Coarse-Grained, Density Dependent Implicit Solvent Model Reliably Reproduces Behavior of a Model Surfactant System

Erik C. Allen, Gregory C. Rutledge*

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA

***Corresponding author:** rutledge@mit.edu

Keywords: Coarse-grain simulation, implicit solvent, density dependence, micelle, surfactant

Abstract:

Density dependent, implicit solvent (DDIS) potentials, the generation of which has been described previously [E.C. Allen and G.C. Rutledge, J. Chem. Phys. **128**, 154115 (2008); E.C. Allen and G.C. Rutledge, J. Chem. Phys. **130**, 034904 (2009)], are used in this work to examine the self-assembly of a model surfactant system. While the measurement of thermodynamic properties in simulations of solvated micelles requires large computational resources or specialized free-energy calculations, the high degree of coarse-graining enabled by the DDIS algorithm allows for the measurement of critical micelle concentration and aggregation number distribution using single processor NVT simulations. In order to evaluate the transferability of potentials derived from the DDIS methodology, the potentials are derived from simulations of simple monomeric solutes and used in the surfactant system without modification. Despite the high degree of

coarse-graining and the simplicity of the fitting simulations, we demonstrate that the coarse-grained DDIS potentials generated by this method reliably reproduce key properties of the underlying surfactant system: the critical micelle concentration, and the average aggregation number. The success of the DDIS algorithm suggests its utility for more realistic surfactant models.

Introduction

A surfactant molecule is composed of a head group that is compatible with the surrounding solvent, for example water, and a tail group that is not. Above a critical concentration in solution, the surfactant molecules spontaneously aggregate into structures known as micelles, in which the head groups surround and shield the tail groups from the solvent.

Theoretical approaches[1-2] treat micellization as arising from a set of competing free energy effects. The transfer of hydrophobic tails from melt to solution promotes micelle formation, but is offset by a surface energy penalty and the loss of surfactant translational entropy. These competing effects create a well-defined free energy minimum as a function of aggregation number, so that the resulting micelles are of nearly uniform size. The concentration at which surfactant molecules spontaneously form micelles is referred to as the critical micelle concentration, or CMC, and in theoretical treatments is directly related to the aforementioned free energy considerations. Both the CMC and the average aggregation number are important properties that characterize the self-assembling nature of surfactants and for which we desire efficient, predictive capabilities.

The literature on particle-based surfactant simulations contains a diverse set of particle representation approaches[3-22], which can be grouped according to two key design decisions: the level of detail used to describe the surfactant molecule itself, and explicit or implicit treatment of the solvent. Explicit solvent simulations[3-19] are computationally

burdensome because of the low value of experimentally measured CMC's (10^{-6}-10^{-2} M). At these concentrations, a significant majority of the simulation cell is composed of the relatively uninteresting solvent. As a result, the direct measurement of the CMC by explicit solvent simulations is impractical, even for very coarse-grained surfactant and solvent representations. Instead, the CMC and aggregation number distribution are generally determined indirectly, through the use of free energy simulations[10,14,15,18].

By reducing the number of particles represented, and thus also the computational burden, implicit solvent simulations of surfactant systems allow investigation of much longer time and length scales. By this approach, Lazaridis *et al*.[21] studied the formation of dodecylphosphocholine (DPC) micelles using the Effective Energy Function[23] (EEF1) implicit solvent model. The simulated CMC was close to the reported experimental value. Von Gottberg *et al.*[22] studied A_2B_2 surfactants in implicit solvent using stochastic dynamics (SD), in which each surfactant bead was intended to represent a Kuhn segment (approximately ten methylene groups) of a polymer chain. They successfully simulated a fully equilibrated micelle size distribution at multiple surfactant concentrations, from which they were able to determine the system CMC.

In this paper, we employ our previously developed coarse-graining algorithm to generate density-dependent, implicit solvent (DDIS) potentials[24,25] for a model surfactant and test their ability to describe surfactant aggregation. The DDIS potentials are specifically derived to reproduce the chemical potentials and radial distribution functions (RDFs) of an underlying all-atom simulation. The advantage of creating an implicit solvent energy

model via coarse-graining is that the results have a clear correspondence to an underlying all-atom energy model that is presumed to be a more accurate representation of the system.

To investigate the applicability of DDIS potentials to surfactant systems, we create and test such potentials for a model surfactant in a Lennard-Jones solvent. The model is a derivation of Smit's[11-13] well-known surfactant model, for which Pool and Bolhuis recently estimated the CMC and aggregation number distribution by means of free energy calculations[18].

As in our previous work, we derive the DDIS potentials from simulations of monomeric solutes in explicit solvent. Our potentials are not derived from mixtures of solute types nor chain simulations. We made this decision because it represents the simplest set of simulations from which to derive DDIS potentials, and a worst case scenario for transferability of the resulting potentials. It allows us to establish a "baseline" for transferability of the DDIS potentials, against which other, more sophisticated derivation methodologies might be compared. We use the results of these previous works to test the accuracy and transferability of our DDIS potentials to this surfactant model.

Theory

Density Notation Conventions

In this work, ρ_S refers to the *total* density of solute particles, where the subscript "S" stands for solute, and may include contributions from different solute types. ρ_I is the density of type I solute particles only. $\{\rho\}$ refers to the set of ρ_I that includes every solute particle type in the system. Finally, ρ without subscript refers to the state point density, considering all particles (solute and solvent) explicitly. A superscript L indicates that the density is the *local* density of solute particles; otherwise, the density refers to the global average density.

Particles of different size contribute to the density in proportion to their volume, which is consistent with the Effective Energy Function[23] implicit solvent model. This convention is chosen because each solute particle reduces the local solvent density roughly in proportion to its volume. Thus, we define the solute density in a specified volume V as

$$\rho_S = \frac{1}{V}\sum_{i=1}^{N}\left(\frac{\sigma_i}{\sigma_W}\right)^3, \tag{1}$$

where N is the number of solute particles in the specified volume, σ_i is the Lennard-Jones radius for particle i, and σ_W is the Lennard-Jones radius of the solvent.

