

This is the author's final, peer-reviewed manuscript as accepted for publication (AAM). The version presented here may differ from the published version, or version of record, available through the publisher's website. This version does not track changes, errata, or withdrawals on the publisher's site.

## Toward a Standard Protocol for Micelle Simulation

MA Johnston, WC Swope, KE Jordan, PB Warren,  
MG Noro, DJ Bray & RL Anderson

### Published version information

**Citation:** "Towards a standard protocol for Micelle simulation." MA Johnston et al. Journal of Physical Chemistry B, Vol. 120, Pt. 26 (2016): 6337-6351.  
doi: [10.1021/acs.jpcb.6b03075](https://doi.org/10.1021/acs.jpcb.6b03075)

This document is the Accepted Manuscript version of a Published Work that appeared in final form in The Journal of Physical Chemistry B, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see:

<http://pubs.acs.org/doi/abs/10.1021/acs.jpcb.6b03075>

Please cite only the published version using the reference above.

This item was retrieved from **ePubs**, the Open Access archive of the Science and Technology Facilities Council, UK. Please contact [epubs@stfc.ac.uk](mailto:epubs@stfc.ac.uk) or go to <http://epubs.stfc.ac.uk/> for further information and policies.

## Toward a Standard Protocol for Micelle Simulation

Michael A. Johnston<sup>(a)</sup>, William C. Swope<sup>(b)\*</sup>, Kirk Jordan<sup>(c)</sup>, Patrick B. Warren<sup>(d)</sup>, and Massimo G. Noro<sup>(d)</sup>, David J. Bray<sup>(e)</sup>, Richard L. Anderson<sup>(e)\*</sup>.

(a) IBM Research Ireland, Dublin, Ireland

(b) IBM Almaden Research Center, San Jose, CA, USA

(c) IBM T.J. Watson Research, Cambridge, MA, USA

(d) Unilever R&D Port Sunlight, Quarry Road East, Bebington, Wirral, CH63 3JW, UK

(e) STFC Hartree Centre, SciTech Daresbury, Warrington, Cheshire WA4 4AD, UK

\* Corresponding Authors: email [richard.anderson@stfc.ac.uk](mailto:richard.anderson@stfc.ac.uk) : tel+44 (0) 1925 603386: email [swope@us.ibm.com](mailto:swope@us.ibm.com) : tel +1-408-927-1706

## **Abstract**

In this paper we present protocols for simulating micelles using dissipative particle dynamics (and in principle molecular dynamics) that we expect to be appropriate for computing micelle properties for a wide range of surfactant molecules. The protocols address challenges in equilibrating and sampling, specifically when kinetics can be very different with changes in surfactant concentration, and with even minor changes in molecular size and structure, even using the same force field parameters. We demonstrate that detection of equilibrium can be automated and is robust, for the molecules in this study and others we have considered. In order to quantify the degree of sampling obtained during simulations, metrics to assess the degree of molecular exchange among micellar material are presented, and the use of correlation times are prescribed to assess sampling and for statistical uncertainty estimates on the relevant simulation observables. We show that the computational challenges facing the measurement of the critical micelle concentration (CMC) are somewhat different for high and low CMC materials. While a specific choice is not recommended here, we demonstrate that various methods give values that are consistent in terms of trends, even if not numerically equivalent.

## Introduction

The formation of micelles in aqueous surfactant solutions is an essential component of many industrial processes. Micelles can be used to facilitate drug delivery, act as friction modifiers and are important in the cleaning behaviour of home and personal care goods (e.g., shampoos, laundry products).<sup>1-8</sup> Micelles even play a significant role in the human body where they are crucial in our ability absorb fat soluble vitamins.<sup>9</sup> A wide variety of experimental techniques, including nuclear magnetic resonance, electron paramagnetic resonance, light scattering and small angle neutron diffraction have be used to monitor the formation, structure and composition of micelles. Self-assembly of surfactants into micelles occurs only when their concentration exceeds the critical micelle concentration (CMC) of the surfactant. The CMC of a surfactant is a crucial property and an important metric for a wide variety of industrial applications involving surfactant adsorption at interfaces, such as foams, emulsions, suspensions, and surface coatings. The CMC can be identified experimentally from a discontinuity or inflection point in plots of some physical properties of the solution as a function of surfactant concentration. Figure 1 demonstrates the changes that occur in a solution at the CMC for sodium dodecyl sulfate (SDS), a common anionic surfactant.<sup>10</sup> The CMC of a surfactant varies with temperature, pressure and often due to the presence of other surface-active molecules. It can be determined using a variety of techniques, e.g., tensiometry, conductometry, and fluorescence spectroscopy.<sup>11</sup> The CMC does not represent a very sharply defined point at which all observables exhibit a transition, a characteristic of so-called first order phase transitions. Consequently, experimental values may vary depending on the measurement technique adopted.

Molecular simulation offers a complementary method to experiment and phenomenological theory in the study of micelles, allowing molecular level resolution, dynamics, and certain thermodynamic quantities to be determined. A number of techniques have been applied to the computational study of micelles over the past three decades ranging from physics-type models in which only the essential features are retained to atomistic models based on realistic potentials.<sup>12-19</sup> A number of excellent lattice-based models also exist.<sup>18-19</sup> Micelle formation is a difficult process to simulate with all-atom methods due to the time scales involved. Dissipative particle dynamics (DPD) has proven to be an effective method to address the length and time scales necessary for a number of important aspects of micelle formation and properties and is being employed in an increasing number of studies concerning micelles.<sup>20-24</sup> Recently Vishnyakov et al., have used DPD in the prediction of CMC values for model surfactants, and Lee et al., studied the effect of chain rigidity on the CMC of surfactant molecules.<sup>20, 21</sup> This work provides in our opinion an example of current best practice of CMC determination by DPD.

The overall ambition of our current work programme is to employ high performance computing and the growing body of experimental data to develop and optimize interaction parameters (force fields) so that models can be systematically improved and validated to the point they are useful for predictive, i.e., chemical engineering applications. This type of approach has been adopted with surprising effect in quantum chemistry as well as in small molecule force field construction (for example, in the Gaussian-n approaches and the OPLS force field family respectively).<sup>25,26</sup> Other notable examples that mesh well with our ambition are recent efforts to develop parameterisation methodology and software tools such as the Force Balance approach of Wang, Martinez and Pande.<sup>27</sup>

To support this ambition, this paper is concerned with the establishment of a protocol for micelle simulation and CMC determination that can be broadly and consistently applied on a wide range of surfactant materials. By broadly applicable we mean the protocol should work for a wide variety of molecules (e.g., large and small; charged/uncharged; branched/linear). Consistent application of our protocol is hoped to allow results from different research groups to be more easily combined (meta-studies) for community-based efforts. Our protocol consists of: methodology choices; specification of observables of interest; construction of starting coordinates; establishment of equilibration; determination of adequacy of sampling. As this is our first attempt at developing a protocol for micelle simulation, we have limited the study to non-ionic linear surfactant molecules. We will apply the protocol to charged surfactant molecules and those with more complex molecular structure in subsequent communications. The outlined protocol is a necessary first step to developing force field parameters that can be validated against experiment for a wide range of molecules. With the consistent development of new force fields comes the opportunity to study outstanding issues in micelle physics.

In the following Methodology section we survey various aspects of the protocols, and the rationale for specific choices made therein. In the Results section, we apply these protocols to three (idealized) molecules with relatively low, medium and high CMC values. Finally, in the Discussion section we summarize the key findings and recommendations.

## 2. Methodology

In this section we describe the elements of the micelle simulation protocols. These include the choice of clustering method, simulation observables, system size considerations, equilibration detection, sampling and uncertainty estimation, and determination of CMC. Finally, details of the simulation carried out in our study are presented. For the underpinning DPD methodology, we refer to the many detailed descriptions in the literature.<sup>28-30</sup> The DPD formalism used for this study follows that of Groot and Warren.<sup>29</sup>

***Clustering procedure.*** The description of a protocol that will support systematic improvement of force field parameters for surfactants begins with a discussion of how surfactant aggregates will be identified. The method chosen should be simple to implement with a minimal set of associated parameters, it should be based only on particle coordinates at a single point in time, rather than velocities or on trajectory history, it should be deterministic and unique in that a given set of particle coordinates should produce a single partitioning of material. Furthermore, to support the goal of generating improved force field parameters for a large variety of molecules, it should work equally well for many types of molecules, large or small, ionic or nonionic, linear or branched.

