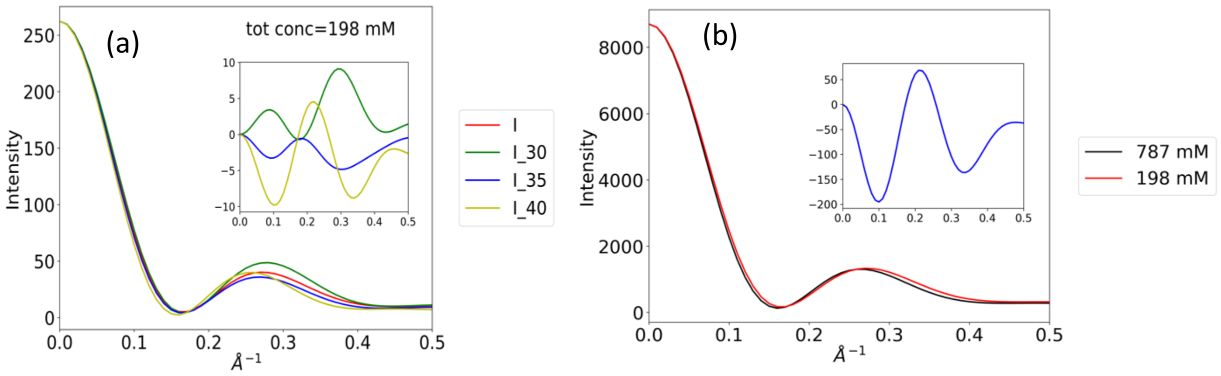


Figure 3.11 (a) Predicted SAXS scattering profile for polydisperse mixture at 198 mM, compared with contributions from individual size components; residual error is shown in inset. (b) Predicted SAXS scattering profile based on composite weightings at 198 mM and 787 mM. All profiles are scaled to have equal $I(q=0)$.

Even if the scattering profile at a single composition is not very sensitive to the breadth of the micelle size distribution, this underlying ΔG_i is reflected in how concentration affects the average micelle size. As we consider how concentration affects the scattering profile, we first note that the contribution from the structure factor (not considered here) will be strongly affected by concentration. Considering the form-factor alone, increasing concentration drives a small shift in distribution of aggregation numbers as shown in Fig. 3.10(a), with the peak size increasing from 33 to 35. The result is a slight shift in the small peak towards lower q in the calculated form factor (Fig. 3.11b).

3.4 Conclusions

In this work we used molecular simulation to study the micellization equilibria of octyl phosphocholine using the HH-Alkane force field and the newly developed Alk-SPC parameters to represent tail-tail and tail-water interactions. PEACH analysis was used to obtain the free energy of association over a broad range of cluster sizes, which followed a dependence on aggregation number that was well-fitted by the quasi-droplet model for micelle assembly. The free energy of association was used to predict the concentration-dependent onset of micellization and cluster size distributions. The critical micelle concentration was in fair agreement with experiment; the HH-Alkane force field showed distinct improvement relative to TraPPE parameters on which it was based, reflecting an improved representation of the hydration free energy of the alkyl tails. Although the optimal cut-off distance to be used in defining a cluster is not clear, a range from 4.5 nm-6.0 nm produced only modest changes in predicted micellization properties. Use of the small-N approach offers simplicity but introduces the risk of neglecting features that emerge from larger-scale simulations, such as the significant tendency for micelles to cluster into higher-order aggregates seen here in a simulation of 250 OPC.

Scattering profiles generated via the WAXSiS server from individual micelle structures were combined, using weights derived from the PEACH analysis, to represent a composite form factor at different OPC concentrations; it was found that only minor differences separated the composite scattering profile from the profile generated using a single typical micelle size. Further study of the dynamics of monomer exchange may be useful combined with these equilibrium calculations to provide details for general kinetic models of micellization.[96]

3.5 Acknowledgments

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Comet cluster at the San Diego Supercomputer Center, which is supported by National Science Foundation Grant No. ACI- 1548562, Allocation No. TG-MCB110144. This work also used the resources of the Cherry L. Emerson Center for Scientific Computation.

Chapter 4 Free Energy of Micellization of Dodecyl Phosphocholine (DPC) from Molecular Simulation: Hybrid PEACH-BAR Method

4.1 Introduction

For surfactants with low critical micelle concentrations (CMC) i.e. less than 10 mM, this procedure is compromised as a result of the slow dissociation rate (relative to 100's of ns) and rare occurrences of unimers and premicelles in a moderately sized solvent bath (with 1000's of waters). To obtain good sampling of clusters of varying size, simulations need to be performed at low concentrations around the CMC; however, the large amount of solvents consumes excessive computational costs. Higher concentration simulations with small box sizes produce trajectories in which a micelle grows to encompass all monomers and remains in that state throughout the trajectory.

To allow the PEACH method to analyze simulations of a wide range of aggregation tendencies, here we combine it with the Bennet Acceptance Ratio (BAR) method for calculating free energy differences between systems under different potential energy functions. We performed MD simulations of the surfactant systems using reduced Lennard-Jones attractions between alkyl tail groups to promote dissociations of monomers from micelles. The cluster free energy for this reference system of weakened interaction strength can be calculated using the PEACH method, since the kinetics of dissociation are faster and since fluctuations in cluster size can be observed at equilibrium in a system with less solvent. Using BAR calculations, the free energy differences between the full strength (FS) and interaction-weakened systems (WS) can then be determined, to recover cluster free energies at full strength of interactions. Here we have applied this PEACH-BAR method to obtain the cluster free energy curve for dodecyl phosphocholine (DPC), for which a CMC of about 1 mM has been reported in experiment.10, 11 In this report, we introduce the concepts and implementation of the PEACH-multistate-BAR method. Then we apply the new method to simulations of DPC, checking the internal consistency of the method using two different WS systems. Comparison of results against an independent free energy calculation obtained through umbrella sampling led us to alter the cluster definition criterion to include headgroup-headgroup interactions. The resulting PEACH-BAR free energy curve is in fair agreement with experiment with respect to the predictions it yields for CMC and cluster size.

A new method to extract the free energy of aggregation vs. aggregate size has been proposed and applied to results generated from molecular dynamics simulations of dodecyl phosphocholine in SPC water. Due to the relatively low dissociation rate and critical micelle concentration, extraction of cluster free energies directly from simulation results using the "partition-enabled analysis of cluster histogram" (PEACH) method is not feasible for these types of surfactants. The new approach applies PEACH to a model with weakened attractions between aggregants, for which the dissociation rate is fast enough to obtain sampling of all cluster sizes, and then recovers the free energy for the fully-attractive force field using the BAR free energy difference method. Results from the PEACH-BAR method were compared with free energy calculations employing umbrella sampling to find the free energy associated with removing a single surfactant from a micelle. The critical micelle concentrations and average cluster sizes are in general agreement with the experimental results. The theoretical predictions of SAXS profile were compared to experimental SAXS results and good agreement confirmed accuracy of the structures generated from simulations.

4.2 Method

4.2.1 Forcefield choices

United-atom (UA) force fields were used for this simulation. Lennard Jones (LJ) parameters for pairwise interactions between CH_x groups with other CH_x groups were taken from TraPPE-UA model[72]. For LJ interactions between CH_x groups and water oxygen, we used parameters from the Alk-SPC model[97], which was parameterized from the TraPPE-UA model[72] to allow us simulate interactions between alkyl tails and SPC water following the development of the HH-Alkane force field[70]. Parameters from the Gromos G54A7 force field[73] [83, 98] were used for other nonbonded interactions and all bonded interactions (bond stretching, bending and torsional potentials).

4.2.2 Simulations of DPC

Simulations are performed with GROMACS 5.0[75] software package. A series of simulations of different concentrations (containing 15-80 DPC with varying numbers of SPC water) were set up to obtain sufficient sampling across the full range of cluster sizes $2 < i < 65$, where i is the number of DPC in the cluster. The total numbers of DPC and solvent for all trajectories are tabulated in SI Table 1. Two WS versions of the force field with weakened attractions between CH_x sites were used to accelerate the monomer exchange process. Specifically, for all CH_x-CH_x interactions, the C_6 coefficient of the r^{-6} Lennard Jones was decreased to 85% or 90% of its full-strength value and the C_{12} parameter was adjusted to maintain the thermal radius according to equation (1)[79] .

$$r_{iw}^{therm} = \sigma_{iw}[\frac{2}{1 + (3kT/\varepsilon)^{1/2}}]^{1/6} = [\frac{2 * C_{12}}{C_6 + (C_6^2 + 4 * C_{12} * 3kT)^{\frac{1}{2}}}]^{1/6} \tag{1}$$

The aim in adjusting these parameters was to non-specifically lower the mean attraction between sites while minimizing perturbation to the equilibrium distribution of distances, with an eye toward achieving good probability overlap between the weakened and full strength structures. To achieve this, we defined the thermal radius as the distance where the potential reaches 3kT instead of 1kT, to avoid large differences in force over frequently accessed ranges of the potential. We use "WS=0.85" and "WS=0.9" to denote the interaction-weakened systems and "FS" to denote the full-strength system. Fig. 1 shows a comparison between the three Lennard-Jones functions (panel A) and their corresponding Boltzmann factors (panel B). With these two decreased strengths, we are able to sample the full cluster size range for reduced strengths and recover the full-strength cluster free energy profile based on BAR method mentioned in the following section. For interaction strength larger than 90%, the attractions between alkyl tails hold the micelles together and monomer exchange events are too rare for us to obtain PEACH statistics. And for interaction strength smaller than 85%, the micelles were held loosely and the overlap of configurational space was not enough for reliable BAR analysis. It is crucial to note that the interactions of all DPC sites with solvent were kept identical in all three systems.

The pressure is maintained at 1 bar and a compressibility of 4.5×10^{-5} by Berendsen barostat[80] with $\tau_P = 2\ ps$. The temperature was maintained at 300K by velocity rescaling thermostat[56] with $\tau_T = 2\ ps$. The md integrator with a 2 fs time step was used for integration of equations of motion. The Verlet[78] cutoff-scheme was applied for short-range non-bonded interactions with a cutoff of 1.2 nm. With the exception of the alterations described above, the Lennard-Jones potential was used to calculate non-bonded interactions with the Lorentz-Berthelot combination rules accounting for different UA groups as shown in equation (2). Coulomb interactions were accounted for by Particle mesh Ewald summation[55] with a real space cutoff of 1.2 nm.

$$\sigma_{ij} = 0.5\left(\sigma_{ii} + \sigma_{ij}\right)$$

$$\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$$

(2)

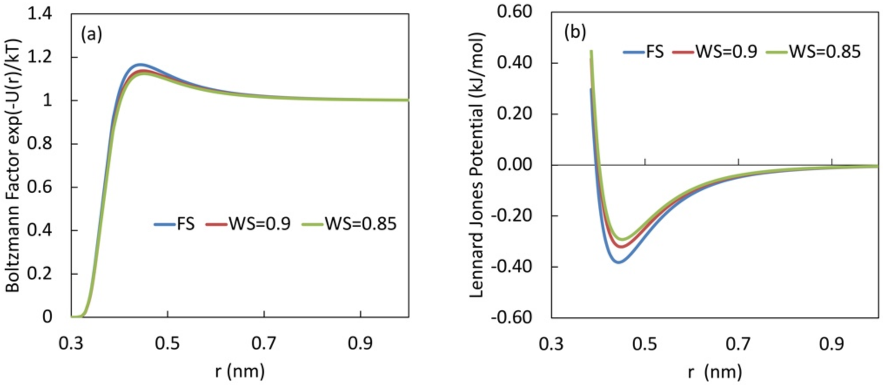


Figure 4.1 (a) LJ potential energy and (b) Boltzmann factors for CW2-CW2 interaction for WS = 0.85 and WS = 0.9, and full-strength system FS.

4.2.3 PEACH-BAR analysis

An 800 ns simulation with a 20 ns pre-equilibration was performed for each trial and used for PEACH analysis. For some trials, the trajectory length and pre-equilibration time were adjusted based on how fast the system reached equilibration and how well the system generate sampling statistics for all cluster sizes observed. For the initial cluster definition, two DPC chains are considered to be of the same micelle if the distance between any of their alkyl tail beads is lower than 0.50 nm. (This cutoff is determined based on results in the previous work for octyl phosphocholine[97]) In a revised cluster definition, the same distance cutoff criterion is used but applied to the distance between any pair of sites on the molecule. The average frequency of occurrence for clusters containing *i* surfactants, denotated as $<n_i>$, was calculated over all frames in the trajectory.

The PEACH method allows extraction of a set of equilibrium constants K_i that optimally reproduces the cluster size distributions $\{<n_i>\}$ observed in one or more simulation trajectories, using a algorithm and an iterative fitting process that we have described previously. The Gibbs free energy of clustering for standard state i-mer can be calculated from {K$_i$} with equation (3)[32]. To represent the output from PEACH in a convenient way, we calculated the Gibbs free energy for a known concentration of i-mer based on equation (4)[32], under the assumption that we have an equal concentration c of i-mer and free DPC monomer.

$$\frac{\Delta G_i^{\circ}}{k_B T} = -ln\, K_{eq,i} \tag{3}$$

$$\frac{\Delta G_i}{k_B T} = -\, ln\, K_{eq,i} - (i-1)\ln\left(\frac{c}{c^{\circ}}\right) \tag{4}$$

With good PEACH fit for interaction-weakened systems “WS = 0.85” and “WS = 0.9”, we can obtain $\Delta G_{i,WS=0.85\ or\ 0.9}$. To find the free energy of clustering in the FS system, we then use the BAR approach shown schematically in Fig 2 and equation (5). $\Delta G_{i,WS=0.9}$ is the Gibbs free energy of forming a same-size micelle in an interaction-weakened system, calculated from PEACH method. $\Delta G_{i,0.9\rightarrow FS}$ is the Gibbs free energy difference for a micelle of size *i* under the interaction-weakened potential versus a full-strength potential, which can be calculated by the BAR (Bennett’s Acceptance Ratio) method[76] or multistate BAR (MBAR) method.[99]

$$\Delta G_{i,FS} = \Delta G_{i,WS=0.9} + \Delta G_{i,0.9\rightarrow FS} - i * \Delta G_{1\,,0.9\rightarrow FS} \tag{5}$$

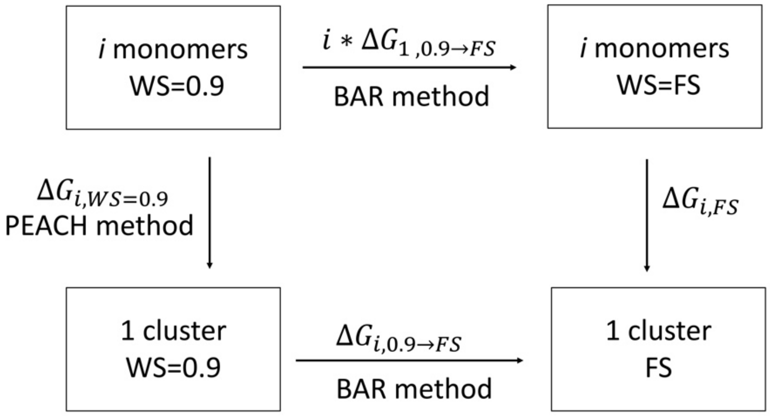


Figure 4.2 Scheme for PEACH-BAR method.

The general idea of free energy methods is to calculate the free energy difference between states of which configuration spaces do not fully overlap, sometimes with intermediate states incorporated to avoid the error from unaccounted sampling. Bennett's Acceptance Ratio method[76] was brought up to calculate free energy difference, ΔA_{ij}, based on configurations of two states i and j. From the derived equations, ΔA_{ij} is calculated by finding solutions for equation (6).[100]

$$\sum_{i=1}^{n_i} \frac{1}{1+\exp\left(\ln\left(\frac{n_i}{n_j}\right)+\beta\Delta U_{ij}-\beta\Delta A\right)} - \sum_{j=1}^{n_j} \frac{1}{1+\exp\left(\ln\left(\frac{n_j}{n_i}\right)+\beta\Delta U_{ji}-\beta\Delta A\right)} = 0 \qquad (6)$$

n_i and n_j are the numbers of samples at each state. ΔU_{ij} is the potential energy difference between sampled state i and weakened state j when the simulations are performed on the configuration space with a distribution proportional to $e^{-\beta U_i}$, with ΔU_{ij} vice versa.

For our purpose, a few micelle sizes (including 1, 3, 8, 9, 14, 15, 19, 20, 24, 25, 29, 30, 34, 35, 39, 40, 44, 45, 49, 50, 54, 55, 59, 60, 64, 65, 69, 70, 74, 75, 79, 80) were chosen to perform BAR analysis. From the simulation trajectory for FS, WS =0.85 and WS=0.9, the frames containing these sizes of micelles was taken out to generate configurations for these cluster sizes. Then for each pair of states (FS and WS=0.85, FS and WS=0.9, WS=0.9 and WS=0.85), we used the rerun option in Gromacs mdrun utility to calculate the potential energy for the configurations of one state under the interaction strength of the other state in the pair and vice versa. Once ΔU_{ij} and ΔU_{ji} were calculated, ΔA_{ij} was obtained by using a bisection method to find solutions for equation (6). To calculate free energy difference using configurations from multiple states, Shirts and Chodera[99] developed codes to implement the multistate BAR method. We also implemented MBAR method with the calculated ΔU_{ij} and ΔU_{ji} to calculate ΔA_{ij} and compare with BAR results.

4.2.4 Umbrella Sampling Analysis

With Gromacs 5.0 package[75], the “pull code” was used to generate configurations for umbrella sampling. One DPC was pulled out from a micelle of (N-1) DPC. The distance between the two groups was determined by their center of mass. From the simulation trajectory of pulling, 25 configurations were selected as the initial configurations for the following umbrella sampling simulations, and later a few windows were added to fill in poorly sampled regions. The configurations were selected to have an increment of ~ 0.1 nm in distance between the pulled DPC and the DPC micelle. Each umbrella sampling takes 100 ns. The simulation setup for pressure, temperature and cutoff schemes are the same as PEACH-BAR simulations. The parameters for setting up potential of mean force (PMF) analysis via umbrella sampling are shown in Table 4.1.