DDIS Potential Review

In our previous work[24], we proposed a density-dependent, implicit solvent (DDIS) potential with the form

$$E_i = \mu\left(\rho_{S,i}^{L}\right) + \frac{1}{2}\sum_{j\neq i} V_{IJ}^{EFF}\left(r_{ij},\rho_{S,i}^{L}\right) \tag{2}$$

where E_i is the effective energy of particle i, V_{IJ}^{EFF} is a pairwise potential between I- and J-type solute particles that is an explicit function of solute particle density in the vicinity of particle i, $\rho_{S,i}^{L}$, and μ is a "self-interaction" term that is also a function of solute particle density. The first term serves to capture the effects of solute-solvent and solvent-solvent interactions in a mean field approximation, while the second term represents solute-solute interactions that are mediated by solvent. The dependence of both terms on local density allows the potential to be cognizant of the greater or lesser influence of solvent in the different regions of a system that is inhomogeneous on the solute length scale, such as that which arises in a system of surfactants that self-assemble into micelles. We also proposed a method to derive such potentials from simulations of solute type I in solvent. This fitting was performed such that the solute-solute RDF and solute excess chemical potential, μ_I^{ex}, are reproduced across all solution compositions. For details of this fitting procedure, the reader is referred to Ref. 24, hereafter referred to as Paper 1.

Mixtures of multiple particle types are treated by a simple mixing rule in which like particle interactions are unmodified, but the local density used is the total local solute

density. DDIS interactions between dissimilar particles are derived from the interactions between pairs of identical particles through the equation

$$V_{IJ}^{EFF}\left(\frac{r}{\sigma_{IJ}};\rho_S^L\right)=V_{IJ}\left(\frac{r}{\sigma_{IJ}}\right)+\frac{\varepsilon_{IJ}}{2}\left(\frac{\Delta V_{II}^{EFF}\left(r/\sigma_I;\rho_S^L\right)}{\varepsilon_I}+\frac{\Delta V_{JJ}^{EFF}\left(r/\sigma_J;\rho_S^L\right)}{\varepsilon_J}\right) \tag{3}$$

where ε_{IJ} and σ_{IJ} are the Lennard-Jones parameters for *I-J* interactions, $V_{IJ}(r)$ is the *I-J* all-atom potential, and

$$\Delta V_{II}^{EFF}\left(\frac{r}{\sigma_I};\rho_S^L\right)=V_{II}\left(\frac{r}{\sigma_I}\right)-V_{II}^{EFF}\left(\frac{r}{\sigma_I};\rho_S^L\right) \tag{4}$$

The mixing rule is clearly a computational simplification – an optimized *I-J* potential could be derived across all density ranges using the DDIS approach. However, deriving the true mixed potentials for an *M*-component system would require *M(M*-1)/2 fitting steps. Ref. 25 discusses the mixing rule and its applicability further, and is hereafter referred to as Paper 2.

Simulation Metrics

The measure of error in the RDF is given by the solute-solute energy, defined as:

$$E_{L,IJ}=\sqrt{\rho_I\rho_J}\int_0^{r_C}V_{IJ}(r)g_{IJ}(r,\{\rho\})4\pi r^2dr \tag{5}$$

where ρ_I is the density of solute type I, V_{IJ} is the exact interaction potential between I- and J-type particles, and g_{IJ} is the I-J RDF. The difference between E_L for a coarse-grained system and that for the all-atom simulation provides a measure of the error in $g(r,\{\rho\})$, relative to the exact $g(r,\{\rho\},\rho)$ for the explicit, all-atom system, in units of energy. The error in the excess chemical potential of particle type I, μ_I^{ex}, is the difference (in units of ε_W, the Lennard-Jones parameter of the solvent) between the target (all-atom) and measured (coarse-grain) values. In this text, both error measurements are presented in terms of the standard error over all simulations. Additionally, we supply figures where appropriate comparing the all-atom and coarse-grained chemical potentials and RDFs, as these offer an intuitive sense of the closeness of fit.

A solute's relative affinity for the solvent can be measured by the free energy of transfer, $\Delta G_I(W \rightarrow I)$, defined as the free energy change associated with transferring a single solute particle from a bath of solvent particles W to a bath of solute particles I. The free energy of transfer can be calculated from the excess chemical potential of solute particles:

$$\Delta G_I(W \rightarrow I) = \mu_I^{ex}\left(\frac{\rho_I}{\rho} = 1\right) - \mu_I^{ex}\left(\frac{\rho_I}{\rho} = 0\right) \tag{6}$$

We characterize the local solute environment by calculating the local number of solute particles, $\langle N_{L,I} \rangle$, defined as:

$$< N_{L,I} >= \rho_I \int_0^{R_C} g(r;\rho_I) 4\pi r^2 dr \tag{7}$$

We define the solute enhancement ratio as the number of neighboring solute particles within a sphere of radius R_c around a reference particle divided by the average number of particles in such a volume. A value less than one indicates a local depletion in solute density, while a value greater than one indicates a local enhancement. A value near 1.0 indicates that the local solute environment is very similar to the global average environment.

Simulation Protocol

All-Atom Model

The underlying all-atom model described by Smit[11-13] is composed of three basic particle types: head, tail, and solvent particles, which are hereafter referred to as H, T, and W respectively. These particles interact via the truncated and shifted Lennard-Jones potential, V^{TS}, where the potential between particles *i* and *j* is described by:

$$V_{ij}^{TS}\left(r_{ij};\varepsilon_{ij},\sigma_{ij},R_{c,ij}\right) = \begin{cases} V_{ij}^{LJ}\left(r_{ij};\varepsilon_{ij},\sigma_{ij}\right) - V_{ij}^{LJ}\left(R_{C,ij};\varepsilon_{ij},\sigma_{ij}\right) & r_{ij} \le R_{C,ij} \\ 0 & r_{ij} > R_{C,ij} \end{cases} \tag{8}$$

with

$$V_{ij}^{LJ}\left(r_{ij};\varepsilon_{ij},\sigma_{ij}\right)=4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{9}$$

where ε_{ij} and σ_{ij} are the Lennard-Jones parameters for *ij* interactions. ε and σ values of dissimilar particles are governed by Lorentz-Berthelot mixing rules. Lennard-Jones parameters for all three particle types are given in Table 1. Equations (8) and (9) allow for the possibility of different cut-off radii ($R_{C,ij}$) for interactions between different particle types. ($R_{C,ij}/\sigma_{ij}$)=2.5 for W-W, H-W, and T-T interactions, and ($R_{C,ij}/\sigma_{ij}$)=$2^{1/6}$ for T-W, H-H, and H-T interactions. These values were chosen to give the H particles a "solvent-philic" behavior, and the T particles a "solvent-phobic" behavior. As indicated in Table 1, $\Delta G_H(W\rightarrow H)$ is positive for H-type particles, indicating that H-type particles prefer a solvent environment relative to one composed of other H-type particles. Conversely, $\Delta G_T(W\rightarrow T)$ is negative for T-type particles.

In many ways, the Smit model is not an ideal candidate for coarse-graining. This is due to both the variation in cutoff radius as a function of interaction type, coupled with the large discrepancy between pure component interactions and the corresponding mixed interactions. However, the availability of high-quality simulation results in the literature is an appealing property of the Smit model, and we utilize it despite these limitations.