Most clustering algorithms are based on some measure of intermolecular distance. This may be based on distances between centers of mass or geometry, distances between tag sites (special sites within each molecule used for this purpose), distances between the closest pair of any sites between the two molecules, or distances between the closest pair of some subset of the sites (such as the solvophobic sites) in the two molecules. Lee et al., and Vishnyakov et al., assumed two molecules belonged to the same cluster if any two of their tail or middle beads overlapped, for example.<sup>20, 21</sup>

For small molecules, the differences among these approaches are usually superficial, since essentially all sites within a small molecule are close to its molecular center, but this will not be the case for larger molecules. Therefore, since the use of the molecular center of mass or geometry, or of a specific tag site is not extendable to larger molecules we only considered criteria based on all or a subset (solvophobic) of the molecular sites. A comparison of these two clustering approaches is shown in Figure 2. The left side of the figure shows the results of clustering based on all sites of each molecule, with two molecules considered to be in the same cluster if any pair of sites is sufficiently close. The right side shows the results of using the same particle coordinates, but based on considerations of just the solvophobic sites. The number of clusters is significantly larger and their size (number of constituent surfactant molecules) significantly smaller when only the solvophobic sites are used. Inspection shows that clustering based on the solvophobic material resolves the larger aggregates into smaller distinct micellar groups, each characterized by compact isolated regions of solvophobic material. This latter picture better supports the notion of a micelle and it allows characterization of the sizes and shapes of supra-micellar aggregates. This choice of clustering based on solvophobic versus all sites, of course, profoundly affects subsequent classification of material into micelles versus free monomers and small submicellar clusters of surfactant, as well as resulting estimates of micelle shape distributions, micelle mean aggregation number and CMC. Another parameter of the clustering algorithm is the site-based cutoff distance used to determine if two molecules are considered close enough to be in the same cluster. We have used the DPD interatomic interaction distance ( $r_c$ ) for this parameter, since it defines intermolecular contact in an



energetic sense and offers a natural distance scale. Small variations of the cluster cutoff distance parameter from this value have not produced noticeable changes in results.

***Simulation observables.*** A number of important observables can be derived after the clustering procedure has been performed on a molecular configuration produced during a simulation. These include the total number of clusters and cluster shape and size distributions. Clusters can be classified as either micellar or submicellar based on the number of surfactant molecules, using a surfactant number cutoff. The choice of this cluster size cutoff parameter is an important and controversial issue that will be addressed with examples later in this paper. Often there is a deep minimum, or even a gap, in the cluster size distribution. In this case the distinction between submicellar aggregates and actual micelles is obvious, and analysis results can be insensitive to the exact choice of the size parameter.

However, we have observed many cases where there is no such obvious minimum, and the choice of cutoff value could be surfactant concentration dependent, may depend on the particular surfactant molecule being simulated, and may depend on choice of force field parameters. The value of the cutoff can affect subsequent computation of the CMC since many metrics are based on the concentration of submicellar material, and can also affect the mean aggregation number since small clusters may be classified differently depending on the value of the size cutoff.

Once clusters have been classified as micellar, aggregates of these micelles can be identified and characterized with respect to their size and shape. Another useful metric that can be computed after clusters have been identified is what we term the cohabitation order parameter,  $Q_1$  (Equation 1 and Supporting Information, section A, for definition). This provides useful information about the cluster size distribution.

$$Q_1 = \frac{1}{2} \left( \frac{\langle s^2 \rangle}{\langle s \rangle} - 1 \right) \quad (1)$$

Note that  $\langle s \rangle$  is the mean aggregation number and  $\langle s^2 \rangle / \langle s \rangle$  is the weight-averaged aggregation number. For uniformly dispersed surfactant, all surfactant molecules are free, or exist in very small clusters, so the value of  $Q_1$  is small at the beginning of a simulation and grows as the clusters form. When larger clusters break apart, the value of  $Q_1$  decreases. Therefore,  $Q_1$  is a useful observable for monitoring the approach to equilibrium and its temporal behavior can be related to processes such as micelle growth, degradation and merging.

**System size.** There should be enough material in the simulations to characterize adequately all the important observables. For example, since the number of micellar clusters is important for characterizing both the CMC and the mean aggregation number, our experience is that simulations should have enough material to form at least five micelles at the total surfactant concentrations where micelles form. In our experience trial concentration scans are necessary, perhaps with a smaller simulation boxes, in order to determine the appropriate system size needed for production simulations.

**Equilibration.** Simulations are usually started with molecular configurations that may be far from equilibrium with respect to aggregation behavior, such as uniformly dispersed surfactant molecules. Over time, micellar aggregates will form and their numbers and sizes will change. For obtaining thermodynamic averages, however, data should not be used until the system has equilibrated. An automated way to signal when the molecular system has reached equilibrium is very important in situations where hundreds of simulations are being performed.

The clustering procedure can either be done on the fly as the simulation proceeds, or by post-processing coordinates saved in trajectory files. Either way, the time series of the observables of interest can be acquired. Figure 3 shows an example of this where the observable is the number of surfactant molecules in submicellar clusters. Since the simulation starts with uniformly dispersed surfactant this metric drops rapidly during the equilibration phase from an initial value of about 1300 and settles to fluctuate near values of about 100. We have found the following analysis protocol useful. First, time series data (collected every 500 DPD time steps) is organized into blocks of 10 samples each, and averages and standard deviations are calculated for each block. A sliding window of 20 such blocks is considered, and a weighted linear least squares (WLLS) fit is calculated using the 20 block averages with the weights set by the 20 block standard deviations. This procedure essentially fits a straight line to the observable as a function of time, over a sliding window of 200 samples (corresponding to 200 000 DPD simulation time steps). The error in the slope can be calculated from the standard deviations used in the WLLS fit. When the simulation starts, the system is far from equilibrium and many observables show a systematic trend. In this non-equilibrated regime, the slopes of the WLLS fits will be statistically significantly different from zero and of some uniform sign. However in an equilibrated system the observables show random fluctuations about constant mean values. In terms of the sliding window WLLS fits, the system is considered to be equilibrated when the slope are zero to within the statistical uncertainty, and remains that way for 20 consecutive blocks. This automated equilibration criterion obviously depends to some extent the choices made for the number of time steps per sampling period (500), the number of sampling periods per block (10), the number of blocks used in the sliding window (20), and the number of consecutive

blocks (20) needed for the system to be considered at equilibrium. Ideally, the sliding window should represent at least a correlation time in the amount of simulation data, but this is not usually known during the equilibration process, so these parameters must be refined heuristically. They are likely to be somewhat dependent on the molecule and thermodynamics conditions, but the values reported above work for the simulations described in this article.

***Sampling and uncertainty estimation.*** After the equilibration phase is complete, the time series data for the observables is accumulated and analyzed to compute averages and standard deviations. Since the data is temporally correlated, this correlation needs to be taken into account in estimating statistical uncertainty. Therefore, autocorrelation functions and their associated correlation times are computed and used to estimate standard deviations of the mean (Supporting Information, section D). For example, the following equations provide the average number,  $M_f$ , and statistical uncertainty,  $\delta(M_f)$ , in the number of surfactant molecules in a simulation with  $M$  total surfactant molecules:

$$M_f = M\bar{\chi} \quad (2)$$

$$\delta(M_f) = \left( M \overline{(\chi - \bar{\chi})^2} (2\tau/D) \right)^{1/2} \quad (3)$$

where  $\chi_{ij}$  is an indicator function which is one if molecule  $i$  is in a submicellar sized cluster at time  $t_j$  after the end of the equilibration phase of the simulation.  $\bar{\chi}$  is the average of this indicator function over all surfactant molecules and over the duration,  $D$ , of the simulation and gives the probability that a surfactant molecule is in a cluster of submicellar size. Also in this expression is the standard deviation of this indicator function and the correlation time,  $\tau$ , for this observable.  $D/2\tau$  is the effective number of uncorrelated samples. The correlation

time is computed as the following integral over the estimate of the normalized fluctuation correlation function:

$$\tau = \left( \frac{1}{2} + \sum_{l=1}^{N-1} \frac{N-l}{N} C_l \right) \Delta t \quad (4)$$

$$C_l = \frac{\langle (\chi_{ij} - \langle \chi \rangle) (\chi_{i(j+l)} - \langle \chi \rangle) \rangle}{\langle (\chi - \langle \chi \rangle)^2 \rangle} \approx \frac{1}{M} \sum_{k=1}^M \frac{1}{N-l} \sum_{m=0}^{N-1-l} \frac{(\chi_{km} - \bar{\chi})(\chi_{k(m+l)} - \bar{\chi})}{(\chi - \bar{\chi})^2} \quad (5)$$

where  $\Delta t$  is the time period between measurements of  $\chi$  and  $D=N\Delta t$ . An example of the result this process is shown in Figure 4.