Table 4.1 Setup parameters for umbrella sampling analysis

Trial	Number of DPC (N)	Number of SPC Water	Box size (nm^3)	Pull rate (nm/ps)	Pull Force Constant ($kJ\cdot mol^{-1}\cdot nm^{-2}$)	Pull Time(ns)	Umbrella sampling windows	Umbrella sampling time (ns)
1	30	10145	335.56	0.0015	1000	2	26	100
2	40	10322	341.41	0.0002	1000	15	33	100

Based on the Weighted Histogram Analysis Method (WHAM)[101], the potential of mean force is calculated from the pull force of each umbrella sampling windows, and the PMF of the fully dissociated monomer is set to zero. The equilibrium constant for the association of the monomer to the remainder of the micelle can be calculated by integrating the PMF over a spherical volume as in equation (7):

$$K_{n-1\to n} = n^{-1}\int_0^{r_{cutoff}} exp\,(-\frac{PMF}{k_BT})(4\pi r^2)dr \tag{7}$$

The factor of n^{-1} is needed to correct for the fact that the umbrella sampling approach treats one monomer as distinguishable from the others. The addition reaction for indistinguishable monomers and for a reaction with one "tagged" monomer are represented as follows respectively:

$$A + A_{n-1} \rightleftharpoons A_n$$
$$A^* + A_{n-1} \rightleftharpoons A^* A_{n-1} \tag{8}$$

Whereas the forward rate constants of both reactions must be identical, the rate constants for the reverse process of the first reaction will be n times greater than that of the second, simply because there are n identical monomers available to dissociate instead of a single tagged monomer. Thus, for an addition reaction in a micelle, the equilibrium constant calculated using potential of mean force with a tagged molecule is n times greater than the true equilibrium constant, and thus the free energy of the reaction is lower by $kT \ln n$. (A different route to the same result is to recognize that the partition function of the cluster A^*A_{n-1} is n times greater than partition function of the cluster A_n.) The same correction should be applied for alchemical routes to calculating the same free energy, but has at times been neglected.[45]

4.2.5 SAXS data prediction

The small angle X-ray scattering (SAXS) profiles for micelles of size N= 30, 40, 50, 60 were predicted by the WAXSiS server[2] based on short MD simulations on each micelle. The configurations from previous simulations of FS interactions were used as initial configurations for WAXSiS analysis. The pair distance distribution function (PDDF) was calculated by the GNOM utility in ATSAS software. [102]

4.3 Results and Discussions

4.2.1 Observations of simulations

The simulations were visualized using VMD[58] and we observed increasing aggregation as the tailgroup interaction strength is increased. As evident in snapshots (Fig. 4.3), the shape of a globular micelle was maintained in all the simulations; however, we observed more frequent monomer exchange events and a higher mean concentration of free monomers in "WS = 0.85" and "WS =0.9" than in FS.

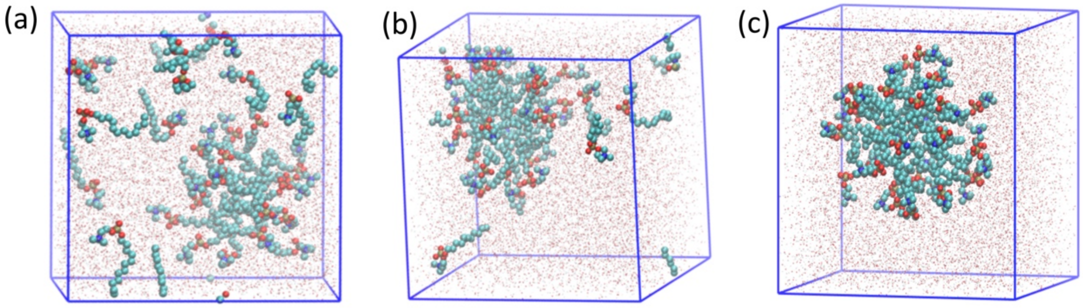


Figure 4.3 VMD view of the simulation box. N=30 (a) WS=0.85 (b) WS= 0.9 (c) FS. For representation purpose, the size ratio of the atoms of DPC and water was set to 6:1.

4.3.2 Cluster Size distribution

Micelle size distributions for the full-strength system and two weakened systems, WS=0.85 and WS=0.9, were calculated from a series of simulation trajectories generated with different total surfactant numbers. The aim of setting up multiple simulations for the same weakened strength was to obtain good sampling statistics for every micelle size across the range of distribution to ease PEACH analysis for cluster free energies. The total surfactant number in each simulation box was chosen to make the cluster size histograms (Fig.4.4) overlap across the trajectories. As the strength of alkane tailgroup interaction increases from 85% to 90% of the full strength, the larger micelles were kept in the solution for a longer time. This trend is consistent with narrower observed cluster size distributions (Fig 4.4) and larger cluster sizes (Appendix V. Fig.1) for simulations of the same total surfactant numbers with increasing tailgroup interaction strengths.

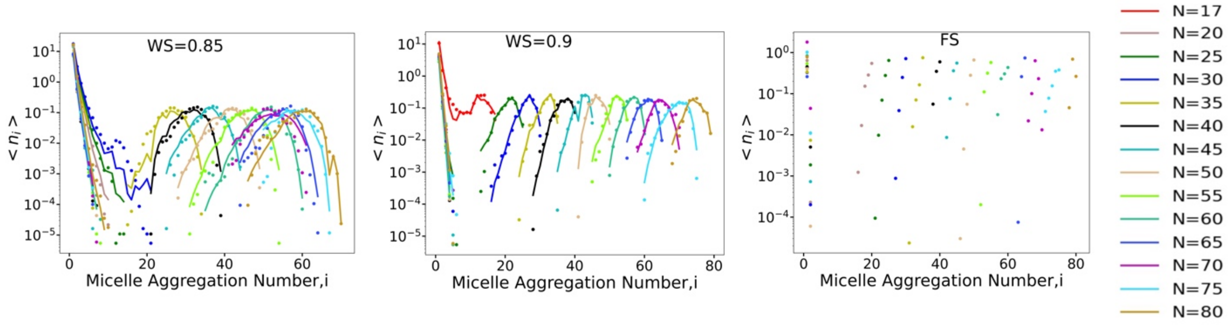


Figure 4.4 Micelle size distribution of simulations (symbols) and the PEACH fit (curves) of DPC with force fields labelled on the figure and number N of DPC indicated by color legend.

4.3.3 Results of PEACH analysis on systems with weakened attractions

With PEACH method, a global iterative fitting procedure was used to determine the set of equilibrium association constants $\{K_i\}$ to optimally fit the observed cluster size distributions. Fittings are shown in Figure 4.4 (a) and (b) for the WS=0.85 and 0.9 series; PEACH analysis of the FS trajectories was not attempted because overlap between histograms from different trajectories was not obtained. It is interesting to consider how the choice of weak reference potential affects the efficiency and reliability of the PEACH results. Our prediction was that the system with weaker interactions (WS=0.85) would prove more useful because the rate of monomer exchange and the overlap between histograms is greater, giving better statistical sampling. In fact, choosing a weak interaction strength has the drawback that some less-stable aggregation numbers are difficult or impossible to observe because they simply fall apart; as a result, the *i*=15-20 range is less-well sampled at WS=0.85 and the *i*>70 range is missing. The PEACH fits for WS=0.9 are generally closer in agreement with the raw data than for WS=0.85. This may reflect better accuracy or may simply be because they are less tightly constrained; each micelle size is represented in more trajectories, on average, at WS=0.85 because the distributions are broader.

Fig.4.5 shows the size-dependent cluster free energy profiles $\{\Delta G_i\}$ for WS=0.85 and WS=0.9, which were derived from the best-fit set of $\{K_\mathrm{i}\}$. These show the expected qualitative trend predicted by simple models[59, 86] and seen for octyl phosphocholine[97].

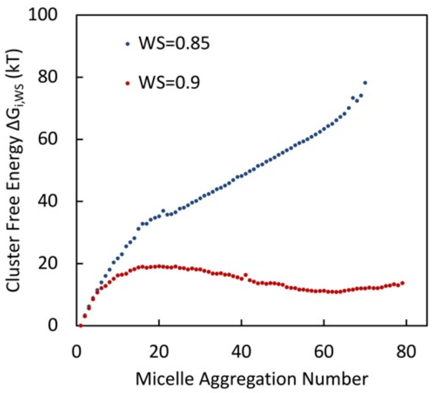


Figure 4.5 Free energy of cluster association ΔG_i versus aggregation number *i* obtained from PEACH analysis for WS=0.85 and WS=0.9, both were calculated for the condition of monomer concentration c = 33.1 mM

Although the range and efficiency of the PEACH-BAR method may depend on the choice of WS parameters, the final results should not; the differences in PEACH cluster free energy for the two WR systems should be counterbalanced by the differences in BAR free energies to yield the same FS system free energy.

4.3.4 Results of BAR calculations of free energy differences

The BAR and MBAR methods were applied to calculate free energy differences between WR and FS potentials for the initial (free monomer) and final (micelle) states of the system. The accuracy of BAR calculations depends on the overlap of configurational distributions sampled by the two systems, which can be illustrated by overlapping histograms of ΔU_{ij} sampled. Overlap was good among all three systems except between WS=0.85 and FS at larger cluster sizes. (Appendix V, Fig.4.2) As a test of self-consistency, we compared the simple BAR difference in free energy between WS=0.85 and FS systems with the sum of BAR differences of WS=0.85 with WS=0.9 and of WS=0.9 with FS, and found them to be essentially identical up to a cluster size of i=60. Similarly, application of the MBAR algorithm to all three systems gave similar results to the application of the simple BAR algorithm and combination of 0.85→0.9 and 0.9→FS steps, within a difference within 1kT. (Fig. 4.6)

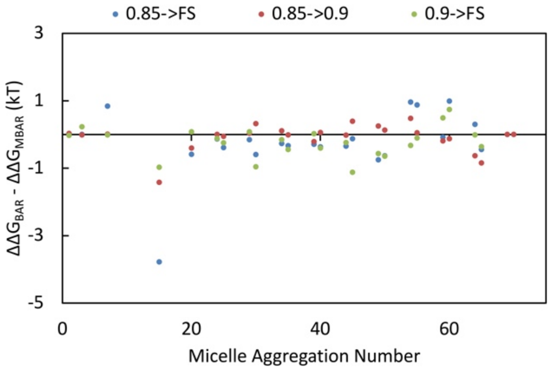


Figure 4.6 The difference of $\Delta\Delta G_{i,WS\rightarrow FS}$ between FS, WS=0.9 and 0.85, calculated from BAR method and MBAR method (with i denoting the micelle aggregation number).

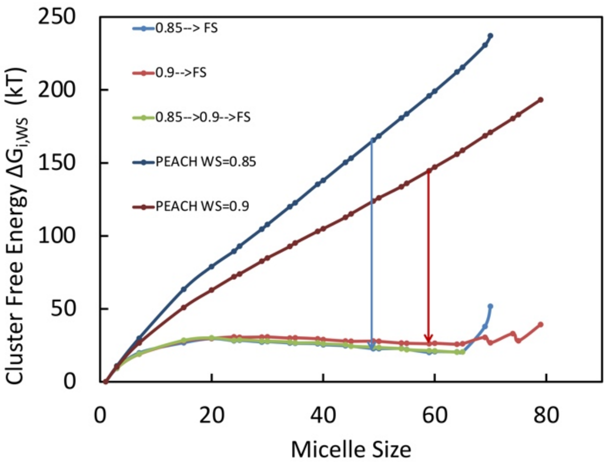


Figure 4.7 Cluster free energy $\Delta G_{i,FS}$ obtained from PEACH-BAR method. All the cluster free energy were calculated for the condition of monomer concentration c=3.3 mM. The arrows showed how the full strength CFE is recovered from CFE of systems with weakened interactions based on the scheme shown in Fig.2.

As shown in Fig.4.7, cluster free energy $\Delta G_{i,FS}$ for cluster size i was therefore calculated by combining $\Delta G_{i,WS}$ for the WS systems from PEACH analysis with the free energy differences across FS and WS force fields from BAR, following the PEACH-BAR scheme shown in Fig.2. The CFE for full strength was calculated from three thermodynamic cycles: (a) "WS=0.9" -> FS (b) "WS=0.85" -> FS (c) "WS=0.85" -> "WS=0.9" -> FS. Agreement among all three curves is good, showing reasonable internal consistency for the PEACH-BAR approach as a whole. In other words, differences in PEACH cluster free energies between the two WS systems are consistent with their differences relative to the FS system calculated using BAR. Cycles (b) and (c) give nearly identical results, reflecting the internal consistency (discussed above) of the BAR calculations across the three force fields. The largest discrepancy among the three sets of results at cluster sizes below 65 is 6 kT, which on a per-monomer basis is well below 1 kT per monomer.

As mentioned above, free energies $\Delta G_{i,FS}$ generated from pathway (b) and (c) are almost identical. Both of them are calculated with the PEACH-generated $\Delta G_{i,WS=0.85}$, plus the BAR-generated free energy difference between states. Because the BAR analysis gives consistent results among the three states, the main source of error causing the difference of the CFE curves generated from the WS=0.85 and WS=0.9 systems is most likely associated with the PEACH-derived CFE of the WS systems. This error might come from sampling issues or from effects of non-ideality; interactions and excluded volume effects involving the free monomers will be more pronounced in the WS=0.85 system because the unimer concentrations are much higher. The error for large cluster sizes (above 60) can be attributed also to poor overlap between BAR energy histograms (see Appendix V Fig. 4.2).

We chose to do further analysis based on the full strength CFE recovered from WS=0.9, as the PEACH analysis gave a better fit and resulted in a smoother free energy curve extending to higher cluster sizes, because BAR energy histogram overlaps were improved, and also because errors due to non-ideality will be lower in the system with lower free monomer concentrations.

4.3.5 Comparison with Umbrella Sampling Calculations

To further validate the PEACH-BAR free energies, we made a comparison with free energy differences calculated using a potential of mean force (PMF) calculation via umbrella sampling (US). The addition or removal of a single monomer to or from a micelle can in principle give information on the difference between ΔG_i and ΔG_{i+1}, or equivalently the equilibrium constant association or dissociation of a single monomer. A few reports have been made of the PMF associated with removing a monomer from a micelle,[41, 43, 44, 103] with the potential defined as a function of the distance between the centers of mass (COM) of the monomer and the remaining components of the micelle. As discussed in the Methods section, to convert this into an equilibrium association constant with units of inverse concentration or volume requires integration of the Boltzmann weight over the volume of the (spherically averaged) micelle, treating the PMF for the surfactant in solution as zero. An additional factor of $1/(i+1)$ needs to be included to account for the indistinguishability of the surfactants. The initial equilibration time of 30 ns allotted to each window was not adequate, so was extended to 100 ns over the windows where the slope of the PMF was steepest.

Table 4.2 Comparison of $\Delta\Delta G_{(i+1)\rightarrow i}$ calculated from umbrella sampling and PEACH-BAR method (based on path “WS=0.9” to FS)

Number of DPC	$\Delta\Delta G_{PEACH-BAR}$ (kT) (tail-only)	$\Delta\Delta G_{PEACH-BAR}$ (kT) (new cluster def)	$\Delta\Delta G_{US}$(kT)
30->29	6.16	6.78	6.57
40->39	6.72	7.50	7.98

Comparison in Table 4.2 shows fair agreement between PEACH-BAR and US values calculated for the smaller clusters (i=29, i+1=30) but significant error (1.26 kT) for the larger clusters (i=39, i+1=40). Very roughly speaking, a discrepancy of this magnitude will lead to an difference in predicted CMC by a factor of $e^{1.26} = 3.5$, which is unacceptably large. In seeking explanations for this discrepancy, we noticed that the PMF curve does not reach a stable value until the DPC molecule is separated by a distance over 2.7 nm from the micelle center of mass (Figure 4.8(b)), and that its value increases by about 3 kT in the range between 2 and 2.5 nm. Observing the structure of the cluster formed when the tagged molecule was constrained in this range of distances, we saw that it clearly did not exhibit tail-tail interactions, but rather was oriented with its headgroup towards the micelle surface. (Strong headgroup-headgroup attractions, which we noted in our previous work on OPC,[97] may be an artifact of the force field.) According to our original cluster definition based on tail sites alone, the tagged molecule would not be classified as belonging to the cluster in calculating histograms for the PEACH-BAR calculations performed above. The contribution from headgroup-headgroup attractions to the affinity of a monomer to a cluster is, however, being counted in the US estimate for that quantity entered in Table 4.2. The original PEACH-BAR calculations therefore were missing a factor that contributes to the stability of aggregates formed.

To test whether excluding headgroup-headgroup attractions accounts for the discrepancy between PEACH-BAR and US results, we re-analyzed the same set of MD trajectories (for WS 0.85, WS 0.95, and FS force fields) to generate new cluster histograms using an all-site cluster criterion with the same 0.5 nm cutoff. The PEACH-BAR analysis was applied to the new histograms to generate new free energy curves. As expected, accounting for headgroup-headgroup contacts resulted in a lowering of the cluster free energies. PEACH fits to the histograms are shown in Fig. 4.9, and are as good or better than for the previous sets. Once again, a high level of consistency between PEACH-BAR free energy curves was obtained from WS 0.85 and WS 0.90 trajectories (Figure 4.10). Under the new cluster definition, PEACH-BAR and US results agree to within 0.5 kT for both cluster sizes (Table 4.2), which given the noise and uncertainties in both methods is reasonable. For the following sections, we will carry out analysis based on this more inclusive cluster definition, which is justified based on thermodynamics instead of on an assumption of what the important interactions should be.

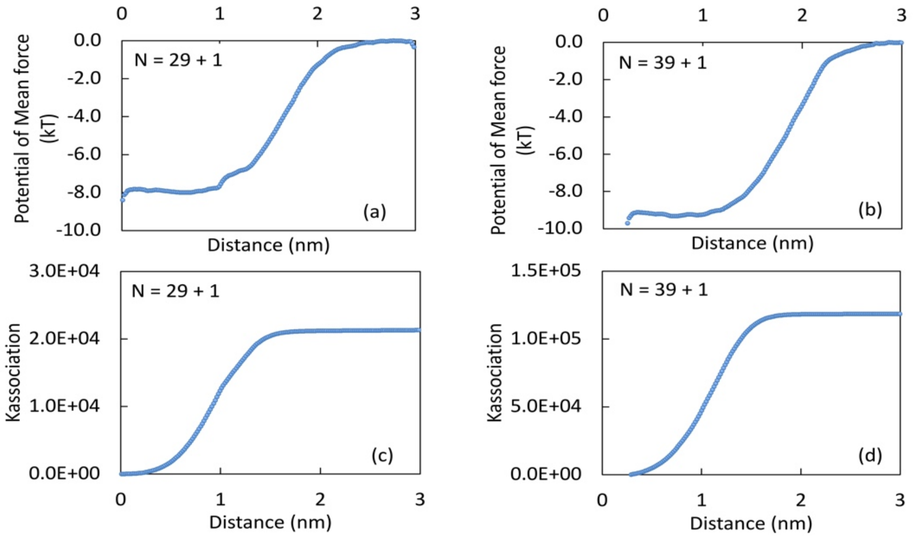


Figure 4.8 Umbrella sampling results. Potential of mean force (PMF) for pulling one DPC out of a micelle of (a) 30 DPC or (b) 40 DPC. Integral of PMF over micelle volume (c) for 30 DPC and (d) for 40 DPC

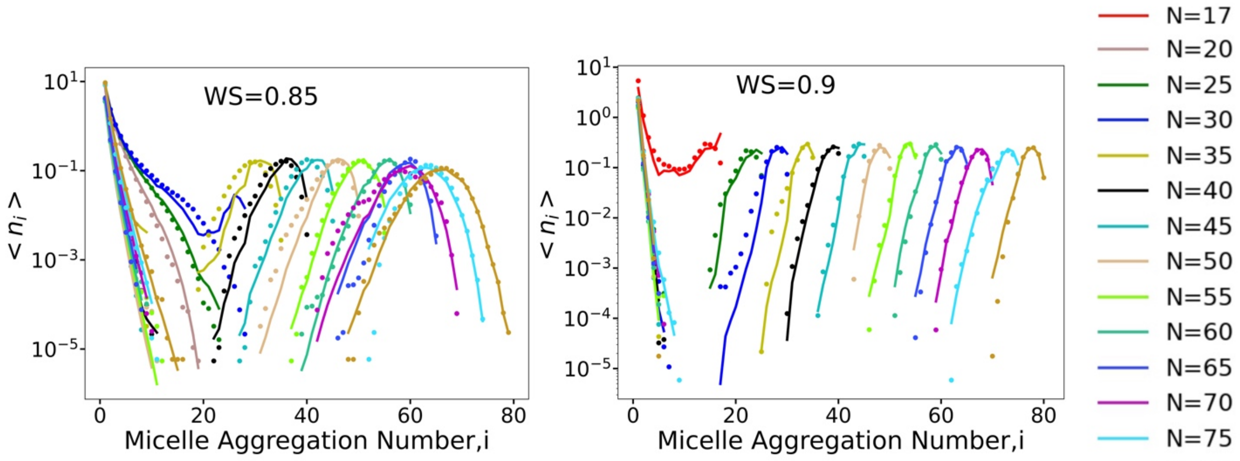


Figure 4.9 Micelle size distribution of simulations (symbols) and the PEACH fit (curves) of DPC for the new cluster definition, with force fields labelled on the figure and number N of DPC indicated by color legend.