Table 1 shows the solute enhancement ratio for all three types of solute particles at ρ_l/ρ=0.5. These results indicate that H-type particles experience a local depletion of other H-type particles, while T-type particles experience a substantial local enhancement

of other T-type particles. The local depletion of H-type particles arises because the radius of H-type particles ($2\sigma_W$) creates a large excluded volume. As a result, the total density within the cutoff radius ($5\sigma_W$) is less than the average solute density.

Surfactant molecules are composed of a single H-type particle connected to multiple T-type particles in a linear fashion, denoted H_1T_M, where M is the number of T-type particles that form the tail. Bonded interactions between I- and J-type particles are governed by a harmonic potential:

$$V_{ij}^{bond}\left(r_{ij}\right) = \frac{1}{2} k_{bond}\left(r_{ij} - r_{eq,IJ}\right)^2 \tag{12}$$

where $k_{bond}=5000\varepsilon/\sigma_W^2$ is the harmonic spring constant, and $r_{eq,IJ}=\sigma_{IJ}$ is the equilibrium bond length. The H-T equilibrium bond length, $r_{eq,HT}$, is $1.5\sigma_W$, while the T-T equilibrium bond length, $r_{eq,TT}$, is $1.0\sigma_W$. The bonded interaction applies only to nearest neighbor particles in the surfactant chain.

Simulation Details

Simulation protocols for all-atom and coarse-grained simulations of monomeric solute are given in Paper 1. Monomeric solute simulations were carried out in the *NVT* ensemble at $T^*=k_BT/\varepsilon_W=1.0$ and $\rho^*=\rho\sigma_W^3=0.60$, where ε_W and σ_W are the Lennard-Jones parameters of the solvent W.

All-atom simulations were carried out in the same manner as the monomeric all-atom simulations in Paper 1, using GROMACS[26] (Version 3.3) Molecular Dynamics simulation software. Free energy calculations in the all-atom simulation were performed using thermodynamic integration (TI)[27]. The standard GROMACS λ switching function was used to integrate between an initial state in which the test particle interactions with the remaining system were switched off and a final state in which the test particle interactions were fully enabled. Switching was applied to nonbonded interactions only, with soft core interactions to avoid singularities and using a soft core interaction parameter α=0.51. A total of 31 λ values were used (λ = [0.00 0.03 0.07 0.10 …. 0.93 0.97 1.00]).

Implicit solvent simulations of surfactants were carried out in the same manner as the monomeric DDIS simulations in Paper 1, using a Monte Carlo code of our own design. Nearest bonded neighbor particles were included in the calculation of local density. In addition to the atom-level translation moves used for monomeric simulations, simulations of surfactants included rigid body chain translation moves, rigid body chain rotation moves, and rotation about individual bonds. The proportion of these moves was 20% atom translation: 60% chain translation:10% chain rotation:10% bond rotation. Simulations were equilibrated for 10^5 cycles, followed by sampling for 4×10^5 cycles. Free energy was calculated using the Bennett Acceptance Ratio method[28]. Five λ values were used (λ = [0.00,0.25,0.50,0.75,1.00]) with an initial state comprising a non-interacting test particle and a final state having a fully interacting test particle. Free energy sampling was performed every two MC cycles, which was sufficient to generate

statistically independent samples, as determined by the autocorrelation function of the measured configurational energy of the solute particles. RDF sampling was performed every 100 MC cycles.

Micelles were identified via a clustering algorithm. Surfactant chains were designated as members of the same cluster if the distance between any tail particles in the surfactant chains was less than $1.5\sigma_T$, where σ_T is the Lennard-Jones radius for tail particles. This is the same algorithm and cutoff distance used previously by van Gottberg et al[22]. The aggregation number distribution is relatively insensitive to small changes in the cutoff distance: a 30% increase (to $1.95\sigma_T$) results in only a 10% increase in the average aggregation number. Because of the different nature of their single micelle simulations, Pool and Bolhuis used a slightly different algorithm to determine aggregation number, but the relevant cutoff distance there was also $1.5\sigma_T$ [18].

Results and Discussion

DDIS Potentials

DDIS potentials were generated for H- and T-type particles in implicit solvent W as described previously[24,25]. Figure 1 demonstrates the ability of the H-H and T-T DDIS potentials to reproduce the excess chemical potential of the equivalent all-atom simulations across a range of solute densities. The standard error in μ^{ex} for H-type particles is 0.06ε, versus 0.37ε for T-type particles.

Figure 2 shows the worst-case reproductions of the all-atom RDFs by the DDIS potentials for the H-type and T-type potentials. As with the excess chemical potential, the level of reproduction obtained for the H-type potential is much better than for the T-type potential. The standard errors in E_L were $E_{L,H}$=0.01ε and $E_{L,T}$=0.35ε. The larger relative error in the solvent-phobic T-type fitting is consistent with our previous work[25]. The characteristics of solvent-phobic particles that give rise to this difficulty of fitting are discussed in the Appendix.

We note here that although the magnitude of errors in the T-type fitting is larger than those in the H-type fitting, the errors do not display any systematic bias. Figure 1 demonstrates that the fitted excess chemical potentials have both positive and negative residuals when compared to the all-atom model.

Figure 3 compares the two-body portion of the coarse-grained potentials for the H-type and T-type potentials at a local density of ρ_I^L/ρ=0.2. At short distances, the "solvent-philic" H-type particles experience a repulsive potential where the attractive well would typically reside. The shape of the H-type potential is reminiscent of the class of potentials referred to as "Hard Core/Soft Shoulder" (HCSS) interactions[29-30]. However, the system behavior generated by the repulsive soft shoulder in this potential differs from that of previous HCSS studies. There, the potential was used at low temperatures, and the repulsive plateau promoted local particle aggregation. In this work, the soft shoulder models the first neighbor shell of implicit solvent, and actually encourages local particle depletion.

In contrast to the H-type particles, the T-type particles experience an attractive well deeper than the all-atom Lennard-Jones potential, which is responsible for increasing solute aggregation as the density of T-type particles decreases. These trends are consistent with Paper 2.

Figure 4 shows the one-body portion of the DDIS potential for H- and T-type particles. As solute density goes to zero, the value of the one-body term approaches $\mu_I^{ex}(\rho_I/\rho=0)$. This equivalence to the excess chemical potential arises because the contribution of pairwise interactions disappears as solute concentration goes to zero, and must therefore be captured by the one-body term. This is consistent with the findings of Paper 1. The profile is relatively flat for low solute densities, changing rapidly as the solute fraction approaches 1. The one-body term increases for H-type particles as density increases, indicating that it is energetically unfavorable to achieve high local densities. In contrast, the one-body term for T-type particles decreases, further promoting aggregation.