In addition to being useful for estimating uncertainties in time averaged quantities, the correlation times for the various observables can also provide insight into the time scales for the underlying physical processes in micellar kinetics. The widely-accepted Aniansson-Wall step-wise association model predicts two time scales,  $\tau_1$  and  $\tau_2$  (in real systems these are typically of the order  $\mu s$  and  $ms$  respectively).<sup>31-33</sup> The fast  $\tau_1$ -process corresponds to monomer-micelle exchange kinetics and the slow  $\tau_2$ -process corresponds to complete micellar turnover. The separation of time scales arises because in stepwise association monomer-micelle exchange events do not change the total number of micelles. Therefore, one has to wait for the fluctuations to take the micelle aggregation number through the minimum in the aggregation number distribution, which is a slow Kramers-like barrier-crossing process. Alternatively, on the time scale of the  $\tau_1$ -process, one might also observe micellar fusion or fission events which also relax the number density.<sup>34</sup> Our simulations are faithful to these kinetics, and all these processes can be observed in principle.

Metrics for assessing the degree to which surfactant material reorganizes among micellar clusters are described in the Supporting Information, section A. A particularly useful metric is the normalized cohabitation correlation function, defined as follows:

$$q(|t - t'|) = \frac{\sum_{i>j=1}^M \theta_{ij}(t)\theta_{ij}(t')}{\sum_{i>j=1}^M \theta_{ij}(t)} \quad (6)$$

where  $\theta_{ij}$  is a cohabitation function that is one if surfactant molecules  $i$  and  $j$  are in the same cluster and zero if they are not. The correlation function gives the conditional probability that two surfactant molecules are in the same cluster at time  $t'$  given that they were in the same cluster at time  $t$ .

For equilibration and thorough sampling it is desirable to simulate long enough to observe all the physical processes that might occur in actual micellar systems. However, this may not be possible and, in fact, it may not even be necessary to obtain an equilibrium distribution of micellar size and shape characteristics. But one should always endeavour to access the  $\tau_2$ -processes which allow the micelle number density to equilibrate, otherwise there is a danger of sampling bias. A simple way to monitor this is to track the total number of micelles.

***Estimating the critical micelle concentration.*** Micellization is a “pseudo” phase transition rather than a true first order phase transition (see Figure 1).<sup>34,35</sup> Therefore, different experiments that monitor and measure different observables might suggest different transition points for the onset of micelle formation.<sup>11</sup> Correspondingly, different observables measured from simulations may also suggest different transition characteristics. Moreover, simulations are performed on relatively small systems, which can result in broadening of transitions that are expected to be true phase transitions in the

thermodynamic limit. These issues complicate the extraction of micellization properties from simulation.

An idealized view of micellization is depicted in Figure 5. Consider the gradual addition of surfactant molecules to solvent. At very low concentration, there is no aggregation of surfactant. In this regime, the free surfactant concentration is equal to the total surfactant concentration. At this low concentration, there are no micelles. With increasing total surfactant concentration, at some point (the critical micelle concentration) micelles begin to form. As more surfactant molecules are added to the mixture, they do not cause an increase in the free surfactant concentration, but, rather, cause an increase in the amount of material in micelles. The micelles in this idealized scenario have a relatively narrow size distribution determined by molecular properties of the surfactant and solvent. Therefore, any additional surfactant causes the production of more micelles rather than merely larger ones, resulting in a linear increase in the micellar population with total surfactant concentration, but no increase in the mean micelle size or in the free surfactant concentration. In this scenario, surfactant molecules are either free (monomers or very small clusters) or in very much larger micelles, and the onset of micelle formation is sharp. In actual simulations, however, there is often a broad distribution of surfactant cluster sizes and it can be difficult to determine at which size they should be called micelles. Even for the micelles, there may be a broad distribution of sizes. The free surfactant, or, rather that which exists in submicellar-sized clusters, may not be constant with increasing total surfactant concentrations. Since a finite size simulation box can contain only an integer number of micelles, there may be a free energy barrier for the formation of micelles from submicellar material requiring that the total surfactant concentration must exceed the CMC before the

first ones are observed. Finally, although we are primarily concerned with the formation of spherical micelles, there are other types of micelles that develop at higher total surfactant concentrations, such as rod-like and worm-like micelles. The transitions forming these types of micelles can be very close to that for formation of the spherical ones, further complicating the situation.

Nonetheless, the idealized behavior of micellization (Figure 5) suggests different ways to extract the CMC. Once a minimum micelle size is established, one may determine bounds on the CMC by looking for the lowest concentration that has at least one micelle and the highest concentration that has none. On the assumption that the free surfactant concentration is constant for total concentrations above the CMC, one may also measure the submicellar surfactant concentration at total concentrations where micelles are observed, or look for the highest submicellar concentration observed near the lowest total concentration that presents micelles. One may perform linear fits of the submicellar concentration as a function of the total surfactant concentration above and below an estimate of the CMC and detect a crossing point. With any of these approaches based on submicellar concentration, even the computation of this concentration may not be straightforward. Santos and Panagiotopolous (2016) have shown that in computing the submicellar concentration as a surrogate for the CMC, one must correct the volume considered to be accessible to the free surfactant by a factor somewhat more than one would expect based on the space occupied by the micellar material.<sup>19</sup>

The idealized micelle behavior also suggests one may perform a linear fit of the total number of micelles as a function of total surfactant concentration and extrapolate this fit to find the zero intercept, representing the highest total surfactant concentration where no



micelles are seen. Some research has looked for a discontinuous response in some observable as a function of surfactant concentration. Santos and Panagiatopolous (2016) have performed grand canonical ensemble simulations (constant chemical potential, volume and temperature) and looked for a discontinuous change in the slope of the internal pressure as a function of surfactant concentration.<sup>19</sup> One may also look for changes in the partial molar volume of the surfactant in isobaric (constant particle number, pressure and temperature) simulations as one increases the total surfactant concentration from below to above the CMC. The partial molar volume of the surfactant can be calculated from a linear fit of the excess volume per solvent bead,  $V_{ex}/N_s$ , to the surfactant composition,  $x$ , expressed as the ratio of the number of surfactant beads to the number of solvent beads:

$$\frac{V_{ex}}{N_s} = \left(\frac{v_m}{n} - v\right)x + (v_s - v) \quad (7)$$

where  $v_m$  and  $v_s$  are the partial volumes of surfactant molecules and solvent beads, respectively, and  $v_s - v$  is the difference between the volume of a solvent bead in the mixture and that in pure solvent, which is very close to zero for dilute solutions.  $n$  is the number of beads per surfactant molecule. The slope of such a linear fit provides  $v_m$ . A detailed explanation concerning the calculation of the partial molar volume is given in the Supporting Information, section C.

Both of the approaches of looking for changes in the behaviour of the pressure (in constant volume simulations) or in volume (in constant pressure simulations) with respect to surfactant concentration are based on the fact that the volume occupied by a surfactant molecule is different if it is surrounded by solvent than if it is part of a micelle. Another approach of this sort is to compute the surfactant chemical potential, which underpins many

of the phenomenological measures shown in Figure 1 (for example, adsorption isotherms and interfacial tensions). As a function of surfactant concentration, the chemical potential of the surfactant is expected to rise until the CMC is reached, then level off (or at least increase much less rapidly) past the CMC.<sup>10</sup> Methods that seek to measure a small change in pressure, molar volume or chemical potential are challenging since the change being measured may be quite small and high precision is required. They may have an advantage over methods that measure surfactant cluster characteristics, however, in that they do not depend on the somewhat arbitrary choices of clustering algorithm, clustering distance parameter, or the cluster size parameter that separates micellar from submicellar clusters.