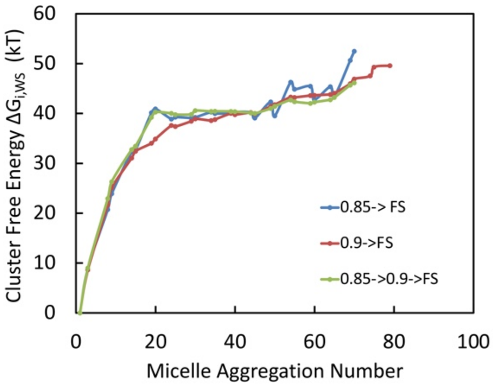


Figure 4.10 Cluster free energy $\Delta G_{i,FS}$ obtained from PEACH-BAR method. All the cluster free energy profiles were calculated for the condition of monomer concentration c=1.1 mM. The arrows showed how the full strength CFE is recovered from CFE of systems with weakened interactions based on the scheme shown in Fig.4.2.

We have not attempted to optimize either PEACH-BAR and US approaches for efficiency, but for a very rough comparison we note that the single point US calculation (yielding the slope of the free energy curve for a cluster at one specific size) required about 1.5 μs of simulation time; to cover the full range of cluster sizes would take roughly ten points. The full set of WS=0.9 →FS PEACH-BAR calculations required about 12 μs total trajectory time, suggesting that the overall computational costs of the two approaches are comparable. The US calculations have the advantage of being able to take advantage of trivial parallelization across umbrella sampling windows.

4.3.6 Phenomenological Modeling of the PEACH-BAR-derived Free Energy Profile

The free energy curves for DPC derived from PEACH-BAR method were fit to phenomenological models, developed to describe the dependence of aggregation free energy on aggregation number. Based on previous literature, the Maibaum/Chandler model[59] and the "quasi-droplet" model[86] were used to fit the CFE curves generated by PEACH-BAR method. Equation (9)[97] shows the general form of the two models.

$$\frac{\Delta G_i}{kT} = A(i^a - 1) - B(i^b - 1) + C(i^c - 1) \tag{9}$$

In the Maibaum/Chandler model[59], each term in the model corresponds to a different component of free energy. The first term Ai^a represents the surface free energy, as constant surface tension times surface area scaling with number of particles to the a = 2/3 power. The second term Bi^b represents the bulk free energy to transfer one surfactant from the solvent into the micelle, scaling linearly with number of surfactants (b=1 and B=$\Delta\mu$, the difference in chemical potential between solvated and micellar surfactant). The third term Ci^c represents the geometrical limitations of packing into a sphere while maintaining headgroups on the surface. The exponent for the third term is still undefined and in original Maibaum/Chandler model, this coefficient was proposed to be 5/3. However, we obtained better fit with c=2 for sodium octyl sulfate[71] and c=4 for octyl phophocholine[97] in previous publications. For the "quasi-droplet" model[86], the parameters were defined as a=1, b=3/2, c=2 were derived from a more complicated calculation that incorporates effects of protrusions of headgroups away from an interior hydrophobic core.

The cluster free energy generated from PEACH-BAR method (using the new cluster definition) was fitted to these two phenomenological models and shown in Fig.11 and Table 3. In contrast to our results on OPC,[97] where the quasi-droplet model gave a better fit, Fig.11 shows no clear advantage for either free energy model. This may reflect the more compact nature of the longer-tail surfactants' micelles, with headgroup protrusions playing less of a role. As shown in Table 3, fit to CFE generated from "WS= 0.9" gives a smaller root mean squared error (RMSE) than the ones from "WS= 0.9". This is consistent with our previous observation that CFE from "WS=0.9" is more reliable due to more overlap of configurations between full strength and "WS=0.9" systems.

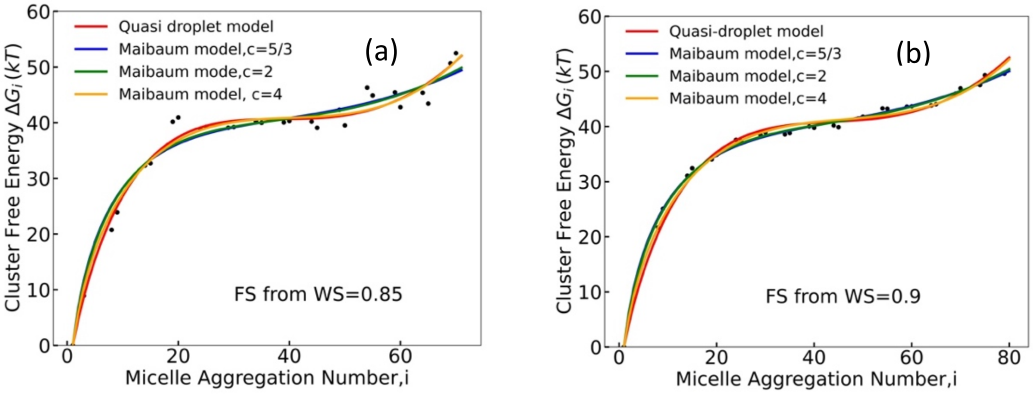


Figure 4.11 The fit of CFE to phenomenological models, with dots indicating the CFE calculated from PEACH-BAR method and lines indicating fitted models. CFE were generated from pathway (a) "WS=0.85" -> FS (b) "WS=0.9" -> FS. The CFE profiles were calculated for the condition of monomer concentration c=1.1 mM.

Table 4.3 Parameters for Maibaum/Chandler model and "quasi-droplet" model (for cluster free energy profile with monomer concentration c = 1.1 mM)

	"WS=0.85" → FS		"WS=0.9" → FS	
	Maibaum/Chandler c= 5/3	Quasi-droplet	Maibaum/Chandler c= 5/3	Quasi-droplet
A	22.2445	7.2496	19.3919	6.0879
B	6.4459	1.5614	5.3540	1.2216

C	0.1163	0.0947	0.0893	0.0695
RMSE	2.2850	1.7644	0.7433	1.0714

Using the fits from "WS=0.9", we could predict the cluster size distributions for all concentrations. The results are qualitatively independent of the fitting function used. To calculate average micelle size, we defined clusters with size larger than 7 to be a micelle. The following equations are used to calculate the average quantities, with $c_{\mathrm{i\_mer}}$ indicating the concentration for cluster of size i.

$$\%DPC\ in\ micelle = \frac{\sum_{i>7} c_{i\_mer} * i}{\sum_{i\geq 1} c_{i\_mer} * i} \tag{10}$$

$$Average\ Micelle\ Size = \frac{\sum_{i>7} c_{i\_mer} * i}{\sum_{i>7} c_{i\_mer}} \tag{11}$$

The aggregation number of DPC has been estimated through a number of experimental approaches. It was reported to be 51-62 from NMR diffusivity measurements[104], 60-80 using two different analyses of SAXS data[91, 92] and between 58 and 70, with a strong concentration dependence, by SANS data.[90] We found an average aggregation number near 75. Fig. 4.12 shows the PEACH-derived micelle statistics predicted for concentrations near experimental CMC and 100 times larger than CMC. No obvious difference was observed for the statistics of these two concentration ranges. The average micelle size is 72 or 75 for 10 mM and 73 or 76 for 100 mM.

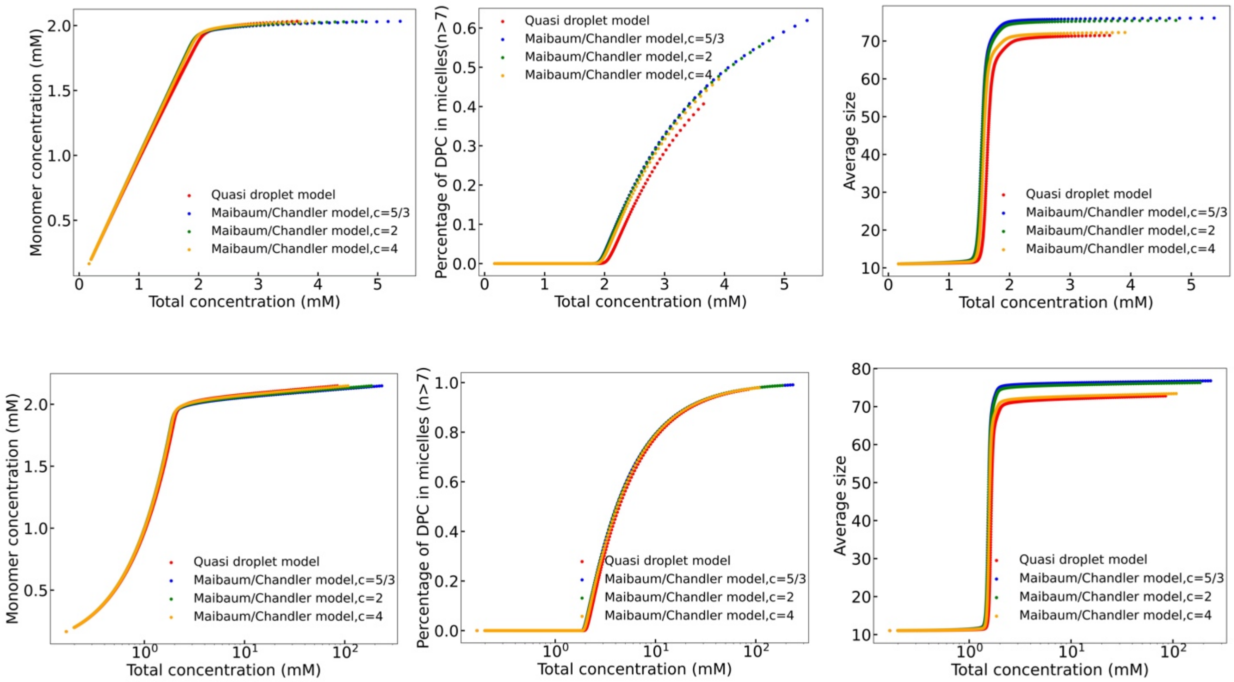


Figure 4.12 Micelle statistics (monomer concentration, DPC% in micelles and average size w.r.t total DPC concentrations) obtained from cluster size distributions predicted by equilibrium constants fit by phenomenological models for concentration range of 5 mM (upper panels) and 100 mM (lower panels)

Based on predictions of quasi-droplet model, CMC was found to be 2.1 mM. (As reported in previous paper[97], the cmc was defined based on the intersection between two lines fitted to a the trace of unimer concentration vs. total concentration, as shown in Fig. 4.13.) A few experimental results reported CMC and aggregation number based on different approaches. CMC was reported to be 1.5 mM by Anatrace, Inc. and other sources reported slightly different numbers. It was reported to be around 0.91 mM by tensiometry data[105] and 0.95 mM by fluorimetry[106] and 1.36 mM by calorimetry[107]. The over-prediction of the CMC relative to experiment is consistent with our results using the same force field for OPC[97], which was in excess by 40% of the value reported by Anatrace, Inc.

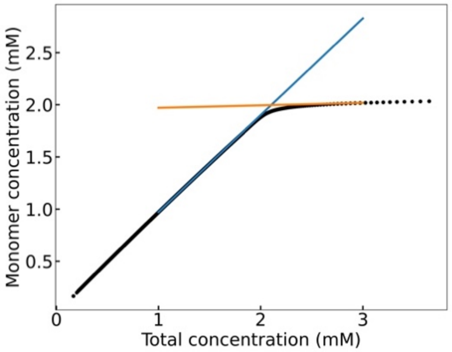


Figure 4.13 Monomer concentrations predicted by equilibrium constants (generated by quasi-droplet model). CMC was determined by finding the intersection points of the two lines on the graph.

4.3.7 Modeling SAXS data

Derivations of micelle aggregation numbers from experiment are based on various types of models, so it is desirable where possible to make direct comparisons between structures generated in simulation with experimental data. Therefore, we predicted the SAXS profile from simulation snapshots using the WAXSiS server[2] and compared to experimental SAXS data.[91, 108] The scattering intensity I(q) is defined as $I(q) = nP(q)S(q)$. P(q) is the averaged form factor of the globular particle and S(q) is the static structure factor describing the spatial distribution of the particles. P(q) is the Fourier transform of the pair distance distribution function (PDDF) p(r). Fig 4.14 shows the intensity and PDDF profiles for micelle sizes N=30,40,50,60. The peaks shifted to lower q (greater correlation distance r) for larger micelles. Lipfert et al. report scattering intensities for various concentrations of DPC (denoted FC-12); the experimental profile is qualitatively similar to ours except for the relative magnitudes of the main peaks at 0 and at 0.2 Å^{-1}; (more explanation about feature at q=0, origins of relative magnitude; or just refer to OPC paper) they observe a peak at 0.18 Å^{-1} and a minimum at 0.09 Å^{-1} slightly lower than in the predictions for N=60, suggesting a somewhat higher micelle size than 60. (As we concluded in a previous paper[97], the scattering profile for polydisperse mixtures of micelles in equilibrium is determined by the dominated species.) Göbl et at.[108] reported similar PDDF profiles for DPC measured by SAXS. In Göbl's paper, the peaks of p(r) shifted to higher distance r for longer alkyl chain. They find positive peaks at 1.0 1.1 and 4.0 4.1 nm, slightly greater than the predicted PDDF profile for micelle size N=60, again suggesting an aggregation number somewhat greater than 60.

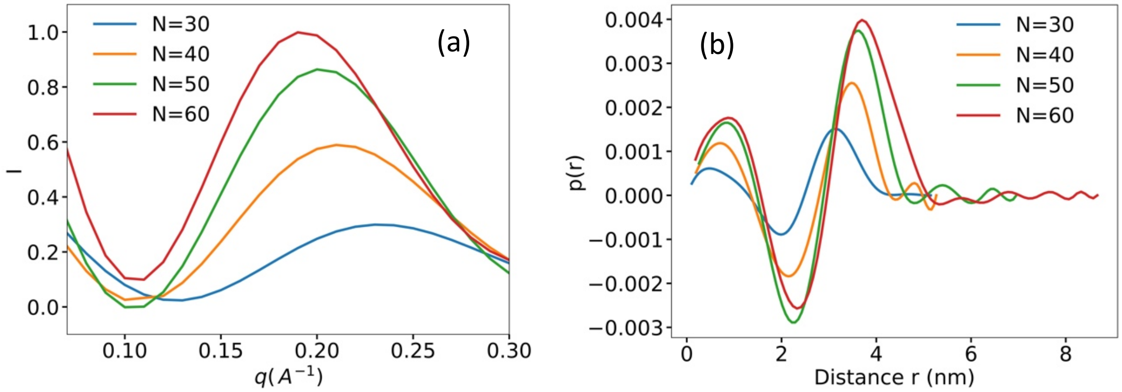


Figure 4.14 (a)SAXS profiles generated by WAXSiS server[2] for cluster size 30,40,50,60 (b) PDDF profiles calculated from the SAXS profile

4.4 Conclusions

In this work, we have presented a new method to calculate cluster free energy for amorphous aggregation with relatively low critical micelle concentration and rare dissociation events. The molecular dynamics simulations were performed for dodecyl phosphocholine with the strength of tailgroup interactions reduced, followed by PEACH analysis to obtain cluster free energy for the interaction weakened systems. The BAR method was used to calculate the difference between the full strength systems and interaction weakened systems. Comparison between results obtained using different degrees of interaction weakening showed internal consistency, and comparison with independent umbrella-sampling free energy calculations also showed reasonable agreement once an appropriate cluster definition was applied. The onset of micellization was predicted at 2 mM, somewhat higher than experimentally determined values as expected given previous results on octyl phosphocholine with the same force field. Average micelle aggregation number was predicted to be about 75, with weak dependence on concentration. Modeled SAXS profiles showed agreement with experiment in the positions of peaks but not in the relative intensities.

4.5 Acknowledgments

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Comet cluster at the San Diego Supercomputer Center, which is supported by National Science Foundation Grant No. ACI-1548562, Allocation No. TG-MCB110144. This work also used the resources of the Cherry L. Emerson Center for Scientific Computation.

Chapter 5. Study of Enthalpy Change for Micellization of Sodium Octanoate

5.1 Introduction

The reversible assembly of surfactants into micelles is a widely observed phenomenon driven by the hydrophobic effects of non-polar tails and hydrophilic effects of polar heads. The competing enthalpic and entropic effects define the process of micellization. However, microscopic insight into the details of what assemblies are present in an equilibrium solution of micelles is not generally available from experiment or simulation. These concentration-dependent distributions, along with insight into enthalpies of formation, are required to predict and interpret the heat of dilution of ionic surfactant solutions as measured using isothermal titration calorimetry (ITC). This chapter presents an application of the PEACH method to model the ITC enthalpograms of long-chain carboxylates (commonly known as soaps) sodium octanoate, based on simulations with atomistic detail. We are interested in how the microscopic properties affect the overall heat effect measured in careful experiments by Medos *et al*,[1] and testing the simplifying assumptions that the experimentalists used to analyze their data.

5.2 Methods

The simulations were performed with Gromacs 5.0 software package.[75] Throughout the simulations, the pressure was maintained at 1 bar with the compressibility of 4.5×10^{-5} by Berendsen barostat with τ_P=5 ps.[77] The temperature was maintained by velocity rescaling thermostat with τ_T=1 ps. We applied md integrator with 2 fs time step. Non-bonded interactions were truncated with Verlet scheme[78] using a cutoff of 1.2 nm. Coulomb interactions were accounted for by particle mesh Ewald summations with a real space cutoff of 1.4 nm.[55] The alkane-

alkane interactions and alkane-water interactions are accounted for by HH-Alkane model[70] (adapted from TrAPPE alkane forcefield[69]) working with TIP4P-2005 water. Parameters from OPLS force field were used to account for LJ parameters and partial charges of ions and carboxylate headgroup following Hess and van der Vegt.[53]

We performed two-component PEACH analysis on each set of simulations, which gives the set of equilibrium constants $\{K_{j,k}\}$ corresponding to each one of the cluster sizes {(j,k)} showing up in simulations at several temperatures. The implementation of two-component PEACH analysis was introduced in a previous publication.[71] Here for the first time we introduce a method to use PEACH to calculate enthalpies of formation in addition to free energies of formation. The $\{K_{j,k,T}\}$ obtained from PEACH analysis on each temperature were used to calculate $\{\Delta G_{j,k,T}\}$. We apply Gibbs-Helmholtz equation to equation (1) to obtain equation (2).

$$\frac{\Delta G_{j,k}\ (T_2)}{T_2} - \frac{\Delta G_{j,k}(T_1)}{T_1} = \Delta H_{j,k}(\frac{1}{T_2} - \frac{1}{T_1})$$

$$\Delta G_{j,k} = \ -RT\ln K_{j,k} \tag{1}$$

$$-\ln K_{j,k,T_2}\ + \ln K_{j,k,T_1} = \frac{\Delta H_{j,k}}{R}(\frac{1}{T_2} - \frac{1}{T_1}) \tag{2}$$

Thus, with the PEACH-derived equilibrium association constants at several temperatures, we performed linear regression for $-\ln K_{j,k}$ with respect to $\frac{1}{T}$ for each cluster size (j,k) and the intercept obtained is $\frac{\Delta H_{j,k}}{R}$, where $\Delta H_{j,k}$ is the enthalpy of association for that cluster to form from j surfactants and k counterions.