Surfactant Results

We performed multiple simulations of 216 surfactant molecules across a range of surfactant concentrations. These simulations demonstrate the true benefit of the implicit solvent approach, since the equivalent all-atom simulation is not computationally feasible. From these simulations, we measured the system CMC and average micelle aggregation number. Pool and Bohuis[18] estimated the CMC's for H_1T_4 and H_1T_5 surfactants using semi-grand canonical simulations and found values of $5\times10^{-6}\sigma_W^{-3}$, and

$6\times10^{-7}\sigma_W^{-3}$, respectively. They suggested an average aggregation number of "≈20" for H_1T_4 and "≈30" for H_1T_5, the use of the ≈ symbol indicating a substantial uncertainty in the true values of these metrics. These results are given in Table 2. Due to the presence of explicit solvent, Pool and Bolhuis were not able to observe these values directly, but required a thermodynamic framework to evaluate the results. Given the differences in simulation type, exact replication of their results is unlikely, but the results provide reasonable guidelines.

We first measured the CMC of the each surfactant type by running a series of single processor simulations at global surfactant concentrations spanning a range around the CMCs reported by Pool and Bolhuis. Figure 5 shows the number density of free surfactant molecules, ρ_1, as a function of the total number density of surfactants, ρ_S. The solid black line is the 45° line corresponding to no aggregation. As Figure 5 demonstrates, the free surfactant density initially increases along the 45° line, but deviates from that line as total surfactant density is increased. This deviation indicates the formation of micellar aggregates. The presence of micelles can be visually confirmed, as is shown in Figure 6. The clustering algorithm used in this work also identifies the presence of three distinct micelles in this snapshot. The decrease in ρ_1 above a certain ρ_S level is consistent with the findings of von Gottberg *et al.*[22], who demonstrate that such a decrease is due to excluded volume effects.

To calculate critical micelle concentration, we follow the convention of von Gottberg *et al.*[22], who define the CMC as the maximum free surfactant concentration across all

simulations. The values obtained for H_1T_4 and H_1T_5 surfactants by this method using the DDIS potentials are $4(1)\times10^{-6}\sigma_W^{-3}$, and $8(1)\times10^{-7}\sigma_W^{-3}$, where the error in the last decimal is given in parenthesis. These CMC values are within 30% of the values reported by Pool and Bolhuis, as shown in Table 2.

While a discrepancy of 30% with the all-atom work of Pool and Bolhuis might at first seem to be a cause for concern, we suggest that it is actually quite respectable. Statistical thermodynamic theories of micelle formation (such as those referenced by Pool and Bolhuis) determine the critical micelle concentration**,** X_{CMC}**,** based on the free energy of micelle formation[1,2]:

$$X_{CMC} = \exp\left[\beta g_{mic}(n^*)\right] \tag{10}$$

where $g_{mic}(n)$ is the free energy of transferring a surfactant molecule from solution to a micelle of size n, and n^* is the most probable aggregation number. Based on Equation (10), the CMC is related to the exponential of free energy of micellization, so that a 30% difference in CMC implies a difference in the free energy of micellization of only $0.25k_BT$, which is comparable in magnitude to the fitting error for the T-type potential. Given the differences in estimation methods between this work and that of Pool and Bolhuis, this level of agreement is as good as can be expected. We remind, however, that the DDIS potentials generated by this coarse-grained methodology are in general temperature-dependent and transferable only over a limited temperature range; this transferability was explored in greater detail in Paper 1. One should exercise caution in

attempting to extract any temperature dependence of the CMC using coarse-grained potentials without additional fitting.

We next measured the aggregation number distribution at a single global surfactant concentration. Sufficient sampling of this distribution was assured by generating 64 independent initial system configurations and averaging the micelle size distribution over all simulations. Figure 7 shows the H_1T_4 micelle size distributions using the DDIS potential for a total surfactant number density of $8\times10^{-6}\sigma_W^{-3}$ and the corresponding H_1T_5 distribution at $1.9\times10^{-6}\sigma_W^{-3}$. In both cases, a large fraction of the surfactant exists as free surfactant (~40%), and the profiles both demonstrate a secondary maximum. These secondary maxima indicate the presence of stable micellar structures. Both the H_1T_4 and H_1T_5 profiles demonstrate a single secondary maximum, indicating a preferred micelle size as predicted by statistical thermodynamic theories.

We calculated the average aggregation number as the number weighted average of all aggregates of size two or larger. For the H_1T_4 and H_1T_5 surfactant systems, the average aggregation numbers were 16 and 34, respectively. These numbers are extremely close to the estimates of ≈20 and ≈30 reported by Pool and Bolhuis.

We also examined the shape of the aggregation number distribution. The inset to Figure 7 shows that the DDIS models for both H_1T_4 and H_1T_5 surfactant systems demonstrate "extended" aggregation number distributions as indicated by the presence of a long tail in the aggregation number distribution, such that the mean micelle size is larger than the

most probable micelle size. Nelson *et al.*[31] demonstrated that such extended tails are associated with the presence of some fraction of cylindrical, or "worm-like", micelles. By contrast, the results of Pool and Bolhuis show a nearly normal probability distribution around the average particle size, indicating mostly spherical micelles. It is unclear at this time from where this difference arises. Although the coarse-graining process could be responsible for the discrepancy, it could also be due to differences in measurement method (Pool and Bolhuis measure the free energy of different micelle sizes at infinite dilution and infer the aggregation number distribution) or to differences in micelle definition (Pool and Bolhuis include a surfactant molecule in the micelle definition if it resides within a certain radius of the center of mass of the existing micelle, while we use a clustering algorithm).

Conclusions

In this work, we have created coarse-grained, density dependent implicit solvent (DDIS) potentials based on underlying all-atom simulations of a truncated and shifted Lennard-Jones model that has previously been shown to demonstrate micellization behavior[11-13]. Coarse-grained potentials were generated as we have previously outlined[24,25]. Potentials were fit to simulations of monomeric (i.e. single repeat units) solute particles in solvent, without explicit consideration of mixtures as inputs to the process. These simple fitting simulations were used to test the transferability of DDIS potentials for simulation environments far from the state point of fitting.

We performed large simulations of surfactant solutions at solute densities near the CMC. The CMCs of the H_1T_4 and H_1T_5 surfactants were within 30% of those estimated by Pool and Bolhuis[18], which implies a difference in free energy of micelle formation of 0.25 k_BT. Additionally, we found average micelle sizes to be extremely close to the estimates of Pool and Bolhuis. For both H_1T_4 and H_1T_5 surfactants, we observed some fraction of large aggregate sizes that were not present in the free energy methods. However, these differences cannot be directly attributed to the coarse-graining process, but may represent differences in simulation methodology or micelle definition.