Many approaches for extracting the CMC rely on the need to perform simulations at extremely low surfactant loadings. This may present practical problems, since there are likely to be very small numbers of micelles, or they may be transient, making precise measurements difficult. To help address this, an alternative operational definition for the CMC is that for which half the surfactant molecule population is in micelles and the other half is in submicellar clusters. Measures of the CMC using this definition are roughly a factor of two higher than those that use a definition based on the first observation of micelles.

Most of the approaches outlined above require a specification of the size a cluster of surfactant must be to be considered to be large enough to be called a micelle. Results of most analyses will be relatively insensitive to this parameter, but only if the cluster size distribution has a deep minimum (or a gap) between the submicellar sizes and the micellar sizes. This often is not the case. In fact, depending on the total surfactant loading, there may not even be a minimum suggesting a separation threshold. We have found that at sufficiently high loadings a minimum can usually be found.

**Target concentration range.** The approaches outlined above for characterizing the formation of spherical micelles call for the need to perform a series of simulations over a range of surfactant concentrations, which span the CMC. The CMC is usually not known in advance, but can be identified from examination of the metrics discussed above. One must find a range of concentrations with predominantly spherical micelles, and recognize when concentrations become so high that tube like micelles and/or wormlike micelles are being produced. Ideally, a concentration range can be found where the number of spherical micelles grows linearly with surfactant concentration but their size is constant (or growing in an expected way, as discussed in Supporting Information, section B). One must also be studying systems large enough to have an adequate amount of surfactant material to allow observation of enough micelles to measure their properties with sufficient precision and to believe that they full-sized micelles representative of equilibrium.

**Model Details.** To test the protocol, three different idealized non-ionic surfactant molecules have been considered in the work presented in this article; H6T6, H4T4 and H4T3 (H=Hydrophilic Head, T=Hydrophobic Tail). In our coarse grained model each H bead can be thought of as representing an ethylene oxide bead ( $\text{CH}_2\text{CH}_2\text{O}$ ), for example, and each tail bead, T, two alkyl groups ( $\text{H}=\text{CH}_2\text{CH}_2$ ). In this sense our model molecules are analogous to the surfactants of the  $\text{C}_n\text{EO}_m$  family. Model surfactant beads are held together by harmonic bonds of the form  $U_S(r) = 0.5\kappa(r-r_0)^2$ . A single bond length of  $r_0 = 0.5 r_c$  was adopted for simplicity with  $\kappa = 50.0 k_B T$ . Vishnyakov et al., have recently shown the importance of chain rigidity upon the CMC of micelles.<sup>20</sup> A harmonic angular potential between pairs of bonds was adopted with form  $U_B(\theta) = 0.5\kappa(\theta-\theta_0)^2$ . An equilibrium angle  $\theta_0 = 180^\circ$  and  $\kappa = 5.0 k_B T$  was adopted. Table I shows molecular volumes calculated from measured properties, for the

relevant components in the hydrocarbon / ethylene oxide / water system. The molecular volume is given by  $v_m = M_W / \rho N_A$ , where  $N_A \approx 6.02 \times 10^{23} \text{ mol}^{-1}$  is Avogadro's number. The headgroup 'EO' component is used to define the bead volume  $v_b$ . Since the bead density in reduced units is  $\rho = 3$ , this means  $r_c^3 \approx 3 \times 65 \text{ \AA}^3$ , or  $r_c \approx 5.8 \text{ \AA}$ . This defines the length scale of our DPD simulations. Table I therefore suggests that one DPD bead can represent 1 H or EO group, or 2.2 water molecules, or  $12/5.8 \approx 2.1$  carbon atoms in a hydrocarbon chain (T Beads). Hence the choice to use a chain of 6 beads the alkyl half of a  $C_{12}EO_6$  or H6T6 model surfactant.

In this study we have aimed to develop a surfactant representation and parameterization strategy that gives rise to a diverse range of micelle behavior upon which to test our protocol. Broadly speaking we have tried to sample low CMC, medium CMC and final high CMC ranges. We have not explicitly attempted to match experimental values; rather we have created models against which our analysis and characterization methods can be tested. The adopted model parameters were as follows. We followed the common DPD community convention and set the conservative repulsion parameters between beads of the same type, i.e.,  $a_{II}$ , as 25 for all species. This choice of self-interaction parameter should strictly only apply for solvent beads representing a single  $H_2O$  molecule, for which it reproduces the compressibility of water. However, this value is frequently adopted for a number of different bead types comprising different chemical species. We recognize this as unsatisfactory and ultimately our aim is to resolve this by the application of our protocol to improve interaction parameters to be reported in subsequent communications. For unlike beads, parameters were chosen to yield a close match to the CMC of  $C_{12}EO_6$  for our H6T6

model. The resulting conservative interactions for unlike species are therefore;  $a_{WH}=25$ ,  $a_{WT}=45$  and  $a_{HT}=30$ .

Cubic simulation cell sizes of initial volume  $64000 r_c^3$  were adopted (unless otherwise stated, e.g., for H4T4 – see Results and Discussion) which corresponds to an edge length of  $\approx 23$  nm. Using a reduced density of  $\rho = 3$  the simulated system volume corresponds to a system size of 192 000 beads per simulation. A DPD time step of 0.04 was adopted and trajectory data was collected every 500 time steps. Simulations were run for  $2 \times 10^6$  time frames. A reduced DPD temperature of  $1 k_B T$  has been adopted throughout. We report on simulations carried out under constraint of constant pressure as this best replicates experimental conditions. NPT simulations were carried out using the Langevin piston approach derived by Jakobsen for DPD simulations.<sup>36, 37</sup> The DPD code contained within the DL\_MESO simulation package was used to perform all simulations.<sup>38</sup>

### 3. Results and Discussion

**Equilibration.** Simulations were performed on the three molecules at approximately ten different concentrations, ranging, in general, from below to well above the CMC. All systems were started with uniformly dispersed surfactant and various observables were computed and used to monitor the approach to equilibrium. Completion of the equilibration phase of the simulation was determined using the protocol described in Methodology (section 2.2). Figure 6 shows the equilibration time determined by the protocol for each surfactant molecule and at each concentration. The equilibration times in this figure are based on the analysis of only one observable, namely, the number of surfactant molecules in submicellar clusters, but the equilibration condition can be made more stringent by including other observables. Although no protocol can be expected to work for all molecules and force fields that might be studied, for the systems in this particular study, inspection shows this protocol to be reasonable and sufficient for determining when in the simulation the equilibration is sufficiently complete.

**System Size.** Exploratory simulations were performed for each molecule at several concentrations using a system size with a volume of approximately  $64000 r_c^3$  and 192000 DPD particles. From the lowest concentrations where micelles were observed up through to the highest concentrations at which the number of micelles showed linear growth as a function of total surfactant concentration, we noted whether there were enough micelle-sized clusters and submicellar surfactant molecules present to meet our criteria for an adequate system size. These conditions were easily met for H4T3. For H4T4, however, this starting system size produced on average only 1.5 micellar sized clusters over the concentration range of interest, causing concern that there was not enough surfactant

material to produce micelles of full equilibrium size. Therefore, for H4T4, larger systems were generated and used for the remainder of the study. These had sizes 1.5 times larger in each linear dimension, with a volume of approximately  $216000 r_c^3$  and 648000 DPD particles, and met our system size criteria for the number of micelles produced. For H6T6, the smaller systems produced four to six micelles at all times over the 2-5% total surfactant concentration range of interest, which was very close to our target of at least 5 micelles. However, over this range there was also clear linear behavior in the number of micelles versus total surfactant concentration and frequent micelle merging and splitting events were observed, suggesting we were seeing equilibrium behavior even with this small system size. Therefore, the smaller system size was used for both H4T3 and H6T6.

***Micellar Properties.*** Figures 7-9 convey the key results of the study. We see in these three molecular systems a range of behaviors. The concentration of surfactant in submicellar clusters (Figures 7(a), 8(a) and 9(a)) reaches some approximate asymptotic values of approximately 0.6%, 0.18% and 0.003%, for H4T3, H4T4 and H6T6, respectively, at large total surfactant concentrations suggesting three different ranges for their CMCs (high, medium and low). The H4T3 and H4T4 submicellar contents increase from small values near zero total surfactant to a maximum, then show a drop with increasing total surfactant concentration. H6T6 actually shows a drop rather than a rise, and then nearly constant concentration with increasing total surfactant. In all three cases, an approach that looks simply at the submicellar surfactant concentration as an estimate of the CMC is problematic for various reasons. For these molecules, one would obtain different estimates from the use of 1) the submicellar concentration at large total concentration limits, 2) the maximum in the submicellar concentration, 3) the total concentration at which there is a change in the slope

of the submicellar concentration, and 4) crossing points of linear fits of the submicellar concentration using estimates of the premicellar and postmicellar total concentration regimes, 5) the total concentration where approximately half of the surfactant is in micellar clusters. However, it is clear from these figures, that there are important ranges of interest for each molecule where changes of behavior are apparent, usually in the 1.5-4% total surfactant region, near where the 50% micellar criterion is reached, and also very near to the change over from one type of linear behavior to another.