To generate the ITC curve, we reproduced the experimental procedure of adding stock solution stepwise. For each state of system after one injection of stock solution, the enthalpy (relative to a fully dissociated solution) is calculated with equation (3)

$$H = \ \sum_{j,k}\Delta H_{j,k} * conc_{j,k} \tag{3}$$

Given a certain total concentration of surfactants, we used a bisectional method to find the optimal set of concentrations for both surfactant and counterion unimers. And the $\{conc_{j,k}\}$ in above equation (3) was calculated based on equation (4)

$$K = \frac{conc_{j,k}}{conc_{j=1}^{j} * conc_{k=1}^{k}} \tag{4}$$

For the ITC curves compared to experiments, ΔH is calculated from equation (5). H_0 and H are the enthalpy for the system before and after one injection of stock solution, n_{stock} is the total amount of surfactant in one injection of stock solution.

$$\Delta H = \frac{q}{n_{stock}} = \frac{H - H_0 - H_{stock}}{n_{stock}} \tag{5}$$

Table 5.1 Simulation setup parameters

Temperature (K)	Number of NaOA	Number of Water
295	40	3200/3812/4500
300	40	3200/3812/4500
305	40	3200/3812/4500

5.3 Results

5.3.1 Simulation and PEACH analysis results

The simulations of NaOA micellization were performed at three temperatures 295K, 300K and 305K (tabulated in Table 5.1). Fig 5.1 shows a snapshot taken from the trajectory of one simulation setup. The cluster size distributions of each simulation conditions were fit with PEACH method to obtain cluster free energy. A comparison of the best PEACH-fit prediction of cluster size distributions with the observed simulation distributions is shown in Fig.5.2. Agreement is worst at the lowest concentration, where the model predicts a higher concentration of larger micelles than observed. This could indicate some non-ideal interaction effects among the clusters, or could be from incomplete equilibration in the most dilute system.

To compare the raw simulation data at different temperatures in a simpler format, the cluster size distribution of micelles in Fig.5.3 shows the total average numbers of micelles $<n_j>$ of the same number of surfactants j by adding up $<n_{j,k}>$ corresponding to all different number of counterions k. We observed a consistent tendency to form micelles of larger sizes as the temperature increases. Cluster free energy generated based on PEACH analysis is shown in Fig.5.4. As is consistent with the cluster size distributions, lower free energy surface was observed for higher temperature at the same concentration, indicating that micelles are easier to form.

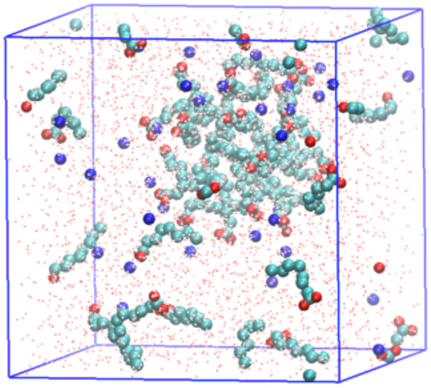


Fig 5.1 Simulation snapshot of 40 sodium octanoate (NaOA) in 3200 SPC water at 300K. Bead colors representations: Cyan-carbon, red-oxygen, blue-sodium ion.

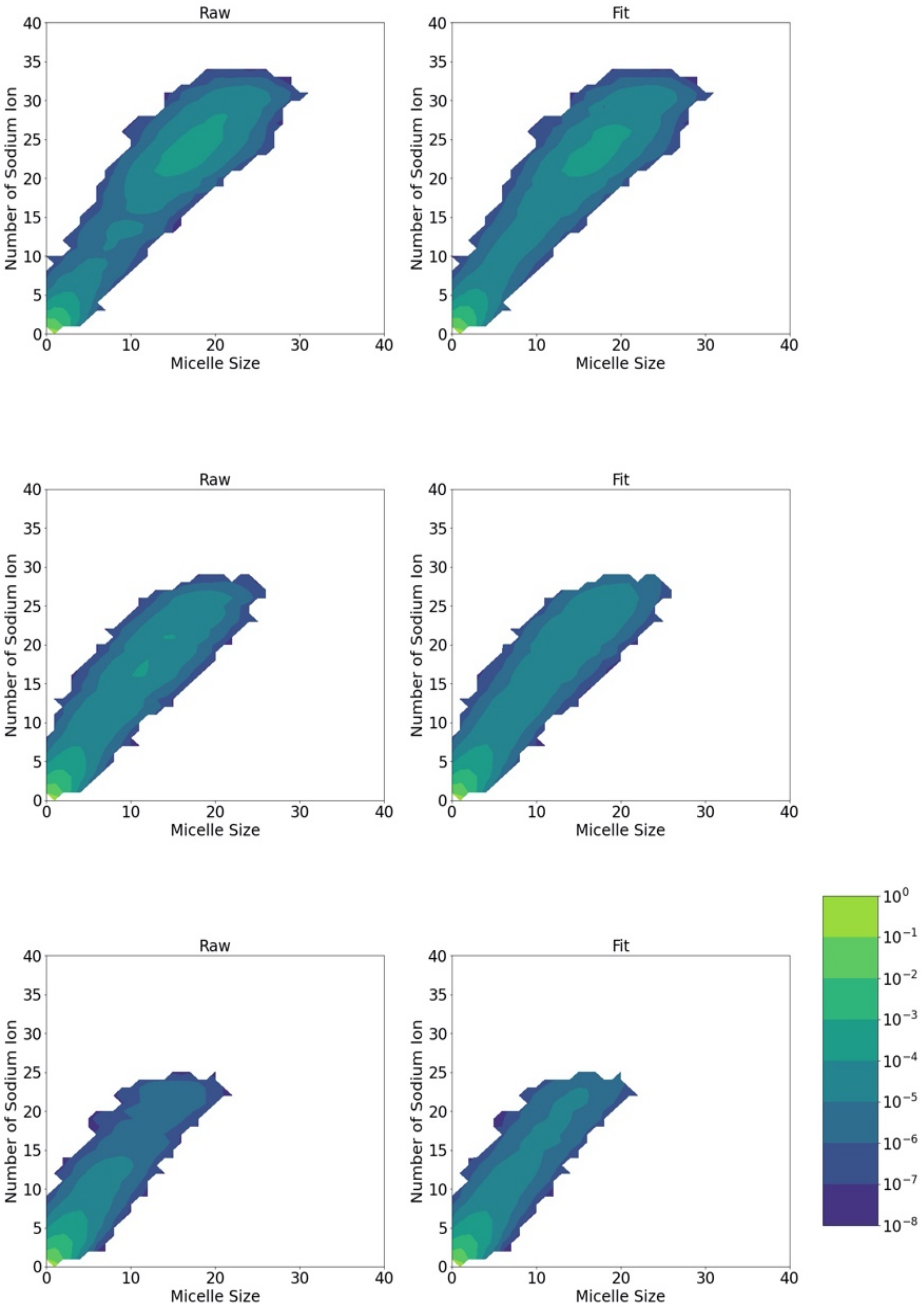


Fig 5.2 Cluster size distributions of 40 NaOA under simulation conditions (a) Number of Water = 3200 (b) Number of Water = 3812 (c) Number of Water = 4500 at 300 K. The left panel shows the simulation distributions, and the right panel shows the PEACH fit distributions.

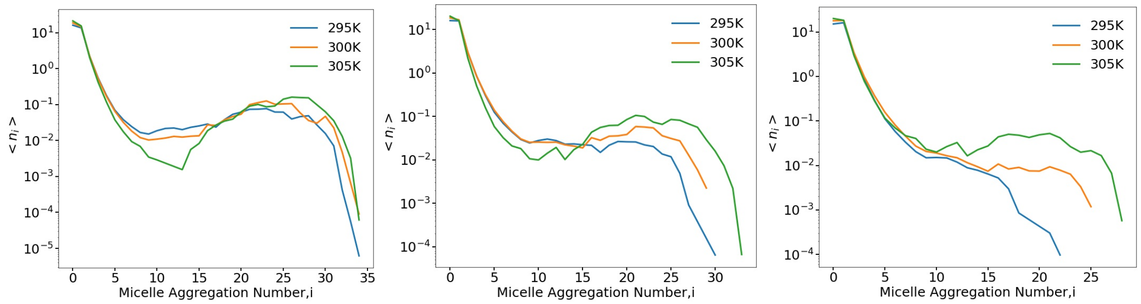


Fig 3. PEACH fit of 40 NaOA under simulation conditions (1) Number of Water = 3200 (2) Number of Water = 3812 (3) Number of Water = 4500.

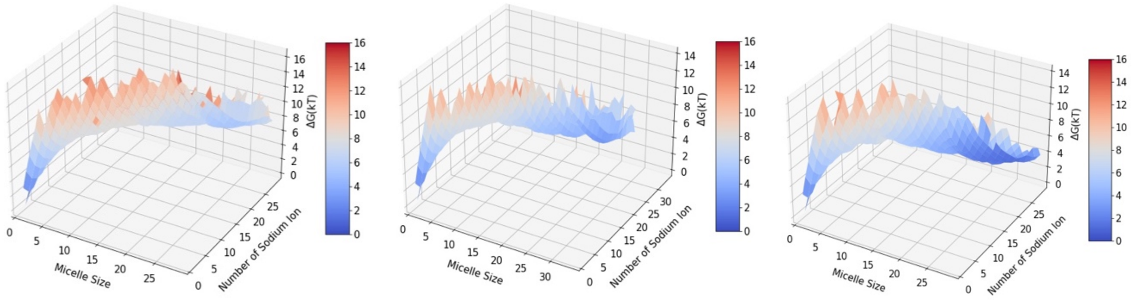


Fig 5.4. Cluster free energy from PEACH fit for 295 K (left), 300K (center), and 305 K (right) for a concentration of $c_{1,0} = 0.1, c_{0,1} = 0.155$.

5.3.2 Micelle Statistics

Based on the cluster free energy, we can calculate the micelle statistics such as monomer concentrations, average micelle size, neutralization percentage and other properties corresponding to the range of concentrations in experiments. These are shown in Fig. 5.5. The critical micelle concentration predicted is in the range of 0.30 to 0.37 mol/L. The cmc derived from the parameters of the two-step model[1], is in the range of 0.33 ~ 0.42 mol/L for a temperature from 328K to 288K, however, the definition of the cmc is slightly different from the conventional definition we use in PEACH analysis. In the two-step model, cmc is defined as the flex point of the plot $\Delta n_{M,2}/\Delta n$, whereas in PEACH analysis, cmc is defined by the maximu of the plot of number of free monomers. From the cmc to or the concentration range we predicted (up to 0.4 mol/L), the average micelle size is in the range of 23~25, significantly greater than the 11±3 derived from the two-step model fit to ITC data; the percentage of neutralization is ~0.7, compared with 0.8±0.1 from the two-step model.[1]

5.3.3 Discussions

As indicated by the temperature trend we discussed before, we observed a lower cmc and higher fraction of micelles for a higher temperature while the percentage of neutralization is mostly not influenced by temperature. In experiments, a negative heat capacity was observed, and therefore, as temperature increases, the onset micellization decreases but the decrease slows down. This is contrary to what we predicted for the onset of fraction of surfactant in micelles based on PEACH-predicted cluster free energy. Our observations, together with the predictions on monomer concentrations, indicated a positive heat capacity.

This discrepancy led us to reconsider the setup of simulations and we assume two major reasons could cause inaccuracies. One possible reason is the imbalanced sampling in trajectory and between several trajectories for one temperature. We obtained 1.2 microseconds of simulations for most trajectories; and based on previous studies, we assume that this trajectory length is enough for us to obtain reliable cluster free energy profiles. We obtained comparable Gibbs free energy to two-step model and this ascertained our assumption about the trajectory length. However, obtaining the correct heat capacity requires a higher accuracy on Gibbs free energy in order to obtain the accurate second derivative of Gibbs free energy. The obstacle to this is that the ratio of number of the associated and dissociated micelles is affected by ratio of the time spent on each configuration and the trajectory we obtained might not represent the most accurate ratio due to the limitation of slow dissociation and association. We discussed about applying Bayesian optimization method to solve the problem of imbalanced sampling between trajectories, however, the sampling in trajectory cannot be optimized. The second possible reason is that the force field does not represent the interactions accurately. Force fields are usually optimized near 300K and less optimized for other temperatures. In our case, we observed that the force field we applied was capable to follow the temperature trend of cluster free energy and approximate the values; but it did not reproduce the heat capacity.

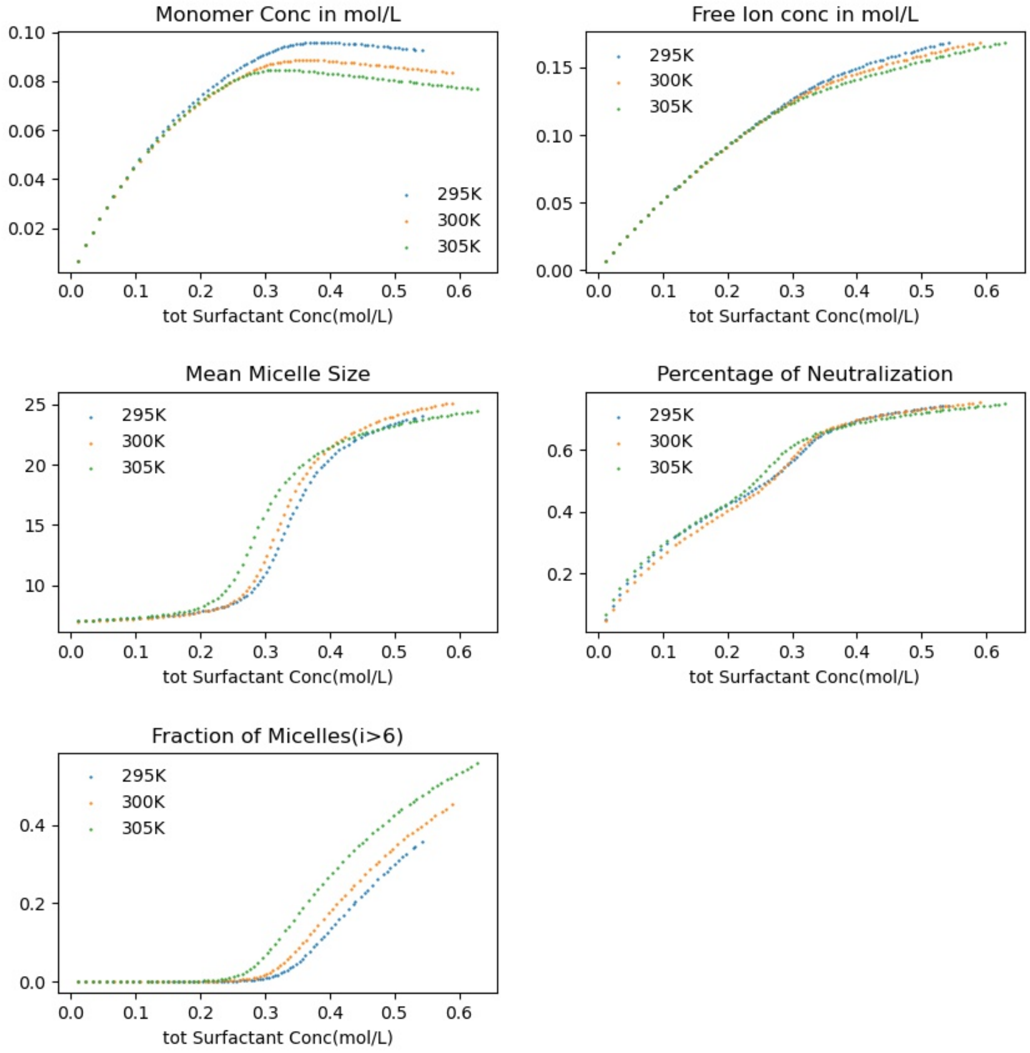


Fig 5.5 Micelle statistics predicted from PEACH analysis.

5.4 Predictions on Enthalpy changes

To obtain the enthalpy change during micellization process, we tried several different strategies. One big obstacle is that in this process, all cluster sizes with certain number of OA^- and Na^+ were taken into considerations and an accurate prediction of equilibrium constant is required for each one of the possible cluster sizes. The subtle change of prediction of equilibrium constants could lead to inaccurate prediction of enthalpy changes. However, using the united atom simulations requires long simulation trajectory to reach a relatively reliable cluster size distributions, even if our PEACH method allows for a much smaller sample size than conventional canonical simulations.

As the change of $\Delta H_{j,k}$ from each temperature is subtle, we first assume no temperature change on $\Delta H_{j,k}$. Applying the Van 't Hoff equations, we obtained $\frac{\Delta H_{j,k}}{R}$ based on equation (2) and calculate enthalpogram from there. Fig.5.6 (a) shows the R^2 of linear regression fitting of three temperatures. As we can see, the fitting is better for large cluster sizes indicating a more linear relation for large cluster sizes. This could mean a lower heat capacity or a reduction in the relative amount of noise affecting the linearity of the plot.

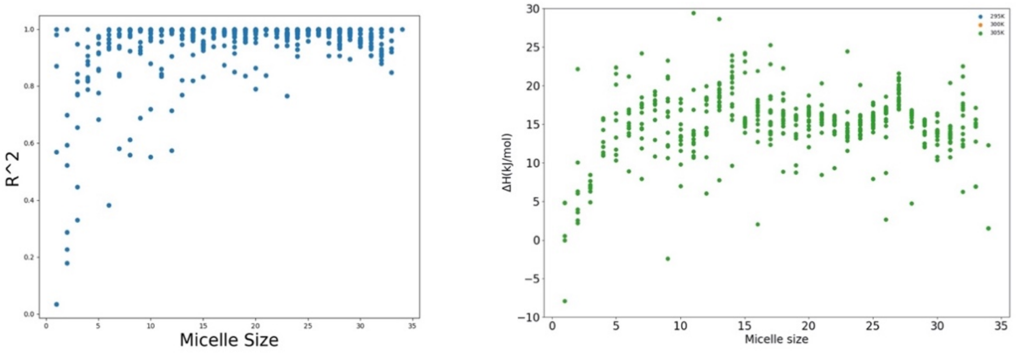


Fig 5.6 (a) R^2 of linear regression fitting of three temperatures (b) The enthalpy change per monomer for each micelle size

The enthalpy changes increases follow the increase of cluster size. And as we look at the enthalpy change per surfactant, we found that for cluster size above 5, the average enthalpy change per monomer upon association is between +15 kJ/mol and +20 kJ/mol. There is a relatively small scatter depending on the number of counterions, suggesting that the transfer of an ion from solution to the micelle surface is not strongly exo- or endo-thermic.