The success of the DDIS fitting framework in reproducing the properties of a simple surfactant model suggests that it may also be of utility with more chemically realistic surfactant systems.

Acknowledgements

This work was supported by the Department of Energy Computational Science Graduate Fellowship Program of the Office of Science and National Nuclear Security Administration in the Department of Energy under contract DE-FG02-97ER25308.

Supplemental Information

The Supplemental Information section contains all the force fields used in this work, presented in table format.

Appendix

A full investigation of the properties of DDIS potentials for highly aggregating particles is beyond the scope of this work, but this section discusses some of the issues associated with the fitting of such potentials. Further study is required to identify best practices.

The fitting of DDIS potentials for highly aggregating (i.e. solvent-phobic) solute particles is challenging, because of the heterogeneous solute distribution in the fitting simulations. More specifically, aggregation of the solute particles implies that the *average* local solute density, $< \rho_S^L >$, is significantly higher than the global solute density ρ_S. For example, for the T-type particles described in this work, a simulation of 100 T-type solute particles in solvent at ρ_S/ρ=0.1 has an average local density $< \rho_S^L > / \rho \approx 0.4$.

In Paper 1, we demonstrated that for a homogeneous distribution of solute particles one could use a global density dependent potential to approximate the local density dependent potential at a local density equal to the *average* local, or global, solute density. When the system is not homogeneous, the average local solute density is not equal to the global solute density, so we modify this rule in one of two ways. The first is to approximate the local density dependent potential at a given local solute density by the global density dependent potential obtained at the same average local solute density. Returning to the T-type simulation with global solute density of ρ_S/ρ=0.1, given above, since the average local density of this simulation is $< \rho_S^L > / \rho \approx 0.4$, the potential derived from this simulation would be used as the local density dependent potential for local density

$< \rho_S^L > / \rho \approx 0.4$. The second approach is to combine Equations 15 and 16 from Paper 1 to obtain Equation (A-1) below:

$$\rho_S \exp\left[\frac{-V_S^{EFF}(r;\rho_S)}{k_B T}\right]\gamma(r,\rho_S) = \sum_i P\left(\rho_S^{L,i}\right)\rho_S^{L,i}\gamma\left(r,\rho_S^{L,i}\right)\exp\left[\frac{-v_S^{EFF}\left(r;\rho_S^{L,i}\right)}{k_B T}\right] \qquad \text{(A-1)}$$

where $\gamma(r,\rho_S)$ is defined as ratio of the radial distribution function to the pairwise potential. Schommers[32] states that this term is only mildly dependent on the potential, so to a first approximation we assume that $\gamma(r,\rho_S)$ is independent of density, that is $\gamma(r,\rho_S) = \gamma(r)$. In this case, the γ terms cancel from each side, and we are left with only terms relating the potential and the radial distribution function.

Equation (A-1) links the global density dependent potential V_S^{EFF} to the weighted average of local density dependent potentials, v_S^{EFF}, where the weighting is given by the distribution of local densities experienced in the simulation, $P\left(\rho_S^{L,i}\right)$. Once a set of m global density dependent potentials is obtained, a set of m equations of the form given by Equation (A-1) can be solved for the m unknown local density dependent potentials.

In this work we attempted both methods of fitting for the T-type potentials. The results were remarkably similar. As such, we suggest that either approach can be considered a suitable fitting method for aggregating solute particles.

A second consideration with respect to solvent-phobic particles is the value of the pairwise potential at the cutoff radius. It is common practice to shift RDF-inverted

potentials up or down by a constant amount to obtain a value of zero at the cutoff radius. While in a density independent potential or a global density dependent potential such a shift would have a no impact on the RDF, in the local density dependent case it produces larger or smaller levels of particle aggregation by lowering or raising the energy of specific local densities relative to others in the same simulation. This effect is most pronounced with solvent-phobic particles which experience a high level of particle aggregation. Thus, the shifts applied to the local density dependent pairwise potentials must be appropriately coupled. Here, we have treated them simply as an additional set of optimization parameters that can be exploited to improve the RDF and excess chemical fit. In this work, we found the best fits were obtained when the pairwise potentials were shifted to zero at the cutoff radius for local densities greater than $\rho_S^L / \rho = 0.2$, but to negative values on the order of -0.05k_BT for lower local densities.

References

1. S. Pubbada and D. Blankschtein, J. Chem. Phys. **92**, 3710 (1990).
2. L. Maibaum, A.R. Dinner, and D. Chandler, J. Phys. Chem. B **108**, 6778 (2004).
3. B. Jonsson, O. Edholm, and O. Teleman, J. Chem. Phys. **85**, 2259 (1986).
4. K. Watanabe, M. Ferrario, and M.L. Klein, J. Chem. Phys. **92**, 819 (1988).
5. A.D. MacKerell, J. Phys. Chem. **99**, 1846 (1995).
6. D.P. Tielman, S.J. Marrink, and H.J.C. Berendsen, Biochim. Biophys. Acta **1331**, 235 (1997).
7. M. Tarek, S. Bandyopadhyay, and M.L. Klein, J. Mol. Liq. **78**, 1 (1998).
8. S.J. Marrink, D.P. Tieleman, and A.E. Mark, J. Phys. Chem. B **104**, 12165 (2000).

9. D.P. Tielman, D. van der Spoel, and H.J.C. Berendsen, J. Phys. Chem. B **104**, 6389 (2000).
10. N. Yoshii, K. Iwahashi, and S. Okazaki, J. Chem. Phys. **124**, 184901 (2006).
11. B. Smit, Phys. Rev. A **37**, 3431 (1988).
12. B. Smit, P.A.J. Hilbers, K. Esselink, L.A.M. Rupert, N.M. van Os, and A.G. Schlijper, J. Phys. Chem. **95**, 6361 (1991).
13. B. Smit, K. Esselink, P.A.J. Hilbers, N.M. van Os, L.A.M. Rupert, and I. Szleifer, Langmuir **9**, 9 (1993).
14. P.G. Bolhuis and D. Frankel, Physica A **244**, 45 (1997).
15. M. Kinoshita and Y. Sugai, Chem. Phys. Lett. **313**, 685 (1999).
16. S.O. Nielsen, C.F. Lopez, G. Srinivas, and M.L. Klein **119**, 7043 (2003).
17. S.J. Marrink, A.H. de Vries, and A.E. Mark **108**, 750 (2004).
18. R. Pool and P.G. Bolhuis, J. Chem. Phys. B **109**, 6650 (2005).
19. M. Hatakeyama and R. Faller, Phys. Chem. Chem. Phys. **9**, 4662 (2007).
20. H. Sinto, S. Marisada, M. Miyahara, and K. Higashitani, Langmuir **20**, 2017 (2004).
21. T. Lazaridis, B. Mallik, and Y. Chen, J. Phys. Chem. B **109**, 15098 (2005).
22. F.K. von Gottberg, K.A. Smith, and T.A. Hatton, J. Chem. Phys. **106**, 9850 (1997).
23. T. Lazaridis and M. Karplus, Prot. Struct. Funct. Genet. **35**, 133 (1999).
24. E.C. Allen and G.C. Rutledge, J. Chem. Phys. **128**, 154115 (2008).
25. E.C. Allen and G.C. Rutledge, J. Chem. Phys. **130**, 034904 (2009).
26. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J.C. Berendsen, J. Comput. Chem. **26**, 1701 (2005).
27. T.P. Straatsma and J.A. McCammon, J. Chem. Phys. **95**, 1175 (1991).