The number of micellar sized clusters (Figures 7(b), 8(b) and 9(b)) shows linear behavior for these molecules near to and slightly above these values, thereby defining the target concentration range. For H4T3, this linear regime is clear and ranges from approximately 3% to 6% total surfactant. For H4T4, the range is from approximately 1.75% to 2.5%; and for H6T6, it is from 2% to about 4%, however somewhat less clear. The figures also show a linear fit to a subset of these points for H4T3 and H4T4. In addition, the same figures show the mean size of the micellar-sized clusters (mean aggregation number). In the regions selected for the linear fit to the micelle count data, the micelle size also shows linear growth with a positive slope.

Although the idealized view of micelle formation in Figure 5 suggests the micelle size (mean aggregation number) should be unchanging, elementary theoretical considerations (Supporting Information, section B) suggest that in fact there should be growth in the mean micelle size with increasing total surfactant concentration, with the rate of increase governed by the width (variance) of the micelle size distribution. As the size distribution narrows, the rate of growth in micelle size is less. The rate of growth in the mean micelle size seen in the simulations is consistent with predictions of the theory. In fact, the rate of growth in this



property expressed as  $d(\ln N)/d(\ln(\text{Concentration}))$  is observed to be 0.22 for H4T3 and 0.28 for H4T4. The values predicted from the theory in Supporting Information, section B, for these slopes are 0.19 and 0.35, respectively.

It is clear that micelle count and size data outside of the target region where linear behavior is observed are falling above or below the linear fit. At the low concentration end, the surfactant concentration may not be high enough to result in full sized micelles, and at the higher end, a different type of micelle might be forming, perhaps signaling a transition from spherical to rod-like shapes. In fact, when rod shaped micelles form, upon increase in the total surfactant concentration, the free energy cost of extending them to greater length may be less than that of forming additional spherical micelles. So, one might expect to see slower linear or even lack of growth in the number of micelles, but linear growth in their size. This seems to occur in H4T3 for total surfactant above 6% (linear growth in micelle size and number), and in H4T4 above 3% (linear growth in size but with constant numbers). For H6T6 this may be occurring at about 4%-5%, but the trend is not so obvious perhaps due to greater uncertainty in the data.

Cluster size distributions for H4T3 and H4T4 (Figures 7(c) and 8(c)) show a sharp peak at small sizes due to the submicellar clusters, a minimum separating submicellar from micellar clusters, then a micelle peak near cluster sizes of about 50 (for H4T3) and 70 (for H4T4). As total concentration is increased, the micellar distributions are seen to broaden and shift to larger sizes, consistent with the cluster number and size data. As suggested by the theory described above, these broader size distributions are consistent with a greater slope in micelle size with increase in surfactant concentration. Tails in this distribution are also consistent with the presence of rod shaped micelles and aggregates of micelles that might not

be resolvable by our cluster identification algorithm. For H6T6 (Figure 9(c)) the distribution is extremely noisy and consists of spikes.

The structure in these distribution functions (especially for H4T4) suggests the presence of aggregates of micelles with surfactant counts that are multiples of the basic micellar unit. For example, with micelles of approximately 75 surfactant molecules (at low total surfactant concentration), there are peaks in the distribution function also at about 150, 220 and 310, corresponding to dimers, trimers and tetramers. Initially we thought these peaks were due to static structures, maybe a manifestation of insufficient equilibration or sampling, and that the structure in the distribution functions would eventually disappear with increased simulation time. For some molecular systems this might be the case. However, these peaks in the distribution are often seen at multiple different concentrations. And for H4T4, visual inspection indicated that micelle sized clusters were aggregating for a short time, then breaking up and diffusing apart, with sometimes-significant exchange of surfactant material during their encounter. These frequent encounters provide a way of equilibrating the size and number of micelles, helping to produce rather narrower size distributions, and also the consequent peaks in the distribution function at multiples of this size. A set of 14 independent simulations using different starting configurations was performed on H4T4 at 1.75% total surfactant to produce a more precise distribution function, shown in Figure 10. One can see that a great deal of structure appears in this distribution, with peaks near 150, 220, 270 and 350. Since peaks at these sizes appear in the distribution functions at other concentrations, they are probably indications of these supra-micellar clusters rather than simply noisy data.

To further investigate the size and shape distribution, the radius of gyration ( $R_g$ ) of all the micelle-sized clusters was measured and it was seen that there were distinct groupings present when the micelle counts were resolved along these two dimensions. Figure 11 shows the micelle size distributions for H4T3 at 3.5% and 6% total surfactant resolved into contributions from micelles with  $R_g < 2$  and  $R_g \geq 2$ . Smaller micelles contribute to a roughly Gaussian shaped peak, whereas larger micelles produce a distribution more skewed to larger sizes. Analysis of the moments of inertia suggests the smaller ones to be primarily spherical or oblate ellipsoids, and the larger ones to be rod shaped prolate ellipsoids. These different profiles can also be seen in the work of Nelson et al.<sup>39</sup>

One potential view is that the large clusters are actually supra-micellar aggregates, and their presence indicates the micelles may be 'sticky' toward one another (i.e., perhaps there is a short-ranged attraction in the inter-micelle interaction potential of mean force). There is some precedence for this in real nonionic micellar systems.<sup>40</sup> Obviously, mischaracterization of supra-micellar aggregates as giant micelles can skew some metrics, such as the mean aggregation number or the micelle number density. Other metrics, such as the total amount of material in micelles, are insensitive to the formation of supra-micellar aggregates. Some care must therefore be taken.

**Kinetics.** The  $q(t)$  cohabitation correlation function (Figure 12, Equation 6) gives an opportunity to analyze the kinetics of the micellar systems studied. This function gives the conditional probability that two molecules are in the same cluster at time  $t$  given that they were in the same cluster at time 0. Rapid decay in this function indicates fast exchange of surfactant material through emission/absorption of submicellar material, or through merging and splitting of micellar aggregates. Figure 13 shows correlation times for two

important observables for each of the three molecules as a function of surfactant concentration. The observables are the number of micelles and number of surfactant molecules in submicellar clusters. Outliers in the correlation times should be noted, especially anomalous decreases relative to those at nearby concentrations, since they signal potential sampling problems. Simulations that are too short will tend to underestimate both the standard deviation and the correlation time, contributing to an underestimate of the uncertainty in the corresponding observable.

The equilibration times, the cohabitation metric and the correlation times are mutually consistent in suggesting that H4T3 has the fastest kinetics and H6T6 the slowest, with H4T4 intermediate. This range of kinetics is also consistent with the amount of noise seen in the distribution functions (Figures 7(c), 8(c) and 9(c)) for these molecules.

***Estimation of the CMC.*** An analysis of the partial molar volume of the surfactant was performed for H4T3. If the volume occupied by a surfactant molecule is different when it is surrounded by solvent (as in free molecules or in submicellar sized clusters) than when it is surrounded by other surfactant molecules (micellar), the partial molar volume of the surfactant molecules should change as one passes from below to above the CMC. Details of this analysis are provided in the Supporting Information, section C, and the result for H4T3 is shown in Figure 14. The figure shows the excess volume per solvent bead as a function of the concentration of surfactant, expressed as the ratio of the numbers of surfactant beads to solvent beads. Two linear regimes are apparent, one with concentrations less than 0.6%, and another with concentrations greater than 0.7%. The two linear fits cross at a concentration of 0.60%, providing an estimate of the CMC. The slopes of these fits give estimates of the volume per surfactant bead relative to the volume of solvent beads in the pure solvent under

the same thermodynamic conditions. We note that below the CMC the slope in the fit is positive, indicating that the surfactant occupies more volume than the equivalent number solvent beads in a pure solvent, but above the CMC, slope is negative and the volume occupied is less. From the slopes of these fits one obtains a surfactant molecular volume of 2.52 (units of  $r_c^3$ ) below the CMC and 2.24 above it.