5.4 Predictions on enthalpogram

Based on the enthalpy change for each micelle size we obtained from linear regression fitting, we can mimic the experimental setup to generate the ITC curves. This allows for more direct comparison between the simulation model and experiment than comparing to quantities inferred from the two-stage model fit. We assume the system to be pure water at the beginning and then we treat the system as a new state every time we add the stock solution. For each new state, we calculate the micelle size distributions and the total enthalpy change as shown in Fig. 7(a). To obtain the curve which can be compared with experimental ITC curve, we use eqn. (5) to know the difference for adding stock solution. Fig 7 (c) shows the enthalpogram for three temperatures 295K, 300K, 305K. Similar to experimental enthalpograms, we predicted the same shape and inflection point at 0.3 mol/L for 300K. We also observed the same range of enthalpy change from -12 kJ/mol to 1 kJ/mol under the same concentration. One limit of this prediction is that, contrary to experiments, we observed that the heat capacity is close to zero for large micelle sizes and a relatively large heat capacity for small micelle sizes. This leads to a different shape for enthalpograms of 295K and 395K; in experiment, there is a noticeable and consistent decrease in the step feature at the cmc as temperature increases.

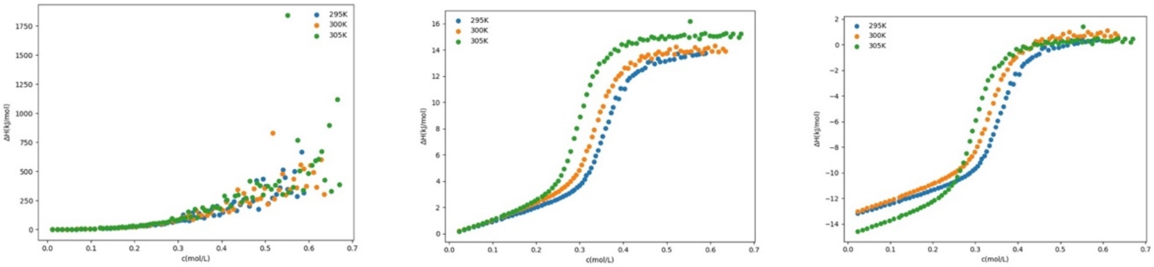


Fig 5.7 (a) Total enthalpy change corresponding to each concentration after adding stock solution (b) ITC curve before deduction of stock solution enthalpy (c) ITC curve after deduction of stock solution enthalpy

5.4 Conclusions

In this chapter, we developed a method to calculate enthalpy change of cluster formation for all cluster sizes shown in simulations based on the cluster free energy for micelles of ionic surfactants. Enthalpograms was generated to compare with experimental ITC curves. Due to the limit of sampling and force field, we did not obtain a good prediction on heat capacity – the change in ΔH with different temperatures. However, the predictions on Gibbs free energy and enthalpy change at 300 K are quite consistent with experiment and the two-step model. This marks the first attempt to model ITC curves for micelle simulation directly from simulation, and points to the strengths and weaknesses of the force field in dealing with the subtle thermodynamics of the hydrophobic effect.

Appendix I. Bayesian optimization method

This section introduces an optimization method which searches for the optimal set of equilibrium constants by finding the set with maximal probability of observing the simulation-generated size distribution. This is a high dimensional optimization, which involves ~2000 dimensions for two-component micelles.

For one temperature, we obtained one trajectory for each simulation box. (a histogram of all cluster sizes shown up in the simulation) The probability of observing one set $\{n_i\}$ of cluster compositions for a system of total number of N in a single, independent observation is shown below.

$$P(\{n_i\}) = (\frac{1}{Q(N)}\prod_i \left(\frac{q_i}{n!}\right)^{<n_i>_{sim}})^{S}$$

S indicates the number of independent sampling, which is unknown and not included in optimization for now. For now, we treat S as 1.

The probabilities that a system with cluster partition functions qi would yield a given sequence of λ independently sampled compositions is therefore

$$P(\{n_i\}_1, \{n_i\}_2, \dots \{n_i\}_\lambda) = \prod\nolimits_{traj\_m}^{traj\_M}\left(\frac{1}{Q(N)}\prod_i \left(\frac{q_i}{n!}\right)^{<n_i>_{sim,traj\_m}}\right)$$

To avoid overflow of K of large numbers, we take the logarithm for the probability

$$\ln P = \sum_m \left(-\ln Q(N) + \sum_i n_{i,m}\left(\ln q_i - \ln n_{i,m}!\right)\right)$$

This can be simplified as

$$\ln P = M\left(-\ln Q(N) + \sum_i \langle n_i \rangle \ln q_i\right) - \sum_m \sum_i n_{i,m} \ln n_{i,m}!$$

where we note that the last term does not depend on the q_i that are being sought.

$$\mathrm{argmax}_{\{q_i\}}\,(\ln P) = \mathrm{argmax}_{\{q_i\}} \sum_{traj\_m} \left(-\ln Q(N) + \sum_i n_{i,m} \big(\ln q_i - \ln n_{i,m}! \big) \right)$$

$$\mathrm{argmax}_{\{q_i\}}\,(\ln P) = \mathrm{argmax}_{\{q_i\}} \left\{ M \left(-\ln Q(N) + \sum_i \langle n_i \rangle \ln q_i \right) - \sum_m \sum_i n_{i,m} \ln n_{i,m}! \right\}$$

Next, we took partial derivatives for lnP with respect to each partition function q_i: with the following relation in mind:

$$\frac{\partial lnP}{\partial lnq_i} = -M \left(\frac{\partial \ln Q}{\partial \ln q_i} - < n_i > \right)$$

Then, applying the relationship

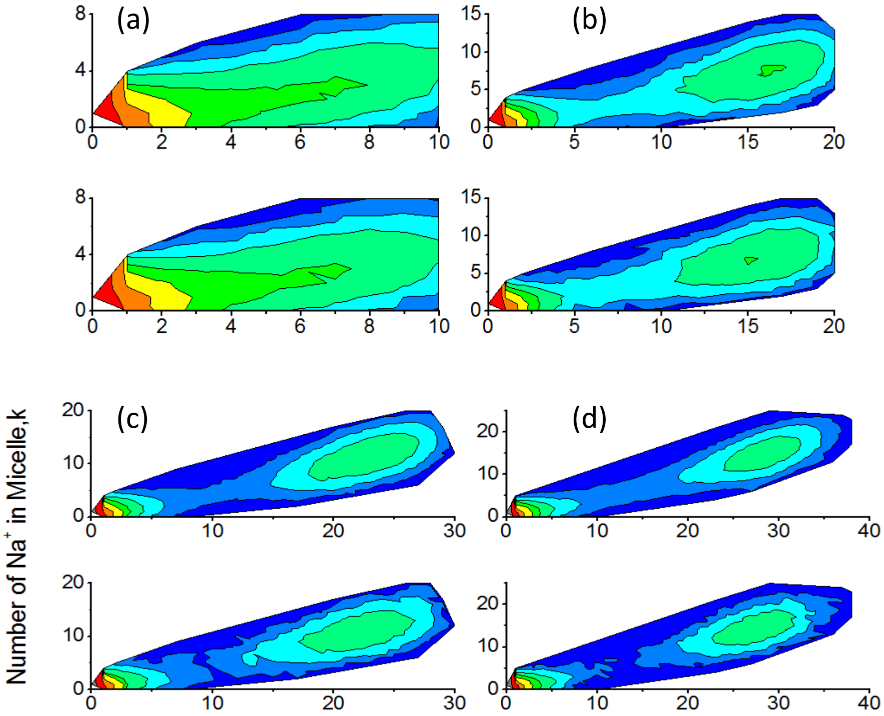
$$\frac{\partial lnQ}{\partial \ln q_i} = < n_i >_{model}$$

where $<n_i>_{model}$ is the ensemble average prediction of numbers of *i*-mers for the set of q_i. For clarity, we now will use $<n_i>_{sim}$ to refer to the observed mean number of *i*-mers over the simulation trajectory.

$$\nabla lnP = \begin{pmatrix} \sum_{traj_m}^{M} (< n_1 >_{sim,traj_m} - < n_1 >_{model}) \\ \cdots \\ \sum_{traj_m}^{M} < n_2 >_{sim,traj_m} - < n_2 >_{model} \end{pmatrix}_{(N+1)^2 \times 1}$$

Appendix II. Supporting Information for Chapter 2

Figure S1. Contour plots for the equilibrium micelle size distribution with respect to surfactant number and sodium ion number for the concentration c= 67.8 mM. (trajectories 1-6 from Table I). Plots are for different numbers of SOS molecules in the simulation box, with (a) SOS=10, (b) SOS =20, (c) SOS =30, (d) SOS=40, (e) SOS=50, (f) SOS=60. The upper panel of each graph is for the fitted distribution while the lower panel is for the simulated distribution. Legend scale represents contours of constant <$m_{j,k}$>.



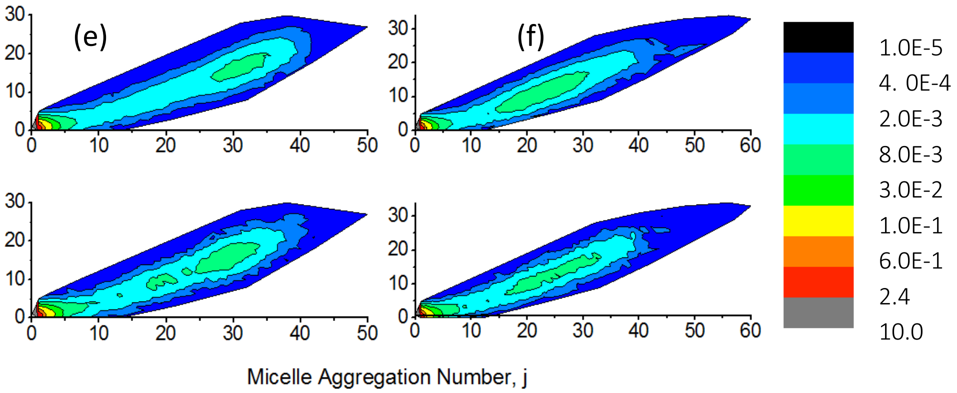
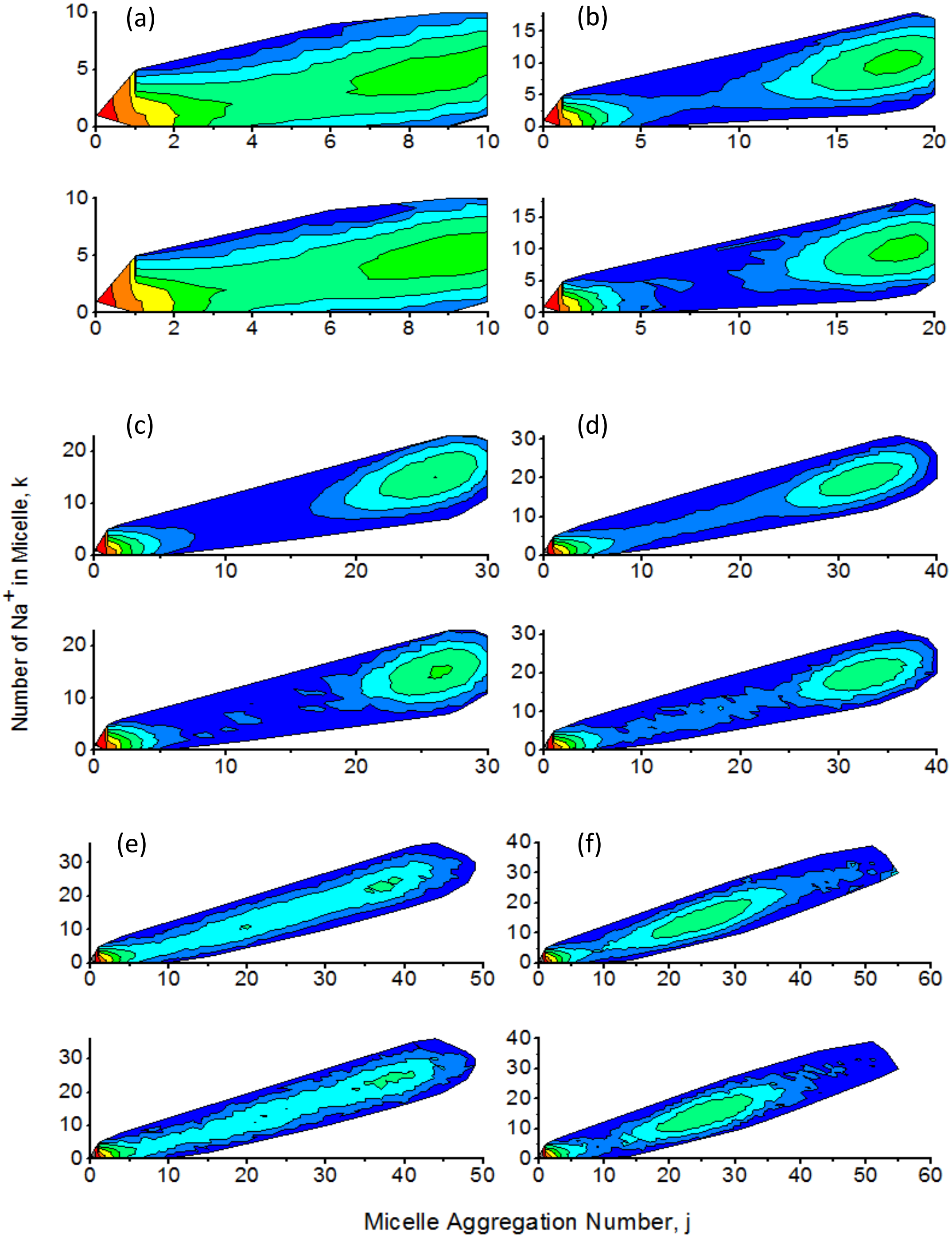
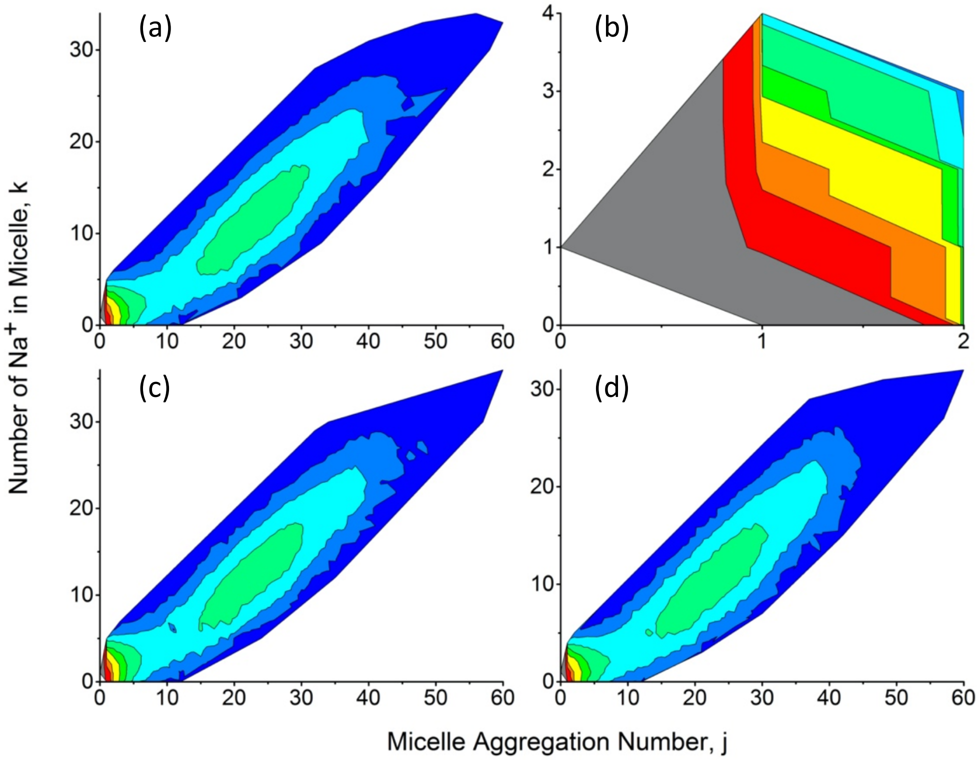


Figure S2. Contour plots for the equilibrium micelle size distribution with respect to surfactant number and sodium ion number for the concentration c= 114.9 mM (trajectories 7-12 from Table I). Plots are for different numbers of SOS molecules in the simulation box, with (a) SOS=10, (b) SOS =20, (c) SOS =30, (d) SOS=40, (e) SOS=50, (f) SOS=60. The upper panel of each graph is for the fitted distribution while the lower panel is for the simulated distribution. See Figure S1 for legend scale.



Cluster size distributions calculated using different cutoff distances

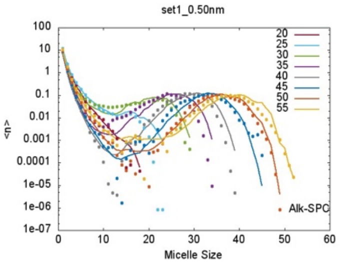
Figure S3. Cluster size distribution for N=60 with the concentration c=67.8 mM. The four panels are for different cutoffs. (a)0.85/1.5 nm (b)0.45/1.5 nm (c) 0.65/1.7 nm (d)0.65/1.3 nm, with the first distance the cutoff used in defining tail-tail association and the second for defining headgroup-counterion association. See Figure S1 for legend scale.

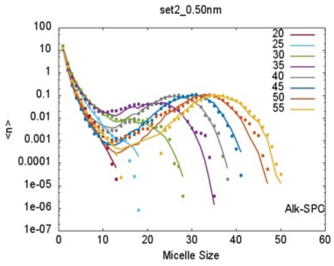


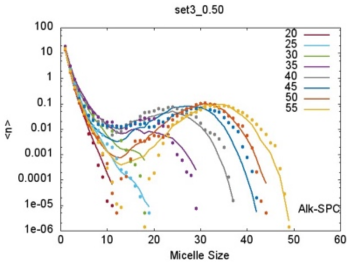
Appendix III. Supporting Information for Chapter 3

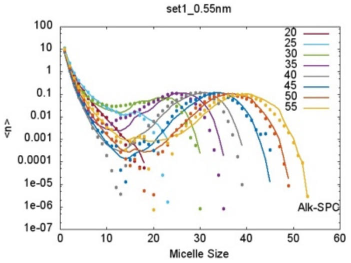
Fig S1. Cluster size distribution for simulations (symbols) and the PEACH fit (curves) for cutoffs 0.50, 0.55, 0.60, 0.65 nm; set number refers to compositions in Table I.