28. C.H. Bennett, J. Comput. Phys. **22**, 245 (1976).

29. G. Malescio and G. Pellicane, Nature Mater. **2**, 97 (2003).

30. M.A. Glaser, G.N. Grason, R.D. Kamien, A. Kosmrlj, C.D. Santangelo, and P. Ziherl, Europhys. Lett. **78**, 46004 (2007).

31. P.H. Nelson, G.C. Rutledge, and T.A. Hatton, J. Chem. Phys. **107**, 10777 (1997).

32. W. Schommers, Phys. Lett. **43**, 157 (1973).

Particle Type	ε_{ii}	σ_{jj}	$\Delta G_I(W \rightarrow I)$	$\langle N_L \rangle / \langle N_L \rangle^{IDEAL}$
H	1.0	2.0	$13.6 \pm 0.1 \varepsilon_W$	0.95
T	1.0	1.0	$-7.04 \pm 0.05 \varepsilon_W$	1.41
W	1.0	1.0	0	1.00

Table 1: Key parameters for all-atom particle types. H-type particles are "solvent-philic", as indicated by the positive free energy of transfer. T-type are "solvent-phobic", as indicated by a negative free energy of transfer and high solute enhancement ratio.

	CMC (σ_w^{-3})		N_{AGG}	
Simulation	***H_1T_4***	***H_1T_5***	***H_1T_4***	***H_1T_5***
All-Atom[18]	5×10^{-6}	6×10^{-7}	≈20	≈30
DDIS	$4(1)\times10^{-6}$	$8(1)\times10^{-7}$	16	34

Table 2: Behavior of H_1T_4 and H_1T_5 solutions using the all-atom and DDIS potentials. Errors in the last decimal is given in parenthesis. All-atom results are taken from Ref. 18, and have been estimated via free-energy methods.

Figure 1: Comparison of excess chemical potential in all-atom (solid line) and coarse-grained (circles) simulations. Left: H-type particles in solvent W. Right: T-type particles in solvent W.

Figure 2: Comparison of worst-case fits of solute RDF in all-atom (solid line) and coarse-grained (circles) simulations. Left: H-type particles in solvent W at ρ_H=0.2. Right: T-type particles in solvent W at ρ_T=0.9.

Figure 3: Coarse-grained two-body term for local solute density ρ_S^L/ρ=0.2 (···H,--T). The dark line shows the T-type Lennard-Jones interaction for comparison.

Figure 4: Coarse-grained one-body term as a function of local solute density (Circles: H, Squares: T). Dashed lines indicate excess chemical potential of H- and T-type particles as solute concentration approaches zero, and illustrate the equivalence of the one-body potential and the excess chemical potential at this concentration.

Figure 5: Free surfactant density (circles) as a function of total surfactant density. Solid line is the 45° line corresponding to a condition of no micellar aggregates. Dashed line indicates the CMC as calculated by the method of von Gottberg *et al.*[22]. Left: H_1T_4; Right: H_1T_5.

Figure 6: Simulation snapshot of 216 H_1T_4 surfactant molecules in implicit solvent showing the formation of micellar aggregates. H-type particles are black, T-type particles are grey. This image shows three distinct micelles, two of which (on the left) are in close proximity; our clustering algorithm is able to distinguish such closely-spaced micelles from a single, elongated micelle.

Figure 7: Micelle aggregation number distribution as measured by a clustering algorithm (Circles: H_1T_4, Crosses: H_1T_5). Inset shows the behavior of the distribution for large micelle sizes.