The decreased volume of surfactant in micellar clusters, relative to that in free and submicellar clusters is unusual. Moreover, one expects the partial molar volume of surfactant in micelles to be similar to that in pure surfactant, which is usually less dense than water, implying a volume that would be greater than that of the solvent, not less, as suggested here. We believe this is a consequence of the force field parameters adopted in these toy models, specifically the self-interaction parameters being set equal. This combined with the relatively short bond-length adopted in our model results in a dense micelle core. There are important consequences for parameterization here; if one attempts to reproduce the end-to-end length of a surfactant monomer at the coarse graining-level adopted (2 water molecules per bead and therefore  $r_c = 5.8 \text{ \AA}$ ) resulting bond lengths will be of the order used in this work. We would recommend that the DPD community begins to embrace self-interaction parameters that are not all set equal (often to 25.0) in order to reproduce the different densities of the various components simulated.

The molar volume analysis is appropriate in the context of NPT simulation methods, where the volume of the system can change in response to changes in composition of the material, and is the counterpart to the observation of a change in the pressure with composition that is seen in the context of constant volume methods.<sup>19</sup> Use of this method for estimating the CMC requires simulations at a couple of very low concentrations (below the

CMC) in order to support a linear fit and long simulations to obtain sufficiently precise measures of the volume.

Another measure of the CMC can be obtained from extrapolation of the linear fit of the number of micelles seen as a function of the total concentration seen in the target concentration range. The x-intercept of this line would be the total concentration at which no micelles are seen but where any increase should produce micelles. However, applying this procedure yields unphysical negative values for the CMCs of H4T3, H4T4 and H6T6. Alternatively, as seen in Figure 15, one may perform a linear fit of the total concentration of surfactant molecules seen in micellar sized clusters as a function of the total surfactant concentration and extrapolate this fit to determine an x-intercept. This is a very stable way to estimate that CMC, and this approach yields values of  $0.65 \pm 0.01\%$  for H4T3,  $0.118 \pm 0.004\%$  for H4T4 and  $0.0019 \pm 0.0007\%$  for H6T6.

A common way to compute the CMC is based on the measure of the free or submicellar surfactant concentration when the total surfactant concentration is above the CMC. Micelles must be present to use this measure. As described by Santos and Panagiotopolous if one wishes to use this approach, one should perform a correction for the volume occupied by the micellar surfactant.<sup>19</sup> When computing the submicellar surfactant concentration, the correction essentially removes volume from the simulation cell that is overlapping or associated with existing micelles and is, therefore, inaccessible to the submicellar material. This has the effect of increasing the concentration of surfactant by an amount that is greater with increasing micellar concentration. Figure 16 shows the effect of this type of correction on the concentration of surfactant in submicellar clusters as a function of total surfactant concentration. The corrections shown in Figure 16 were based on an

estimate of the total volume of all the surfactant, rather than just that due to micelles. The rationale for this is the same as used in derivations of the van der Waals equation of state for weakly attracting soft spheres. For these systems the volume accessible to gas particles is less than the container volume by that occupied by the gas particles themselves. Figure 16 shows that this correction improves the behavior of the concentration of surfactant in submicellar clusters, making it more constant with total surfactant concentration.

After this correction an estimate of the CMC can be obtained by performing a weighted average of this concentration over the concentration range studied. This yields  $0.727 \pm 0.008\%$ ,  $0.118 \pm 0.002\%$ ,  $0.0027 \pm 0.0006\%$  for H4T3, H4T4 and H6T6, respectively. Another estimate of the CMC may be obtained by using the average of the volume corrected submicellar surfactant concentration over the smaller target range. This estimate gives values of  $0.71 \pm 0.01\%$ ,  $0.119 \pm 0.003\%$ ,  $0.0021 \pm 0.0005\%$  respectively for the three systems. One may also use the maximum in the volume corrected submicellar surfactant concentration, and this yields values of  $0.751 \pm 0.005\%$ ,  $0.124 \pm 0.005\%$  and  $0.0044 \pm 0.0008\%$  for the three molecules. Finally, using the total concentration at which half the surfactant material is in micellar clusters gives CMC values of  $1.56 \pm 0.07\%$ ,  $0.336 \pm 0.005\%$  and  $0.13 \pm 0.08\%$  for the three molecules. As mentioned earlier, CMC values evaluated this last way are generally about twice as large as those measured by the other methods.

In general the value of the CMCs reported from our simulations are in-line with the experimental values for C6E4 ( $\sim 0.8\%$ ), C8E4 ( $\sim 0.2\%$ ) and C12E6 ( $\sim 0.003\%$ ) to which our toy models are loosely linked. Values taken from a typical surfactant supplier at [www.anatrace.com](http://www.anatrace.com).

***Estimate of the mean aggregation number.*** As described in the Supporting Information, section B, and as observed in these simulations, the average micelle size depends on the total surfactant concentration unless the width of the micellar cluster size distribution is very narrow. We feel that a reasonable value to report for this metric is the mean micelle size averaged over the concentrations that make up the target concentration range, and report an uncertainty that is half the size range. Over this range, the concentrations are high enough above the CMC to observe at least a small ensemble of micelles with sizes representative of equilibrium, but low enough that the population consists primarily of spherical micelles rather than much larger rods and wormlike micelles. Using this approach, one obtains mean aggregation numbers of  $55\pm 3$ ,  $119\pm 4$ ,  $87\pm 10$  for H4T3, H4T4, and H6T6, respectively.

#### **4. Conclusion**

We have reported here a suite of protocols for simulating micelles using dissipative particle dynamics that we expect to be appropriate for computing micelle properties for a wide range of surfactant molecules. In future work we intend to apply these protocols consistently to a number of systems for the purposes of force field validation and parameter optimization in order to develop a set of models that can be useful for subsequent predictions.

The protocols address challenges in equilibrating and sampling, specifically when kinetics can be very different with changes in surfactant concentration, and with even minor changes in molecular size and structure, even using the same force field parameters. Detection of equilibrium can be automated and is robust for the molecules of this study and others we have considered. In order to quantify the degree of sampling obtained during



simulations, metrics to assess the degree of molecular exchange among micellar material are presented, and the use of correlation times are prescribed to assess sampling and for statistical uncertainty estimates on the relevant simulation observables.

Computational challenges are somewhat different for high and low CMC materials. For low CMC material, one may not observe enough free surfactant to characterize precisely its concentration, requiring potentially larger and longer simulations. For high CMC material, one may not observe enough micelles to get a precise count for computing the mean aggregation number, requiring larger and longer simulations for these cases as well.

The focus in this work has been to characterize the formation and properties of spherical micelles. For this one must identify a rather narrow target range in the total surfactant concentration. Below this range there may be no micelles to observe, and above it one may be including a host of other transitions to different micelles shapes and perhaps even surfactant mesophases, which will complicate or obscure the extraction of the spherical micellar properties of interest.

Evidence from this works suggests that if the force field parameters are to be transferable, i.e., intended to be used in the context of a wide variety of molecule types and sizes, the fitting or training set needs to be similarly broad, since properties have been shown to be very sensitive to molecule size. Even in going from H4T3 to H6T6 there are large differences in both thermodynamic and kinetic properties.

There are challenges to measuring the CMC from simulation, with different reasonable approaches giving different values. While a specific choice is not recommended here, it is shown that various methods give values that are consistent in terms of trends, even if not numerically equivalent. Therefore, for force field tuning and material design application, it

may be adequate simply to pick an approach for estimating the CMC that is numerically robust and easy to compute and then use it consistently.

The large-scale simulations reported in this article are ‘un-steered’ in the sense that we simply observe what happens over a long period of time from a random start. This is ideal for automated scans and parameter refinement. Obviously, there is an important role to be played also by steered computational methods such as, *inter alia*, umbrella sampling, and forward-flux sampling.<sup>41-44</sup> For example our un-steered simulations reveal the presence of supra-micellar aggregates in the H4T4 system, but umbrella sampling could be used to investigate the inter-micelle interaction in this system and thereby confirm whether a short-range attraction (stickiness) is driving the formation of these aggregates. We note that the presence of supra-micellar aggregates raises the interesting possibility that they facilitate the exchange of surfactants between micelles. This is a kinetic process not envisaged in the commonly accepted Aniansson-Wall model. These and other open questions are left for future work.