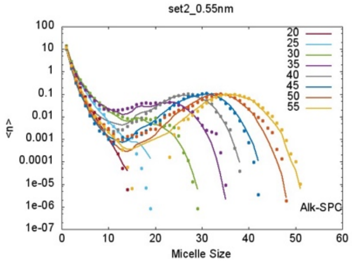
Alk-SPC

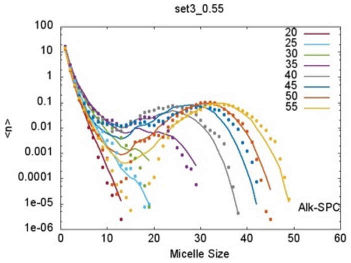

set1_0.50nm
20
25
30
35
40
45
50
55
Alk-SPC
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
<n>
0
10
20
30
40
50
60
Micelle Size

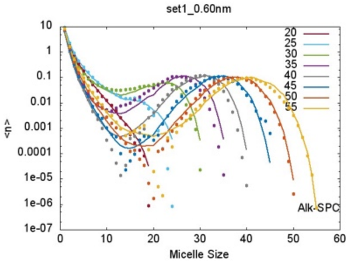

set2_0.50nm
20
25
30
35
40
45
50
55
Alk-SPC
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
<n>
0
10
20
30
40
50
60
Micelle Size

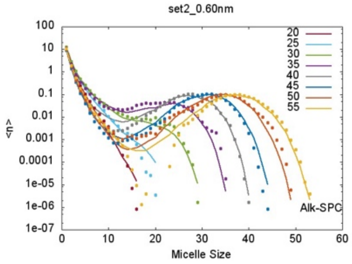

set3_0.50
20
25
30
35
40
45
50
55
Alk-SPC
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
<n>
0
10
20
30
40
50
60
Micelle Size

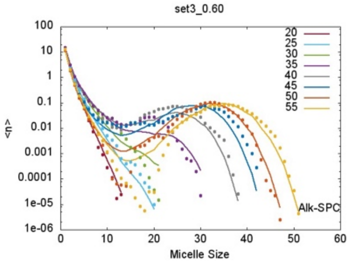

set1_0.55nm
20
25
30
35
40
45
50
55
Alk-SPC
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
<n>
0
10
20
30
40
50
60
Micelle Size

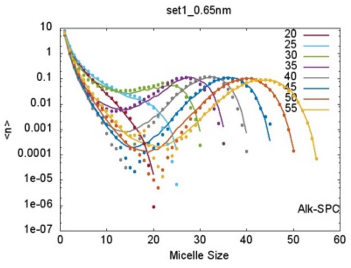

set2_0.55nm
20
25
30
35
40
45
50
55
Alk-SPC
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
<n>
0
10
20
30
40
50
60
Micelle Size

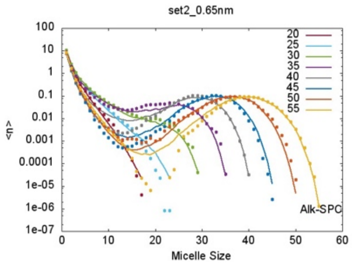

set3_0.55
20
25
30
35
40
45
50
55
Alk-SPC
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
<n>
0
10
20
30
40
50
60
Micelle Size

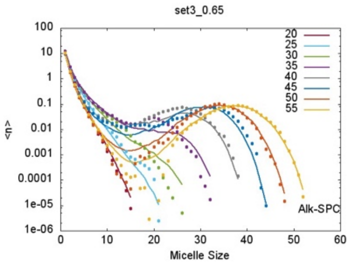

set1_0.60nm
20
25
30
35
40
45
50
55
Alk-SPC
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
<n>
0
10
20
30
40
50
60
Micelle Size


set2_0.60nm
20
25
30
35
40
45
50
55
Alk-SPC
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
<n>
0
10
20
30
40
50
60
Micelle Size


set3_0.60
20
25
30
35
40
45
50
55
Alk-SPC
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
<n>
0
10
20
30
40
50
60
Micelle Size

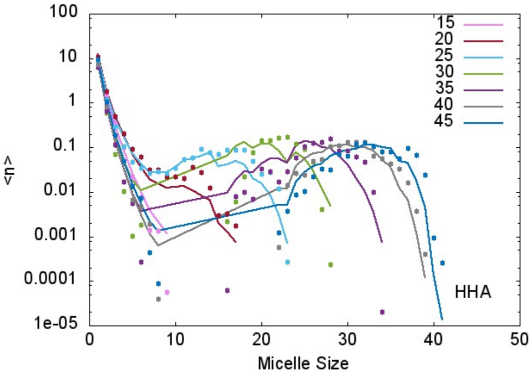

set1_0.65nm
20
25
30
35
40
45
50
55
Alk-SPC
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
<n>
0
10
20
30
40
50
60
Micelle Size


set2_0.65nm
20
25
30
35
40
45
50
55
Alk-SPC
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
<n>
0
10
20
30
40
50
60
Micelle Size

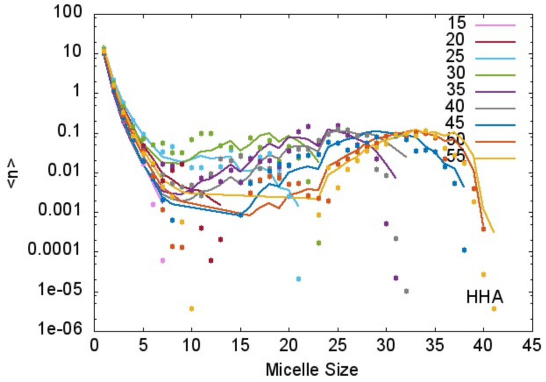

set3_0.65
20
25
30
35
40
45
50
55
Alk-SPC
100
10
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
<n>
0
10
20
30
40
50
60
Micelle Size

HHA

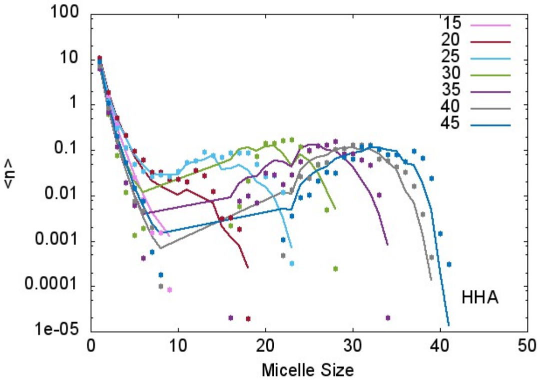
set1_0.50nm



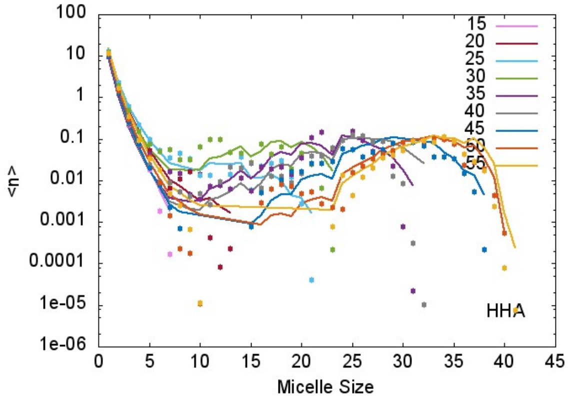
set2_0.50nm



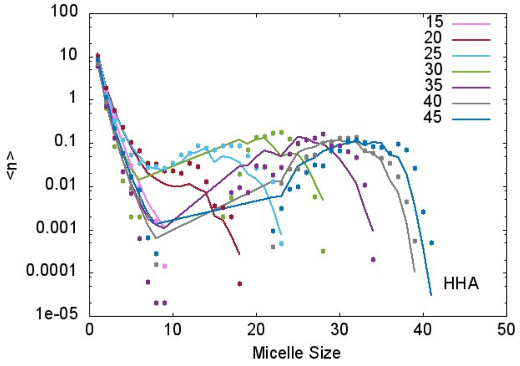
set1_0.55nm



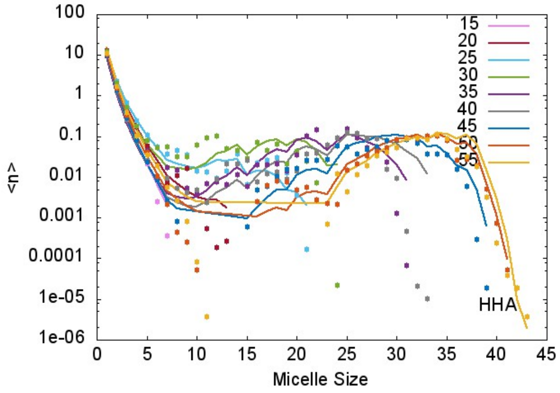
set2_0.55nm



set1_0.60nm



set2_0.60nm



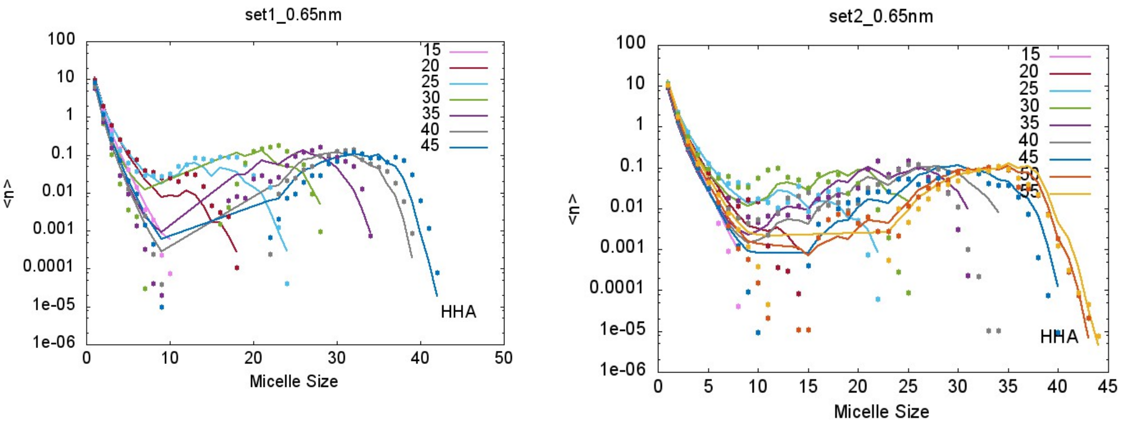


Figure S2: Comparison of dynamics of fluctuation of largest micelle over first 200 ns of simulation for Alk-SPC force field (red) and HHA force field (black). For both force fields, system is composed of 40 OPC and 3812 waters, and cluster sizes are plotted at 10 ps intervals.

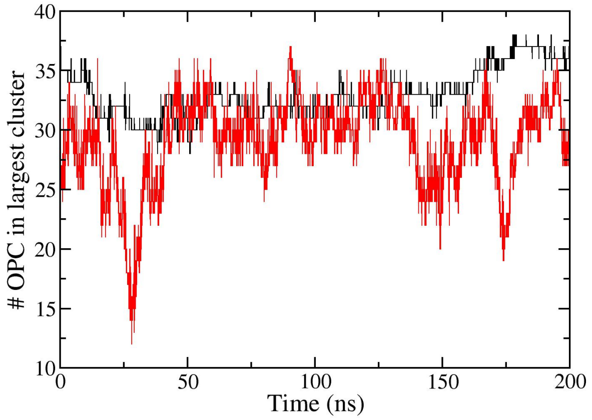
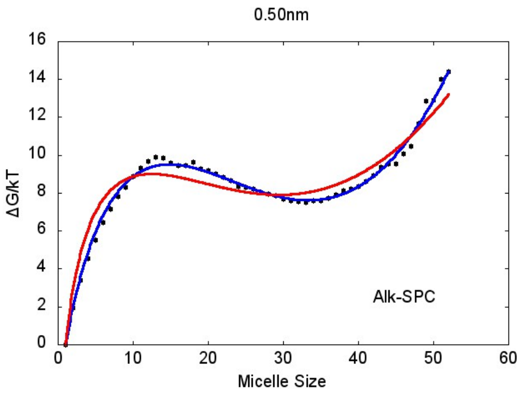
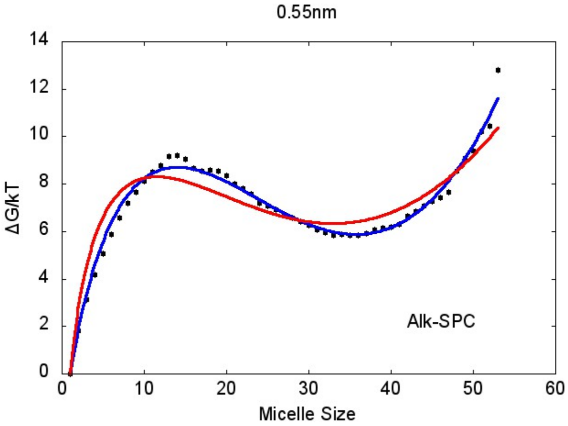
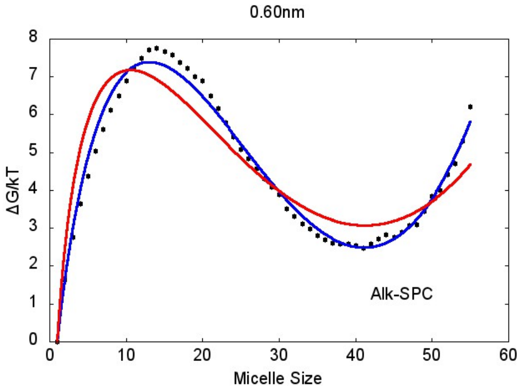


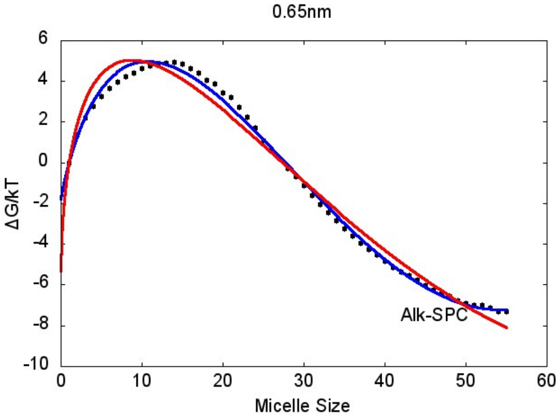
Figure S3. Fit cluster free energy curve (symbols) to two models (a) OPC simulated with Alk-SPC ($c_{monomer}$=116 mM)(b) OPC simulated with HHA ($c_{monomer}$=116 mM), with red curve for Maibaum-Chandler model and blue curve for “quasi-droplet” model.

Alk-SPC

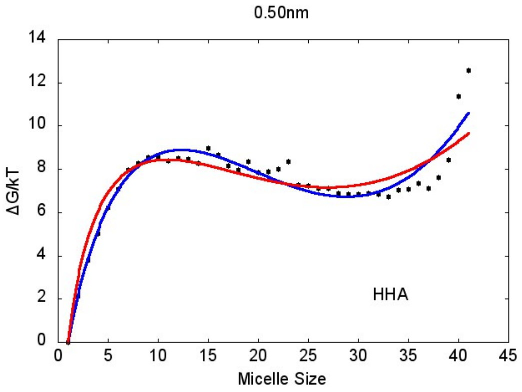

0.50nm
ΔG/kT
0
2
4
6
8
10
12
14
16
Micelle Size
0
10
20
30
40
50
60
Alk-SPC

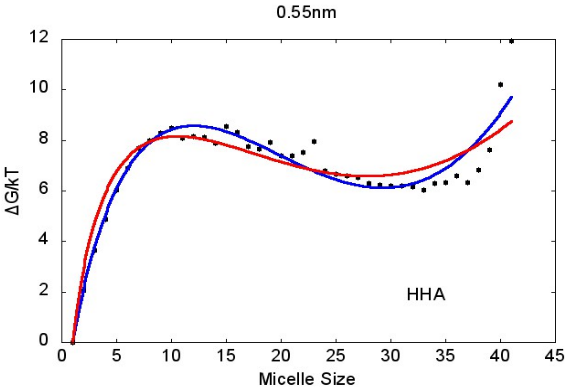

0.55nm
ΔG/kT
0
2
4
6
8
10
12
14
Micelle Size
0
10
20
30
40
50
60
Alk-SPC

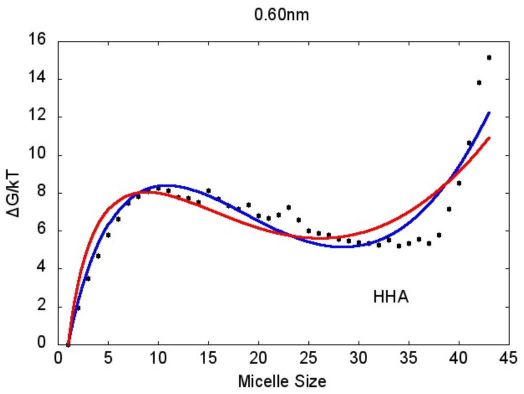

0.60nm
ΔG/kT
0
1
2
3
4
5
6
7
8
Micelle Size
0
10
20
30
40
50
60
Alk-SPC

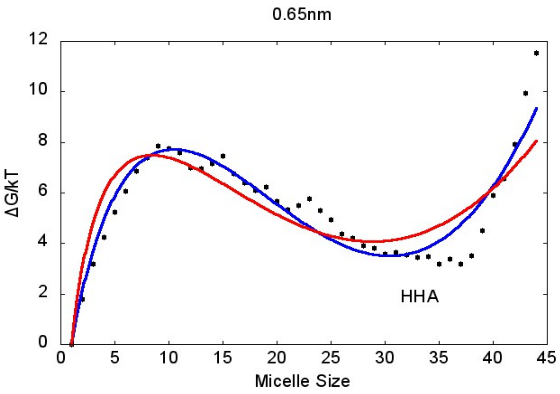

0.65nm
ΔG/kT
-10
-8
-6
-4
-2
0
2
4
6
Micelle Size
0
10
20
30
40
50
60
Alk-SPC

HHA


0.50nm
ΔG/kT
0
2
4
6
8
10
12
14
Micelle Size
0
5
10
15
20
25
30
35
40
45
HHA


0.55nm
ΔG/kT
0
2
4
6
8
10
12
Micelle Size
0
5
10
15
20
25
30
35
40
45
HHA





$$\frac{\Delta G(i)}{kT} = A(i^a - 1) - B \times \left(i^b - 1\right) + C \times (i^c - 1)$$

a=1, b=3/2, and c=2

Table S1: Best-fit parameters for quasi-droplet model as applied to PEACH-derived cluster free energies with different force fields and cutoffs r_{cut}.

Cutoff/nm	Force Field	Parameters			
		A	B	C	RMS
0.50	Alk-SPC	3.7197	1.07384	0.0837218	0.236877
	HHA	4.33884	1.36368	0.115297	0.603107
0.55	Alk-SPC	3.51592	1.01825	0.0785707	0.290286
	HHA	4.26495	1.34579	0.11373	0.615484

0.60	Alk-SPC	3.09374	0.892887	0.0668185	0.249387
	HHA	4.68232	1.53398	0.133444	0.993096
0.65	Alk-SPC	2.42344	0.71269	0.0502369	0.260727
	HHA	4.37147	1.42875	0.122445	0.738317

Figure S4: Snapshot of 35 OPC/3812 SPC system showing tails sites in yellow, headgroup sites in elemental colors (cyan=CH_x, red=O, brown=P, blue=N), and "neighbor bonds" connecting pairs of tail sites closer less than r_{cut}=0.65 nm in black. Purple sites mark the tail of an OPC that is exterior to a large cluster but connected via a "neighbor bond". Solvents near the two neighboring sites are shown as thin sticks.