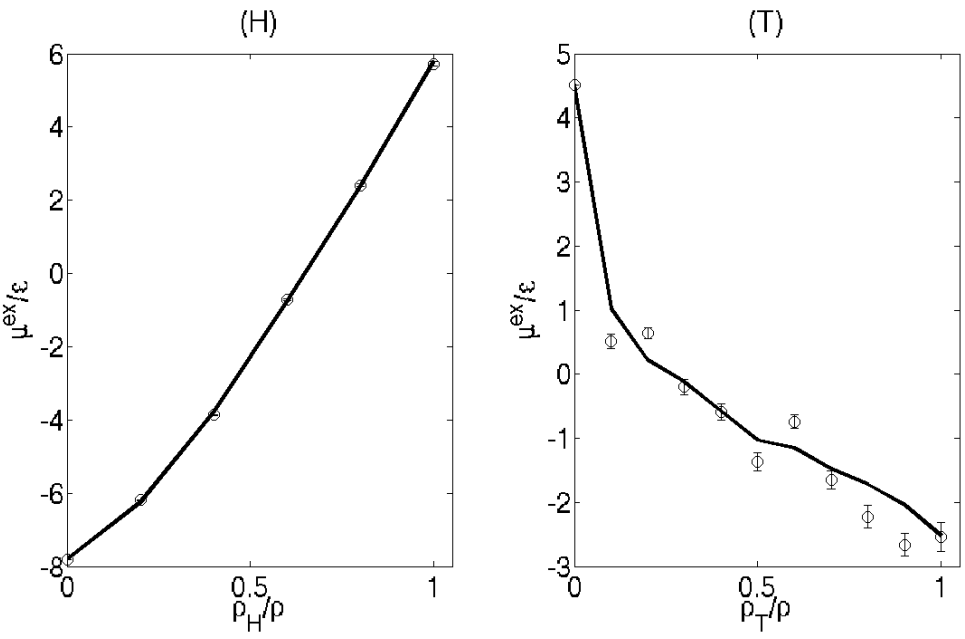

Figure 1

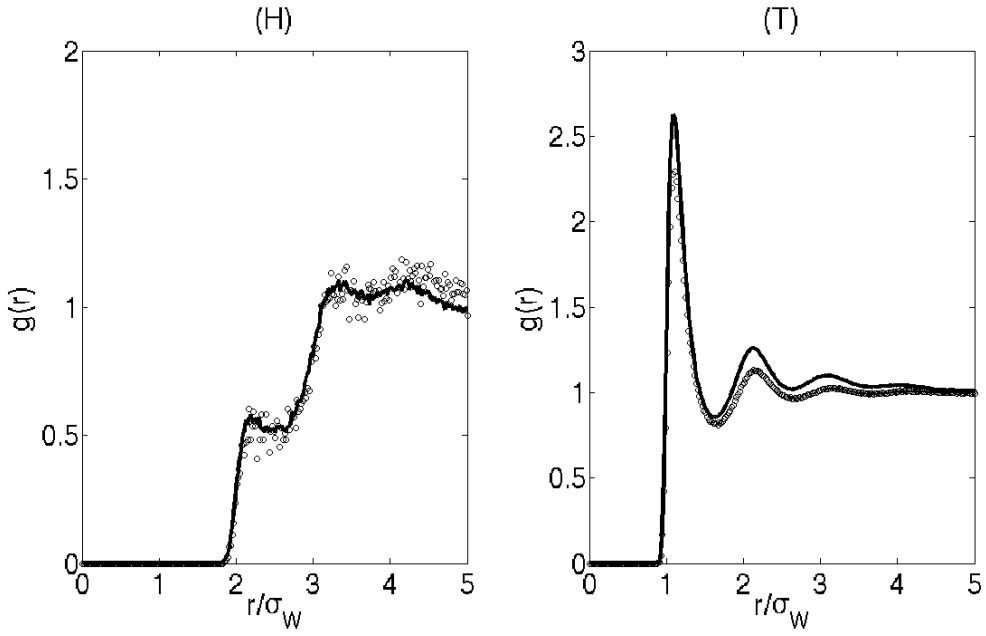

Figure 2

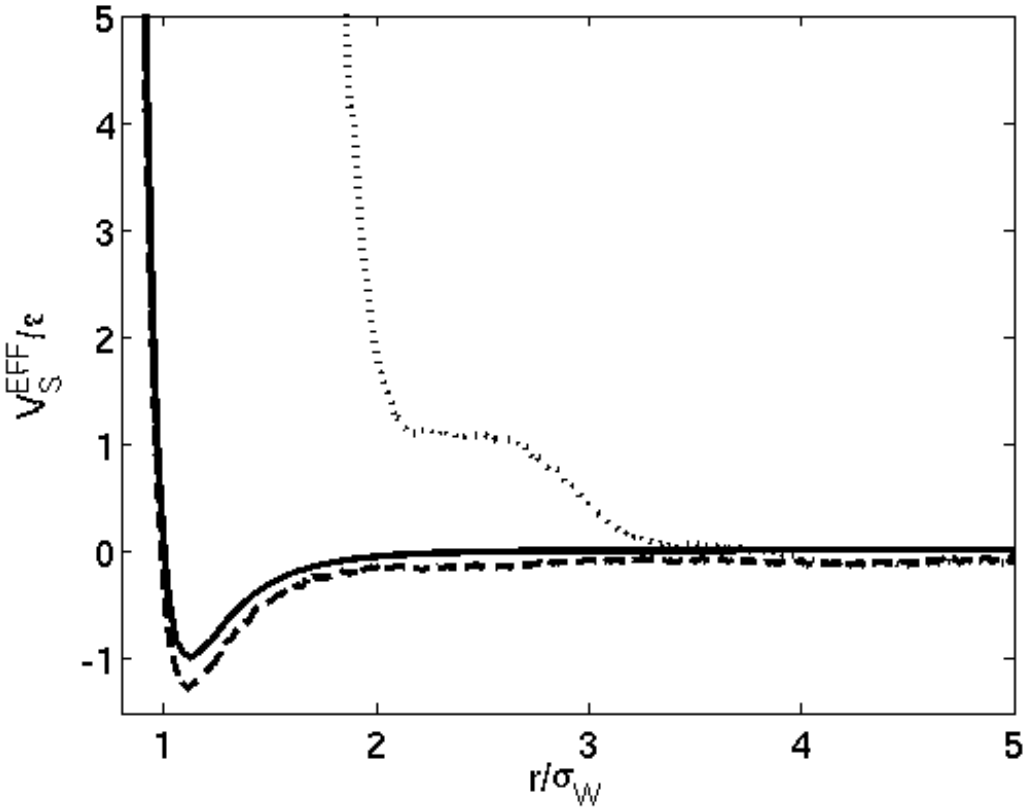

Figure 3

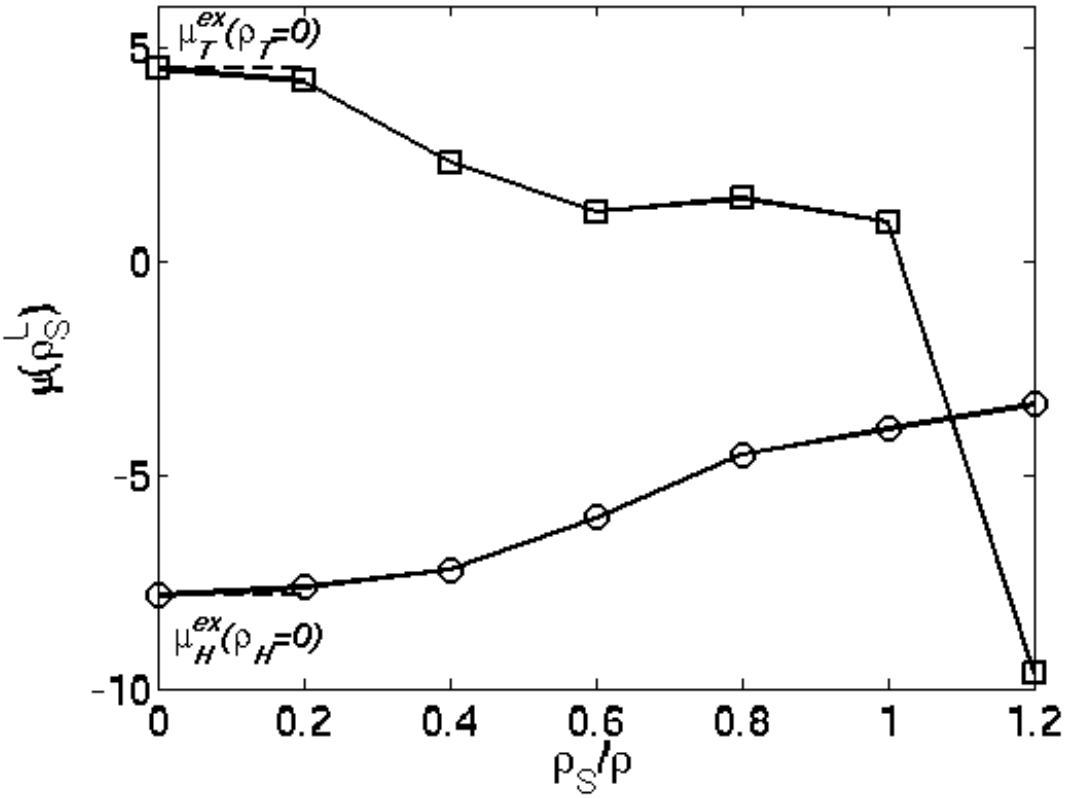
$\mu_T^{ex}(\rho_T=0)$
$\mu_H^{ex}(\rho_H=0)$
$\mu^L(\rho_S)$
ρ_S/ρ
5
0
-5
-10
0
0.2
0.4
0.6
0.8
1
1.2