## **Acknowledgements**

We thank the STFC Hartree Centre for supporting this effort by allowing human resource and computational infrastructure to be dedicated to this work. We thank Andrea Ferrante for his insightful discussions at an early part of the protocol development and also Ed Pyzer-Knapp for his help in the paper-writing phase. This work has been partially funded by Innovate UK project 101712 (Massimo Noro, Patrick Warren, David Bray and Richard Anderson) and the authors are grateful to the members of the UK Computer Aided Formulation (CAF) consortium project for their input.

## **Associated Content**

### **Supporting Information**

Description and derivation of key equations; metrics to assess sampling; relationships for concentration dependence of mean micelle size; procedure for computation of partial molar volumes of surfactant; equations used for computing statistical uncertainty estimates of time series data.

## References

- (1) Ahmad, Z.; Shah, A.; Siddiq, M.; Kraatz, H. Polymeric Micelles as Drug Delivery Vehicles. *RSC Advances* **2014**, *4*, 17028-17038.
- (2) Kedar, U.; Phutane, P.; Shidhaye, S.; Kadam, V. Advances In Polymeric Micelles for Drug Delivery and Tumor Targeting. *Nanomedicine* **2010**, *6*, 714-729.
- (3) Kwon, G.; Okano, T. Polymeric Micelles As New Drug Carriers. *Adv. Drug Deliver. Rev.* **1996**, *21*, 107-116.
- (4) Yokoyama, M. Polymeric Micelles as Drug Carriers: Their Lights and Shadows. *J. Drug Target.* **2014**, *22*, 576-583.
- (5) Zheng, R.; Liu, G.; Devlin, M.; Hux, K.; Jao, T. Friction Reduction of Lubricant Base Oil by Micelles and Crosslinked Micelles of Block Copolymers. *Tribol. T.* **2009**, *53*, 97-107.
- (6) Spikes, H. Friction Modifier Additives. *Tribol. Lett.* **2015**, *60*, 5.
- (7) Zana, R.; Kaler, E. *Giant Micelles*; CRC Press: Boca Raton, **2007**.
- (8) Attwood, D. The Effect of Electrolyte on the Micellar Properties of an Anionic-Nonionic Detergent in Aqueous Solution. *Kolloid Z. Z. Polym.* **1969**, *232*, 788-792.
- (9) Thompson, G. Absorption of Fat-Soluble Vitamins and Sterols. *J. Clin. Pathol.* **1971**, *5*, 85-89.
- (10) Presto, W.; Preston, W. Some Correlating Principles of Detergent Action. *J. Phys. Chem.* **1948**, *52*, 84-97.
- (11) Chakraborty, T.; Chakraborty, I.; Ghosh, S. The Methods of Determination of Critical Micellar Concentrations of the Amphiphilic Systems in Aqueous Medium. *Arab. J. Chem.* **2011**, *4*, 265-270.

- (12) Smit, B.; Hilbers, P.; Esselink, K.; Rupert, L.; van Os, N.; Schlijper, A. Computer Simulations of a Water/Oil Interface in the Presence Of Micelles. *Nature* **1990**, *348*, 624-625.
- (13) Jury, S.; Bladon, P.; Cates, M.; Krishna, S.; Hagen, M.; Ruddock, N.; Warren, P. Simulation of Amphiphilic Mesophases Using Dissipative Particle Dynamics. *Phys. Chem. Chem. Phys.* **1999**, *1*, 2051-2056.
- (14) Prinsen, P.; Warren, P.; Michels, M. Mesoscale Simulations of Surfactant Dissolution and Mesophase Formation. *Phys. Rev. Lett.* **2002**, *89*.
- (15) Tang, X.; Koenig, P.; Larson, R. Molecular Dynamics Simulations of Sodium Dodecyl Sulfate Micelles in Water—The Effect Of The Force field. *J. Phys. Chem. B* **2014**, *118*, 3864-3880.
- (16) Roussel, G.; Michaux, C.; Perpète, E. Multiscale Molecular Dynamics Simulations of Sodium Dodecyl Sulfate Micelles: From Coarse-Grained to All-Atom Resolution. *J. Mol. Model.* **2014**, *20*, 1-8.
- (17) Chun, B.; Choi, J.; Jang, S. Molecular Dynamics Simulation Study of Sodium Dodecyl Sulfate Micelle: Water Penetration and Sodium Dodecyl Sulfate Dissociation. *Colloid. Surface. A* **2015**, *474*, 36-43.
- (18) Floriano, M.; Caponetti, E.; Panagiotopoulos, A. Micellization In Model Surfactant Systems. *Langmuir* **1999**, *15*, 3143-3151.
- (19) Santos, A.; Panagiotopoulos, A. Determination of the Critical Micelle Concentration in Simulations of Surfactant Systems. *J. Chem. Phys.* **2016**, *144*, 044709.
- (20) Lee, M.; Vishnyakov, A.; Neimark, A. Calculations of Critical Micelle Concentration by Dissipative Particle Dynamics Simulations: The Role of Chain Rigidity. *J. Phys. Chem. B* **2013**, *117*, 10304-10310.

- (21) Vishnyakov, A.; Lee, M.; Neimark, A. Prediction of the Critical Micelle Concentration of Nonionic Surfactants by Dissipative Particle Dynamics Simulations. *J. Phys. Chem. Lett.* **2013**, *4*, 797-802.
- (22) Dhakal, S.; Sureshkumar, R. Topology, Length Scales, and Energetics of Surfactant Micelles. *J. Chem. Phys.* **2015**, *143*, 024905.
- (23) Dhakal, S.; Sureshkumar, R. Uniaxial Extension of Surfactant Micelles: Counterion Mediated Chain Stiffening and a Mechanism of Rupture by Flow-Induced Energy Redistribution. *ACS Macro Lett.* **2016**, *5*, 108-111.
- (24) Guo, X.; Zhang, L.; Wu, Z.; Qian, Y. Dissipative Particle Dynamics Studies on Microstructure of Ph-Sensitive Micelles for Sustained Drug Delivery. *Macromolecules* **2010**, *43*, 7839-7844.
- (25) Curtiss, L.; Redfern, P.; Raghavachari, K. Gaussian-4 Theory. *J. Chem. Phys.* **2007**, *126*, 084108.
- (26) Jorgensen, W.; Maxwell, D.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.
- (27) Wang, L.; Martinez, T.; Pande, V. Building Force fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885-1891.
- (28) Hoogerbrugge, P.; Koelman, J. Simulating Microscopic Hydrodynamic Phenomena with Dissipative Particle Dynamics. *Europhys. Lett.* **1992**, *19*, 155-160.
- (29) Groot, R.; Warren, P. Dissipative Particle Dynamics: Bridging the Gap Between Atomistic and Mesoscopic Simulation. *J. Chem. Phys.* **1997**, *107*, 4423.

- (30) Moeendarbary, E.; NG, T.; Zangeneh, M. Dissipative Particle Dynamics: Introduction, Methodology and Complex Fluid Applications — A Review. *Int. J. Appl. Mechanics* **2009**, *01*, 737-763.
- (31) Aniansson, E.; Wall, S. Kinetics of Step-Wise Micelle Association. Correction and Improvement. *J. Phys. Chem.* **1975**, *79*, 857-858.
- (32) Kale, K.; Zana, R. Effect of the Nature of the Counterion on the Volume Change Upon Micellization of Ionic Detergents in Aqueous Solutions. *J. Colloid Interf. Sci.* **1977**, *61*, 312-322.
- (33) Patist, A.; Oh, S.; Leung, R.; Shah, D. Kinetics Of Micellization: Its Significance to Technological Processes. *Colloid. Surface. A* **2001**, *176*, 3-16.
- (34) Pool, R.; Bolhuis, P. Prediction of an Autocatalytic Replication Mechanism for Micelle Formation. *Phys. Rev. Lett.* **2006**, *97*.
- (35) Shinoda, K.; Hutchinson, E. Pseudo-Phase Separation Model for Thermodynamic Calculations on Micellar Solutions 1. *J. Phys. Chem.-US* **1962**, *66*, 577-582.
- (36) Jakobsen, A. Constant-Pressure and Constant-Surface Tension Simulations in Dissipative Particle Dynamics. *J. Chem. Phys.* **2005**, *122*, 124901.
- (37) Jakobsen, A. Erratum: "Constant-Pressure and Constant-Surface Tension Simulations in Dissipative Particle Dynamics *J. Chem. Phys.* **2006**, *125*, 029901.
- (38) Seaton, M.; Anderson, R.; Metz, S.; Smith, W. DL\_MESO: Highly Scalable Mesoscale Simulations. *Mol. Simulat.* **2013**, *39*, 796-821.
- (39) Nelson, P.; Rutledge, G.; Hatton, T. On the Size and Shape of Self-Assembled Micelles. *J. Chem. Phys.* **1997**, *107*, 10777-10781.