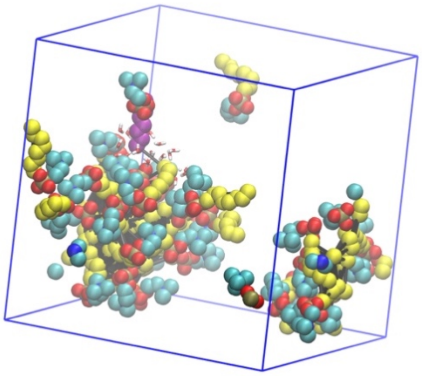
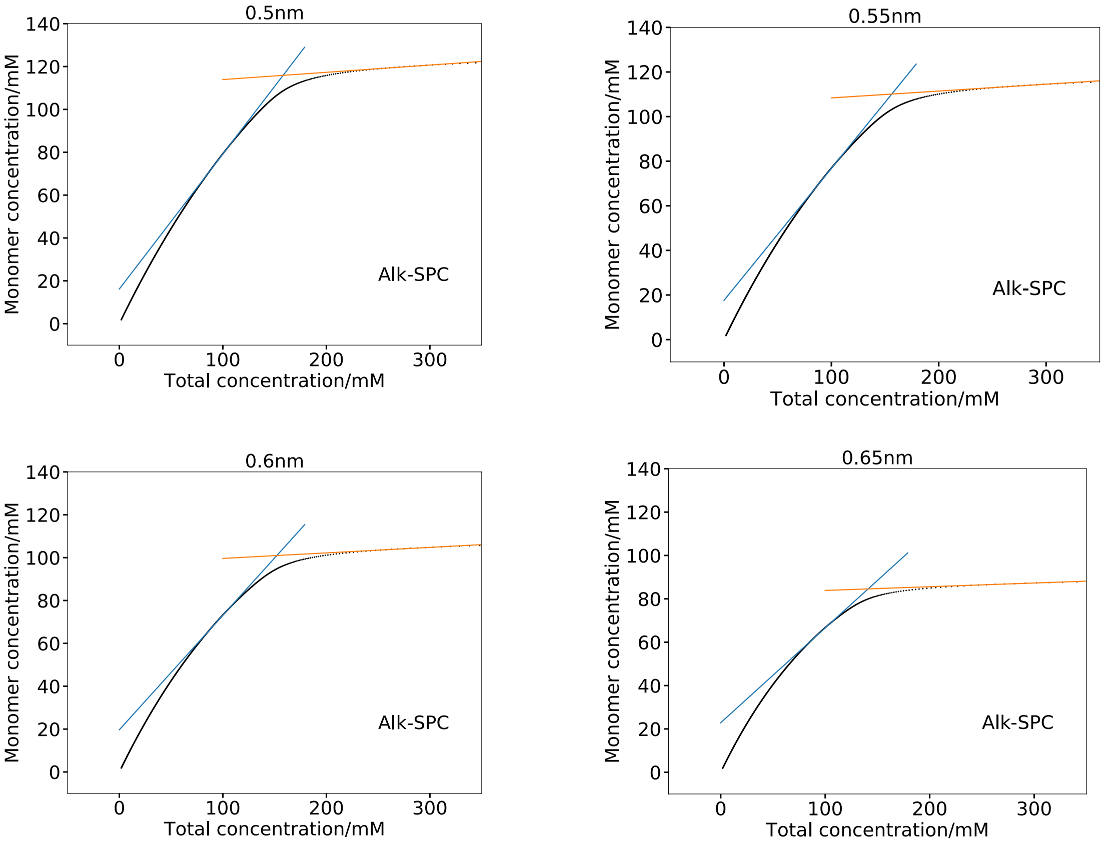
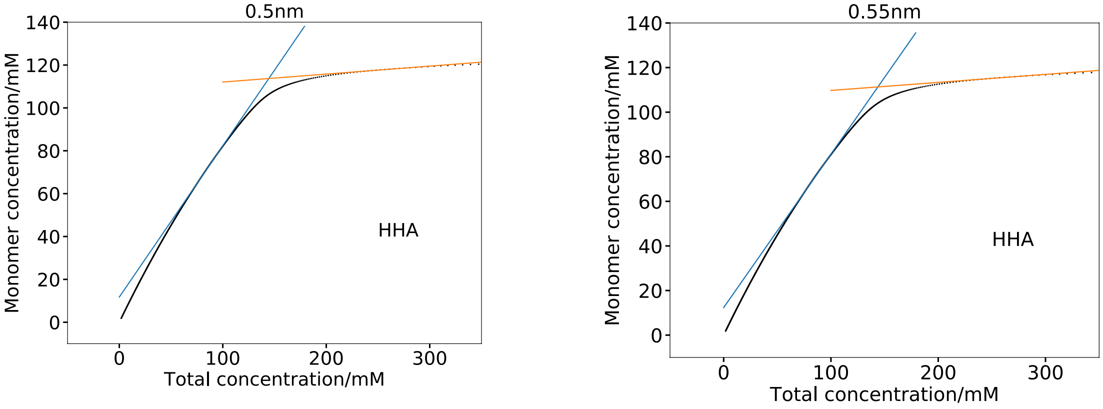
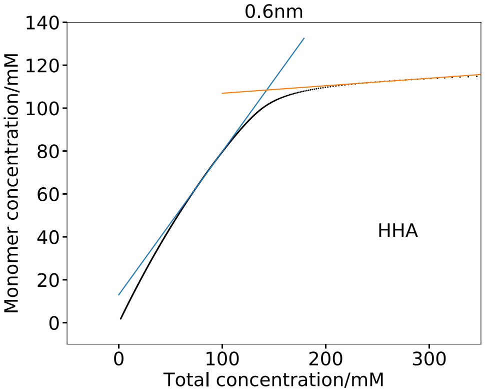


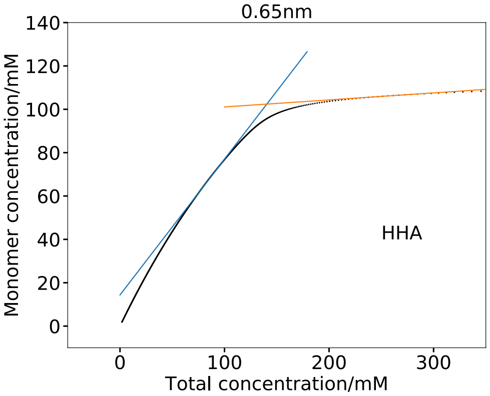
Figure S5. Free monomer concentration vs. total OPC concentration predicted from cluster free energy profile for different force fields, labelled with cluster criteria r_{cut} values in nm. Lines are fits whose intersections are used to estimate cmc.



HHA




0.6nm
Monomer concentration/mM
Total concentration/mM
HHA
0
20
40
60
80
100
120
140
0
100
200
300


0.65nm
Monomer concentration/mM
Total concentration/mM
HHA
0
20
40
60
80
100
120
140
0
100
200
300

Appendix IV. Supporting Information for Chapter 4

SI Table 1. Simulation setup for DPC micellization

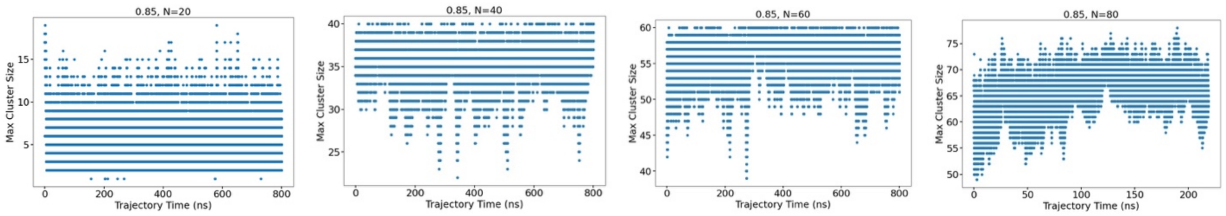
N_DPC	N_SPC water	Effective Volume (nm^3)	Trajectory (ns)
WS = 0.85			
1	6109	187.11867	50
20	6109	187.11867	800
25	6109	187.11867	800
30	6109	187.11867	800
35	6109	187.11867	800
40	6109	187.11867	800
45	6109	187.11867	800
50	6109	187.11867	800
55	6109	187.11867	800
60	6109	187.11867	800
65	6922	212.02086	800
70	11235	344.12805	800
75	11235	344.12805	800
80	11245	344.43435	800
WS = 0.9			
1	6109	187.11867	50
17	6109	187.11867	400
25	6109	187.11867	800
30	6109	187.11867	800

35	6109	187.11867	800
40	6109	187.11867	800
45	6109	187.11867	800
50	6109	187.11867	400
55	6109	187.11867	800
60	6109	187.11867	800
65	6922	212.02086	400
70	7000	214.41	400
75	7000	214.41	400
80	7000	214.41	400
FS			
1	6109	187.11867	50
15	6109	187.11867	400
20	6109	187.11867	200
25	6109	187.11867	400
30	6109	187.11867	300
35	6109	187.11867	400
40	6109	187.11867	300
45	6109	187.11867	400
50	6109	187.11867	450
55	6109	187.11867	100
60	6109	187.11867	100
65	6922	212.02086	140
70	11639	356.50257	100
75	11235	344.12805	150

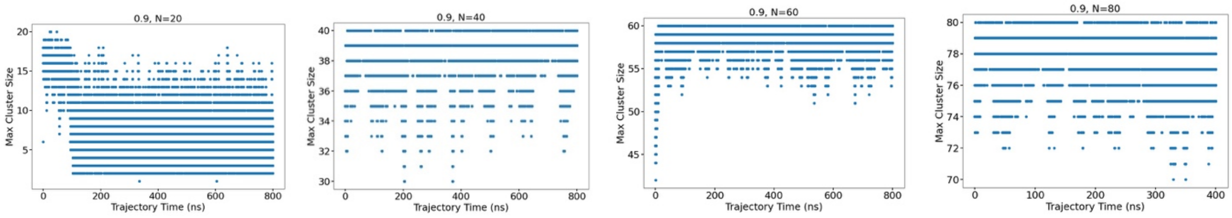
80	11430	350.1009	90

SI Figure 1. Maximum cluster size distributions for systems of N_DPC =20, 40, 60, 80

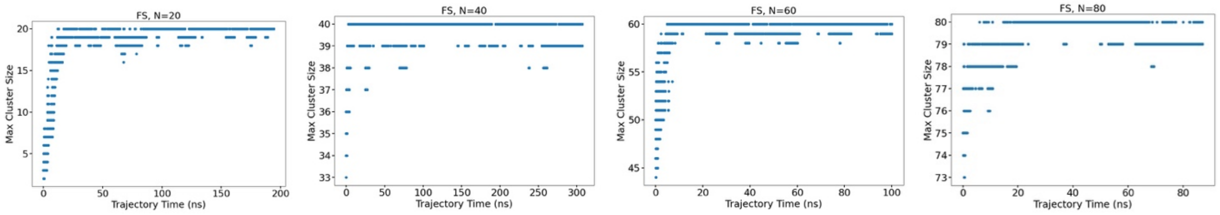
WS=0.85



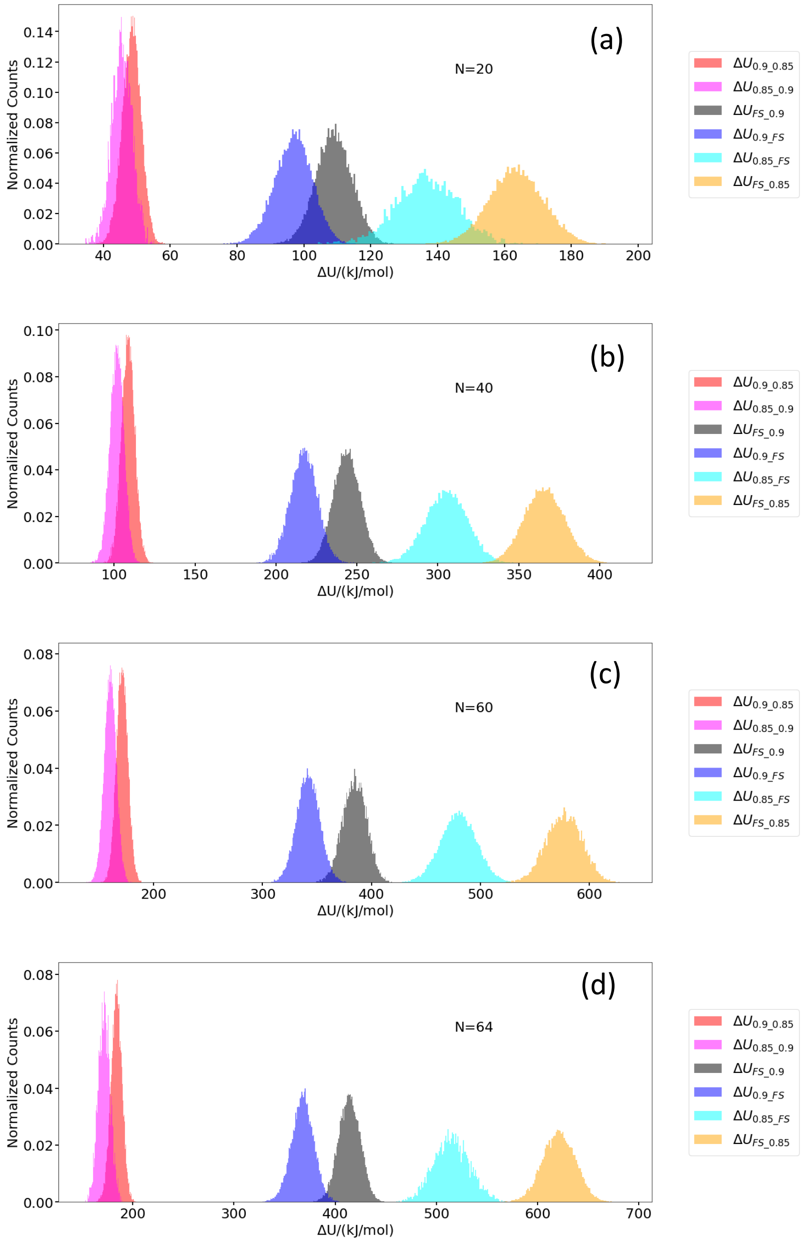
WS=0.9



FS



SI Figure 2. Distribution of the potential energy difference ΔU_{ij} and ΔU_{ji} between paired states for (a)N=20 (b)N=40 (c)N=60 (d) N=64. For the notations, $\Delta U_{ij} = U_{jj} - U_{ij}$, which denotes the potential energy difference between a configuration generated and sampled in j state and a configuration in i state and resampled in j state


(a)
N=20
Normalized Counts
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
40
60
80
100
120
140
160
180
200
ΔU/(kJ/mol)
$\Delta U_{0.9\_0.85}$
$\Delta U_{0.85\_0.9}$
$\Delta U_{FS\_0.9}$
$\Delta U_{0.9\_FS}$
$\Delta U_{0.85\_FS}$
$\Delta U_{FS\_0.85}$
(b)
N=40
Normalized Counts
0.00
0.02
0.04
0.06
0.08
0.10
100
150
200
250
300
350
400
ΔU/(kJ/mol)
$\Delta U_{0.9\_0.85}$
$\Delta U_{0.85\_0.9}$
$\Delta U_{FS\_0.9}$
$\Delta U_{0.9\_FS}$
$\Delta U_{0.85\_FS}$
$\Delta U_{FS\_0.85}$
(c)
N=60
Normalized Counts
0.00
0.02
0.04
0.06
0.08
200
300
400
500
600
ΔU/(kJ/mol)
$\Delta U_{0.9\_0.85}$
$\Delta U_{0.85\_0.9}$
$\Delta U_{FS\_0.9}$
$\Delta U_{0.9\_FS}$
$\Delta U_{0.85\_FS}$
$\Delta U_{FS\_0.85}$
(d)
N=64
Normalized Counts
0.00
0.02
0.04
0.06
0.08
200
300
400
500
600
700
ΔU/(kJ/mol)
$\Delta U_{0.9\_0.85}$
$\Delta U_{0.85\_0.9}$
$\Delta U_{FS\_0.9}$
$\Delta U_{0.9\_FS}$
$\Delta U_{0.85\_FS}$
$\Delta U_{FS\_0.85}$

References

1. Medoš, Z. i.; Bešter-Rogač, M., Two-step micellization model: the case of long-chain carboxylates in water. *Langmuir* **2017,** *33* (31), 7722-7731.
2. Knight, C. J.; Hub, J. S., WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic Acids Res.* **2015,** *43* (Web Server issue), W225-W230.
3. Vinogradova, O.; Sönnichsen, F.; Sanders, I. C. R., *J. Biomol. NMR* **1998,** *11* (4), 381-386.
4. Eastoe, J.; Hollamby, M. J.; Hudson, L., Recent advances in nanoparticle synthesis with reversed micelles. *Adv. Colloid Interface Sci.* **2006,** *128-130*, 5-15.
5. Johnston, M. A.; Swope, W. C.; Jordan, K. E.; Warren, P. B.; Noro, M. G.; Bray, D. J.; Anderson, R. L., Toward a Standard Protocol for Micelle Simulation. *J. Phys. Chem. B* **2016,** *120* (26), 6337-6351.
6. Kindt, J. T., Accounting for Finite-Number Effects on Cluster Size Distributions in Simulations of Equilibrium Aggregation. *J. Chem. Theory Comput.* **2013,** *9* (1), 147-152.
7. ten Wolde, P. R.; Frenkel, D., Computer simulation study of gas–liquid nucleation in a Lennard-Jones system. *J. Chem. Phys.* **1998,** *109* (22), 9901-9918.
8. Chen, B.; Siepmann, J. I.; Oh, K. J.; Klein, M. L., Aggregation-volume-bias Monte Carlo simulations of vapor-liquid nucleation barriers for Lennard-Jonesium. *J. Chem. Phys.* **2001,** *115* (23), 10903-10913.
9. Stubbs, J. M.; Potoff, J. J.; Siepmann, J. I., Transferable Potentials for Phase Equilibria. 6. United-Atom Description for Ethers, Glycols, Ketones, and Aldehydes. *J. Phys. Chem. B* **2004,** *108* (45), 17596-17605.
10. Oxtoby, D. W., Homogeneous nucleation: theory and experiment. *J. Phys.: Condens. Matter* **1992,** *4* (38), 7627-7650.
11. Wolde, P. R. t.; Frenkel, D., Computer simulation study of gas–liquid nucleation in a Lennard-Jones system. *J. Chem. Phys.* **1998,** *109* (22), 9901-9918.
12. Oh, K.; Zeng, X. C., Formation free energy of clusters in vapor-liquid nucleation: A Monte Carlo simulation study. *J. Chem. Phys.* **1999,** *110* (9), 4471-4476.
13. Kathmann, S. M.; Schenter, G. K.; Garrett, B. C.; Chen, B.; Siepmann, J. I., Thermodynamics and Kinetics of Nanoclusters Controlling Gas-to-Particle Nucleation. *The Journal of Physical Chemistry C* **2009,** *113* (24), 10354-10370.
14. Ben-Shaul, A.; Gelbart, W. M., Statistical thermodynamics of amphiphile self-assembly: structure and phase transitions in micellar solutions. In *Micelles, Membranes, Microemulsions, and Monolayers*, Springer: 1994; pp 1-104.
15. Srinivasan, V.; Blankschtein, D., Effect of counterion binding on micellar solution behavior: 1. Molecular− thermodynamic theory of micellization of ionic surfactants. *Langmuir* **2003,** *19* (23), 9932-9945.
16. Aliaga, C.; Bravo-Moraga, F.; Gonzalez-Nilo, D.; Márquez, S.; Lühr, S.; Mena, G.; Rezende, M. C., Location of TEMPO derivatives in micelles: subtle effect of the probe orientation. *Food Chem.* **2016,** *192*, 395-401.
17. Cheng, X.; Kim, J.-K.; Kim, Y.; Bowie, J. U.; Im, W., Molecular dynamics simulation strategies for protein–micelle complexes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016,** *1858* (7, Part B), 1566-1572.