Figure 4

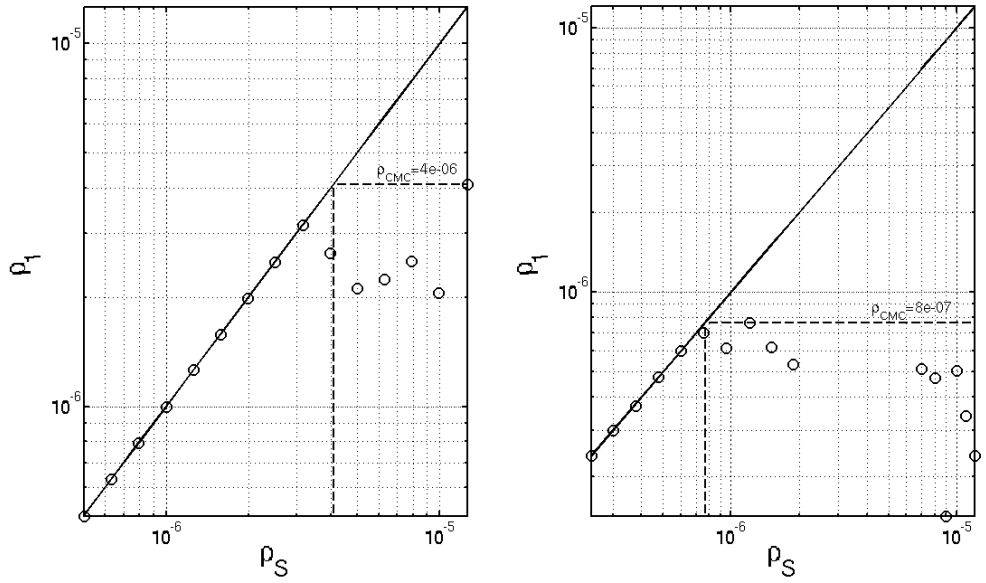
10^{-5}
10^{-6}
ρ_1
ρ_{CMC}=4e-06
10^{-6}
10^{-5}
ρ_S
10^{-5}
10^{-6}
ρ_1
ρ_{CMC}=8e-07
10^{-6}
10^{-5}
ρ_S

Figure 5

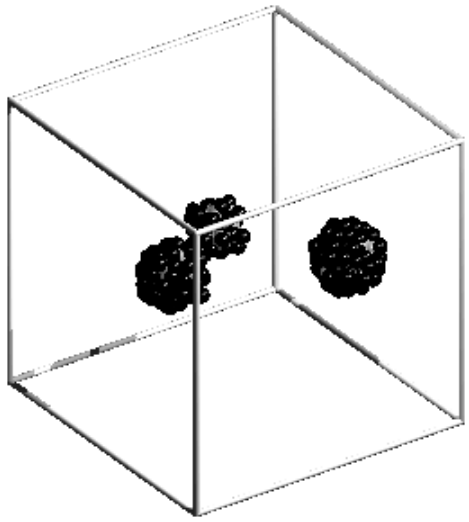

Figure 6

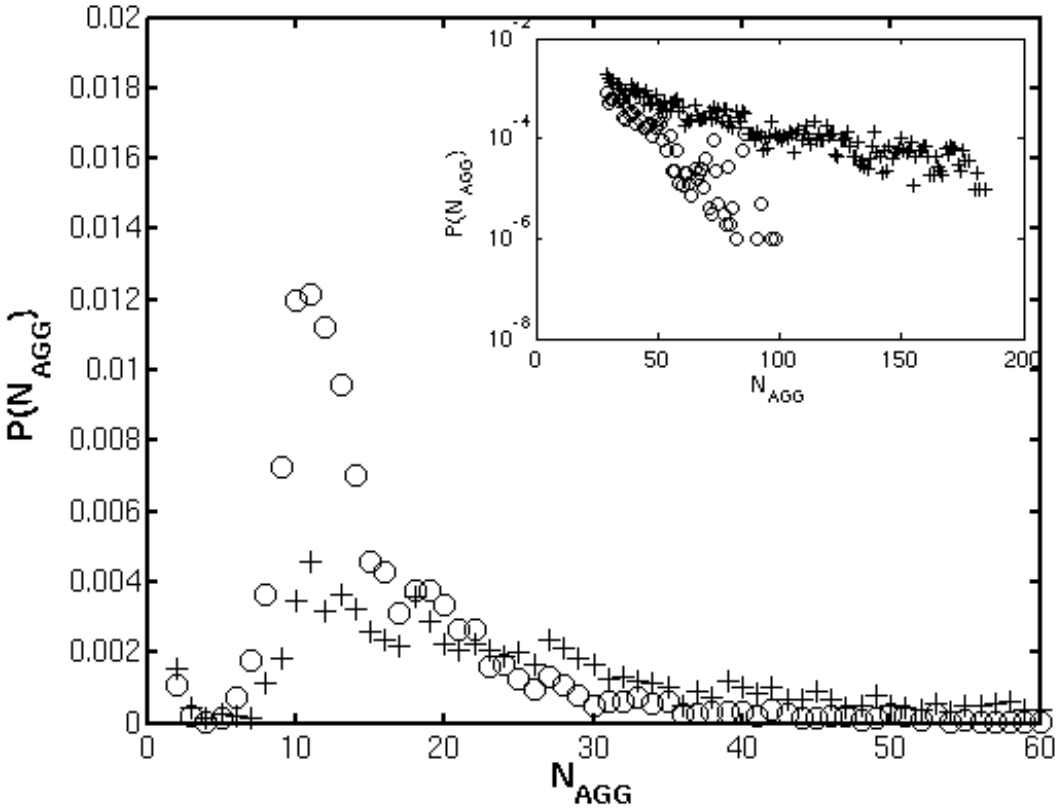
0.02
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
P(N_AGG)
0
10
20
30
40
50
60
N_AGG
10^-2
10^-4
10^-6
10^-8
P(N_AGG)
0
50
100
150
200
N_AGG

Figure 7