18. Date, A.; Ishizuka, R.; Matubayasi, N., Energetics of nonpolar and polar compounds in cationic, anionic, and nonionic micelles studied by all-atom molecular dynamics simulation combined with a theory of solutions. *Physical Chemistry Chemical Physics* **2016,** *18* (19), 13223-13231.
19. De Nicola, A.; Kawakatsu, T.; Rosano, C.; Celino, M.; Rocco, M.; Milano, G., Self-Assembly of triton X-100 in water solutions: A multiscale simulation study linking mesoscale to atomistic models. *J. Chem. Theory Comput.* **2015,** *11* (10), 4959-4971.
20. Gujt, J.; Bešter-Rogač, M.; Spohr, E., Structure and stability of long rod-like dodecyltrimethylammonium chloride micelles in solutions of hydroxybenzoates: a molecular dynamics simulation study. *Langmuir* **2016,** *32* (32), 8275-8286.
21. Kawada, S.; Komori, M.; Fujimoto, K.; Yoshii, N.; Okazaki, S., Molecular dynamics study of the formation mechanisms of ionic SDS and nonionic C12E8 micelles and n-dodecane droplets. *Chem. Phys. Lett.* **2016,** *646*, 36-40.
22. Lapelosa, M.; Patapoff, T. W.; Zarraga, I. E., Molecular simulations of micellar aggregation of polysorbate 20 ester fractions and their interaction with N-phenyl-1-naphthylamine dye. *Biophys. Chem.* **2016,** *213*, 17-24.
23. Liang, X.; Marchi, M.; Guo, C.; Dang, Z.; Abel, S., Atomistic simulation of solubilization of polycyclic aromatic hydrocarbons in a sodium dodecyl sulfate micelle. *Langmuir* **2016,** *32* (15), 3645-3654.
24. Wei, Y.; Liu, G.; Wang, Z.; Yuan, S., Molecular dynamics study on the aggregation behaviour of different positional isomers of sodium dodecyl benzenesulphonate. *RSC advances* **2016,** *6* (55), 49708-49716.
25. Méndez-Bermúdez, J. G.; Dominguez, H., Structural changes of a sodium dodecyl sulfate (SDS) micelle induced by alcohol molecules. *J. Mol. Model.* **2016,** *22* (1), 33.
26. Mustan, F.; Ivanova, A.; Madjarova, G.; Tcholakova, S.; Denkov, N., Molecular dynamics simulation of the aggregation patterns in aqueous solutions of bile salts at physiological conditions. *J. Phys. Chem. B* **2015,** *119* (51), 15631-15643.
27. Raffaini, G.; Mazzaglia, A.; Ganazzoli, F., Aggregation behaviour of amphiphilic cyclodextrins: the nucleation stage by atomistic molecular dynamics simulations. *Beilstein journal of organic chemistry* **2015,** *11* (1), 2459-2473.
28. Santos, D. P.; Longo, R. L., Molecular dynamics simulations of specific anion adsorption on sulfobetaine (SB3-14) micelles. *J. Phys. Chem. B* **2016,** *120* (10), 2771-2780.
29. Sulatha, M. S.; Natarajan, U., Molecular dynamics simulations of adsorption of poly (acrylic acid) and poly (methacrylic acid) on dodecyltrimethylammonium chloride micelle in water: effect of charge density. *J. Phys. Chem. B* **2015,** *119* (38), 12526-12539.
30. Verma, R.; Mishra, A.; Mitchell-Koch, K. R., Molecular modeling of cetylpyridinium bromide, a cationic surfactant, in solutions and micelle. *J. Chem. Theory Comput.* **2015,** *11* (11), 5415-5425.
31. Wang, L.; Fujimoto, K.; Yoshii, N.; Okazaki, S., A molecular dynamics study of the breathing and deforming modes of the spherical ionic SDS and nonionic C12E8 micelles. *J. Chem. Phys.* **2016,** *144* (3), 034903.
32. Patel, L. A.; Kindt, J. T., Cluster Free Energies from Simple Simulations of Small Numbers of Aggregants: Nucleation of Liquid MTBE from Vapor and Aqueous Phases. *J. Chem. Theory Comput.* **2017,** *13* (3), 1023-1033.

33. Ahlgren, S.; Ono, K., Addition and counting: the arithmetic of partitions. *Notices of the AMS* **2001,** *48* (9), 978-984.

34. Auluck, F. In *On partitions of bipartite numbers*, Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge University Press: 1953; pp 72-83.

35. Bureaux, J. In *Partitions of large unbalanced bipartites*, Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge University Press: 2014; pp 469-487.

36. Sanders, S. A.; Sammalkorpi, M.; Panagiotopoulos, A. Z., Atomistic Simulations of Micellization of Sodium Hexyl, Heptyl, Octyl, and Nonyl Sulfates. *J. Phys. Chem. B* **2012,** *116* (8), 2430-2437.

37. Jusufi, A.; LeBard, D. N.; Levine, B. G.; Klein, M. L., Surfactant concentration effects on micellar properties. *J. Phys. Chem. B* **2012,** *116* (3), 987-991.

38. Quina, F. H.; Nassar, P. M.; Bonilha, J. B. S.; Bales, B. L., Growth of Sodium Dodecyl Sulfate Micelles with Detergent Concentration. *The Journal of Physical Chemistry* **1995,** *99* (46), 17028-17031.

39. Burov, S. V.; Shchekin, A. K., Aggregation work at polydisperse micellization: Ideal solution and “dressed micelle” models comparing to molecular dynamics simulations. *J. Chem. Phys.* **2010,** *133* (24), 244109.

40. Arnarez, C.; Uusitalo, J. J.; Masman, M. F.; Ingólfsson, H. I.; De Jong, D. H.; Melo, M. N.; Periole, X.; De Vries, A. H.; Marrink, S. J., Dry Martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. *J. Chem. Theory Comput.* **2015,** *11* (1), 260-275.

41. Yuan, F.; Wang, S.; Larson, R. G., Potentials of Mean Force and Escape Times of Surfactants from Micelles and Hydrophobic Surfaces Using Molecular Dynamics Simulations. *Langmuir* **2015,** *31* (4), 1336-1343.

42. Yuan, F.; Larson, R. G., Multiscale molecular dynamics simulations of model hydrophobically modified ethylene oxide urethane micelles. *J. Phys. Chem. B* **2015,** *119* (38), 12540-12551.

43. Mohan, G.; Kopelevich, D. I., A multiscale model for kinetics of formation and disintegration of spherical micelles. *J. Chem. Phys.* **2008,** *128* (4), 044905.

44. Bernardino, K.; de Moura, A. F., Aggregation Thermodynamics of Sodium Octanoate Micelles Studied by Means of Molecular Dynamics Simulations. *J. Phys. Chem. B* **2013,** *117* (24), 7324-7334.

45. 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* (18), 184901.

46. Stojmenović, I.; Zoghbi, A., Fast algorithms for genegrating integer partitions. *International journal of computer mathematics* **1998,** *70* (2), 319-332.

47. Faa di Bruno, F., Sullo sviluppo delle funzioni. *Annali di scienze matematiche e fisiche* **1855,** *6*, 479-480.

48. Andrews, G. E., Number Theory. *Number Theory; Dover Books: Mineola, NY,* **1994**.

49. McQuarrie, D. A., Statistical Thermodynamics. *University Science Books: Mill Valley* **1973**.

50. Ono, K. R., L.; Schneider, R. , *In Exploring the Riemann Zeta-Function: 190 Years from Riemann's Birth; Montgomery, H., Nikeghbali, A., Rassias, M., Eds.; Springer: New York,* **2017**

.

51. Muller-Kirsten, H. J., *Basics of statistical physics*. World Scientific Publishing Company: 2013.

52. Hardy, M., Combinatorics of partial derivatives. *arXiv preprint math/0601149* **2006**.

53. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E., GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008,** *4* (3), 435-447.

54. Wang, S.; Larson, R. G., Coarse-Grained Molecular Dynamics Simulation of Self-Assembly and Surface Adsorption of Ionic Surfactants Using an Implicit Water Model. *Langmuir* **2015,** *31* (4), 1262-1271.

55. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.

56. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126* (1), 014101.

57. Santos, A. P.; Panagiotopoulos, A. Z., Determination of the critical micelle concentration in simulations of surfactant systems. *J. Chem. Phys.* **2016,** *144* (4), 044709.

58. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graphics* **1996,** *14* (1), 33-38.

59. Maibaum, L.; Dinner, A. R.; Chandler, D., Micelle Formation and the Hydrophobic Effect. *J. Phys. Chem. B* **2004,** *108* (21), 6778-6781.

60. Girshick, S. L.; Chiu, C. P., Kinetic nucleation theory: A new expression for the rate of homogeneous nucleation from an ideal supersaturated vapor. *J. Chem. Phys.* **1990,** *93* (2), 1273-1277.

61. Gettins, J.; Jobling, P. L.; Walsh, M. F.; Wyn-Jones, E., Ultrasonic relaxation of surfactants in water and deuterium oxide solutions. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics* **1980,** *76*, 794-802.

62. Fletcher, P. D.; Reinsborough, V. C., Effect of short-chained sodium alkylsulfate micelles on the kinetics of metal complex formation. *Can. J. Chem.* **1981,** *59* (9), 1361-1367.

63. Jönsson, B.; Edholm, O.; Teleman, O., Molecular dynamics simulations of a sodium octanoate micelle in aqueous solution. *J. Chem. Phys.* **1986,** *85* (4), 2259-2271.

64. Shelley, J.; Watanabe, K.; Klein, M. L., Simulation of sodium octanoate micelles in aqueous solution. *Electrochim. Acta* **1991,** *36* (11-12), 1729-1734.

65. Brodskaya, E. N., Computer simulations of micellar systems. *Colloid J.* **2012,** *74* (2), 154-171.

66. Wymore, T.; Gao, X. F.; Wong, T. C., Molecular dynamics simulation of the structure and dynamics of a dodecylphosphocholine micelle in aqueous solution. *J. Mol. Struct.* **1999,** *485-486*, 195-210.

67. Tieleman, D. P.; van der Spoel, D.; Berendsen, H. J. C., Molecular Dynamics Simulations of Dodecylphosphocholine Micelles at Three Different Aggregate Sizes: Micellar Structure and Chain Relaxation. *J. Phys. Chem. B* **2000,** *104* (27), 6380-6388.

68. Abel, S.; Dupradeau, F.-Y.; Marchi, M., Molecular Dynamics Simulations of a Characteristic DPC Micelle in Water. *J. Chem. Theory Comput.* **2012,** *8* (11), 4610-4623.

69. Martin, M. G.; Siepmann, J. I., Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998,** *102* (14), 2569-2577.

70. Ashbaugh, H. S.; Liu, L.; Surampudi, L. N., Optimization of linear and branched alkane interactions with water to simulate hydrophobic hydration. *J. Chem. Phys.* **2011,** *135* (5), 054510.

71. Zhang, X.; Patel, L. A.; Beckwith, O.; Schneider, R.; Weeden, C. J.; Kindt, J. T., Extracting Aggregation Free Energies of Mixed Clusters from Simulations of Small Systems: Application to Ionic Surfactant Micelles. *J. Chem. Theory Comput.* **2017,** *13* (11), 5195-5206.

72. Martin, M. G.; Siepmann, J. I., Predicting multicomponent phase equilibria and free energies of transfer for alkanes by molecular simulation. *Journal of the American Chemical Society* **1997,** *119* (38), 8921-8924.

73. Poger, D.; Van Gunsteren, W. F.; Mark, A. E., A new force field for simulating phosphatidylcholine bilayers. *J. Comput. Chem.* **2010,** *31* (6), 1117-1125.

74. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F., A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force‐field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004,** *25* (13), 1656-1676.

75. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.

76. Bennett, C. H., Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976,** *22* (2), 245-268.

77. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.

78. Páll, S.; Hess, B., A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* **2013,** *184* (12), 2641-2650.

79. Ben-Amotz, D.; Stell, G., Reformulation of Weeks−Chandler−Andersen Perturbation Theory Directly in Terms of a Hard-Sphere Reference System. *J. Phys. Chem. B* **2004,** *108* (21), 6877-6882.

80. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.

81. Dmitri, I. S.; Michel, H. J. K., Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **2003,** *66* (10), 1735.

82. Abascal, J. L.; Vega, C., A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005,** *123* (23), 234505.

83. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F., Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011,** *40* (7), 843.

84. Mao, Y.; Zhang, Y., Thermal conductivity, shear viscosity and specific heat of rigid water models. *Chem. Phys. Lett.* **2012,** *542*, 37-41.

85. Rusanov, A. I.; Kuni, F. M.; Grinin, A. P.; Shchekin, A. K., Thermodynamic Characteristics of Micellization in the Droplet Model of Surfactant Spherical Molecular Aggregate. *Colloid J.* **2002,** *64* (5), 605-615.

86. Grinin, A. P.; Rusanov, A. I.; Kuni, F. M.; Shchekin, A. K., Thermodynamic Characteristics of a Spherical Molecular Surfactant Aggregate in a Quasi-Droplet Model. *Colloid J.* **2003,** *65* (2), 145-154.

87. Gregory D. Smith, V. G. P., An Inexpensive Glass Capillary 'Tensiometer' for Determining the Critical Micelle Concentration of Surfactants. *Chemical Educator* **2009,** *14* (6), 239–242
88. Tanford, C., Theory of micelle formation in aqueous solutions. *The Journal of Physical Chemistry* **1974,** *78* (24), 2469-2479.
89. Rharbi, Y.; Li, M.; Winnik, M. A.; Hahn, K. G., Temperature dependence of fusion and fragmentation kinetics of Triton X-100 micelles. *Journal of the American Chemical Society* **2000,** *122* (26), 6242-6251.
90. Pambou, E.; Crewe, J.; Yaseen, M.; Padia, F. N.; Rogers, S.; Wang, D.; Xu, H.; Lu, J. R., Structural features of micelles of zwitterionic dodecyl-phosphocholine (C12PC) surfactants studied by small-angle neutron scattering. *Langmuir* **2015,** *31* (36), 9781-9789.
91. Lipfert, J.; Columbus, L.; Chu, V. B.; Lesley, S. A.; Doniach, S., Size and Shape of Detergent Micelles Determined by Small-Angle X-ray Scattering. *J. Phys. Chem. B* **2007,** *111* (43), 12427-12438.
92. Oliver, R. C.; Lipfert, J.; Fox, D. A.; Lo, R. H.; Doniach, S.; Columbus, L., Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group. *PLoS ONE* **2013,** *8* (5), e62488.
93. Jusufi, A.; Kohlmeyer, A.; Sztucki, M.; Narayanan, T.; Ballauff, M., Self-Assembly of Charged Surfactants: Full Comparison of Molecular Simulations and Scattering Experiments. *Langmuir* **2012,** *28* (51), 17632-17641.
94. Faramarzi, S.; Bonnett, B.; Scaggs, C. A.; Hoffmaster, A.; Grodi, D.; Harvey, E.; Mertz, B., Molecular Dynamics Simulations as a Tool for Accurate Determination of Surfactant Micelle Properties. *Langmuir* **2017,** *33* (38), 9934-9943.
95. Chen, P.-c.; Hub, Jochen S., Validating Solution Ensembles from Molecular Dynamics Simulation by Wide-Angle X-ray Scattering Data. *Biophys. J.* **2014,** *107* (2), 435-447.
96. Diamant, H.; Andelman, D., Free energy approach to micellization and aggregation: Equilibrium, metastability, and kinetics. *Current Opinion in Colloid & Interface Science* **2016,** *22*, 94-98.
97. Zhang, X.; Arce Nunez, J. G.; Kindt, J. T., Derivation of micelle size-dependent free energies of aggregation for octyl phosphocholine from molecular dynamics simulation. *Fluid Phase Equilibria* **2019,** *485*, 83-93.
98. de Ruiter, A.; Boresch, S.; Oostenbrink, C., Comparison of thermodynamic integration and Bennett acceptance ratio for calculating relative protein‐ligand binding free energies. *J. Comput. Chem.* **2013,** *34* (12), 1024-1034.
99. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008,** *129* (12), 124105.
100. Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S., Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett.* **2003,** *91* (14), 140601.
101. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., THE WEIGHTED HISTOGRAM ANALYSIS METHOD FOR FREE-ENERGY CALCULATIONS ON BIOMOLECULES .1. THE METHOD. *J. Comput. Chem.* **1992,** *13* (8), 1011-1021.
102. Franke, D.; Petoukhov, M.; Konarev, P.; Panjkovich, A.; Tuukkanen, A.; Mertens, H.; Kikhney, A.; Hajizadeh, N.; Franklin, J.; Jeffries, C., ATSAS 2.8: a comprehensive data analysis

suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **2017,** *50* (4), 1212-1225.

103. Ahn, Y. N.; Mohan, G.; Kopelevich, D. I., Collective degrees of freedom involved in absorption and desorption of surfactant molecules in spherical non-ionic micelles. *J. Chem. Phys.* **2012,** *137* (16), 164902.

104. Lauterwein, J.; Bösch, C.; Brown, L. R.; Wüthrich, K., Physicochemical studies of the protein-lipid interactions in melittin-containing micelles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1979,** *556* (2), 244-264.

105. Yaseen, M.; Wang, Y.; Su, T. J.; Lu, J. R., Surface adsorption of zwitterionic surfactants: n-alkyl phosphocholines characterised by surface tensiometry and neutron reflection. *J. Colloid Interface Sci.* **2005,** *288* (2), 361-370.

106. Palladino, P.; Rossi, F.; Ragone, R., Effective critical micellar concentration of a zwitterionic detergent: a fluorimetric study on n-dodecyl phosphocholine. *J. Fluoresc.* **2010,** *20* (1), 191-196.

107. Sikorska, E.; Wyrzykowski, D.; Szutkowski, K.; Greber, K.; Lubecka, E. A.; Zhukov, I., Thermodynamics, size, and dynamics of zwitterionic dodecylphosphocholine and anionic sodium dodecyl sulfate mixed micelles. *J. Therm. Anal. Calorim.* **2016,** *123* (1), 511-523.

108. Göbl, C.; Dulle, M.; Hohlweg, W.; Grossauer, J.; Falsone, S. F.; Glatter, O.; Zangger, K., Influence of Phosphocholine Alkyl Chain Length on Peptide−Micelle Interactions and Micellar Size and Shape. *J. Phys. Chem. B* **2010,** *114* (13), 4717-4724.