- (40) Mallamace, F.; Gambadauro, P.; Micali, N.; Tartaglia, P.; Liao, C.; Chen, S. Kinetic Glass Transition In a Micellar System with Short-Range Attractive Interaction. *Phys. Rev. Lett.* **2000**, *84*, 5431-5434.
- (41) Torrie, G.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comp. Phys.* **1977**, *23*, 187-199.
- (42) Pool, R.; Bolhuis, P. Sampling the Kinetic Pathways of a Micelle Fusion and Fission Transition. *J. Chem. Phys.* **2007**, *126*, 244703.
- (43) Dellago, C.; Bolhuis, P.; Csajka, F.; Chandler, D. Transition Path Sampling and the Calculation of Rate Constants. *J. Chem. Phys.* **1998**, *108*, 1964-1977.
- (44) Allen, R.; Warren, P.; ten Wolde, P. Sampling Rare Switching Events in Biochemical Networks. *Phys. Rev. Lett.* **2005**, *94*, 018104.

## Tables

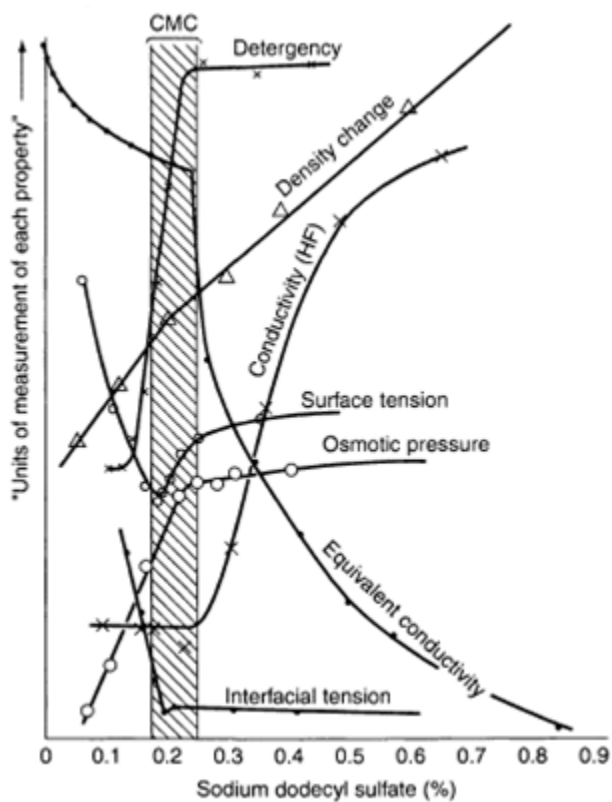
**Table I:** Molecular properties of ethylene oxide, water, and a representative hydrocarbon.

The last column gives the molecular volume in terms of the DPD bead volume.

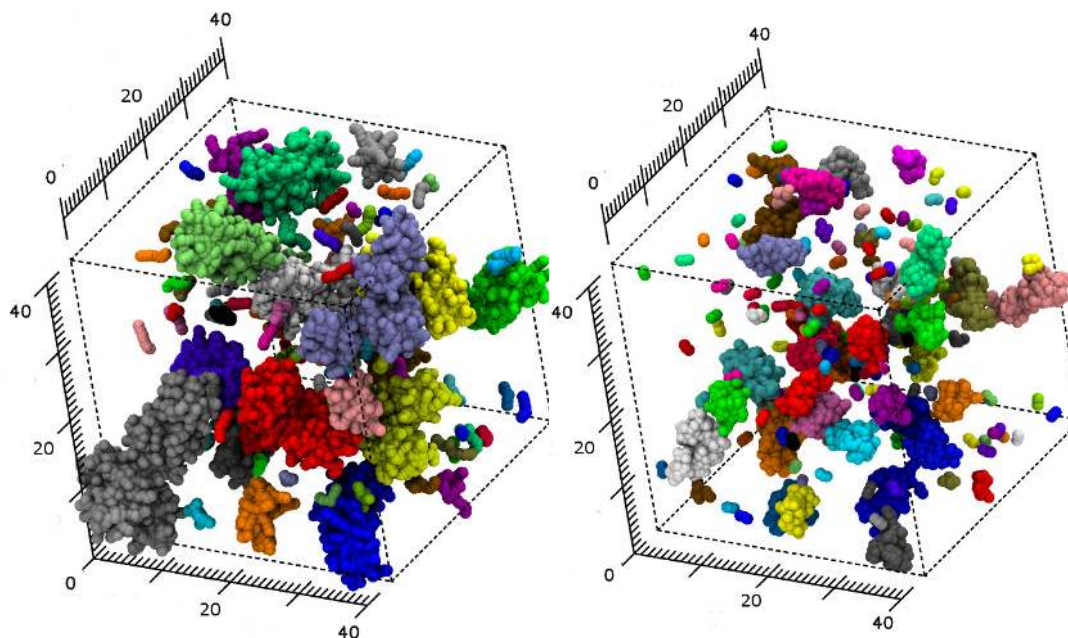
	$M_W$ (kg mol <sup>-1</sup> )	$\rho_{expt}$ (kg m <sup>-3</sup> )	$v_m$ (Å <sup>3</sup> )	$v_m/v_b$
CH <sub>2</sub> CH <sub>2</sub> O	0.044	1125	65	=1
H <sub>2</sub> O	0.018	998	30	2.2
C <sub>12</sub> H <sub>26</sub>	0.170	750	376	5.8



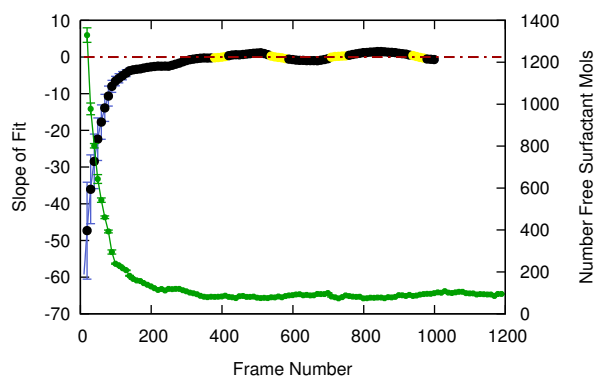
## Figure & Captions



**Figure 1** Changes in a number of physical properties for an aqueous solution of sodium dodecyl sulfate as the concentration of solution is increased from below to above the CMC (reprinted with permission from American Chemical Society - Preston, W. C, Some correlating principles of detergent action, Presto, W.; Preston, W. Some Correlating Principles Of Detergent Action. *J. Phys. Chem.* 1948, 52, 84-97.)<sup>10</sup>

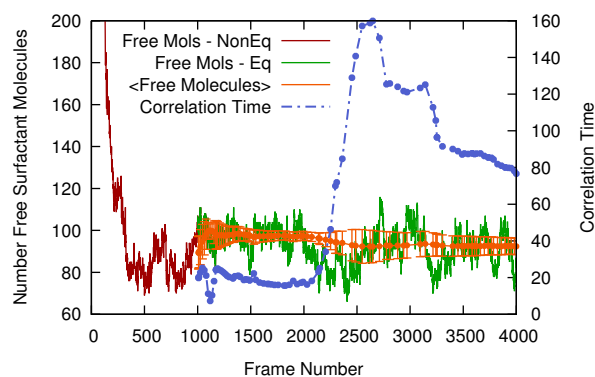


**Figure 2** Left: clustering based on all sites of each molecule (111 clusters) and all sites of the surfactant molecules are shown; Right: clustering based on solvophobic sites of each molecule (167 clusters) and only the solvophobic sites of the molecules are shown. The molecular system was H4T3 with an 8% surfactant concentration.

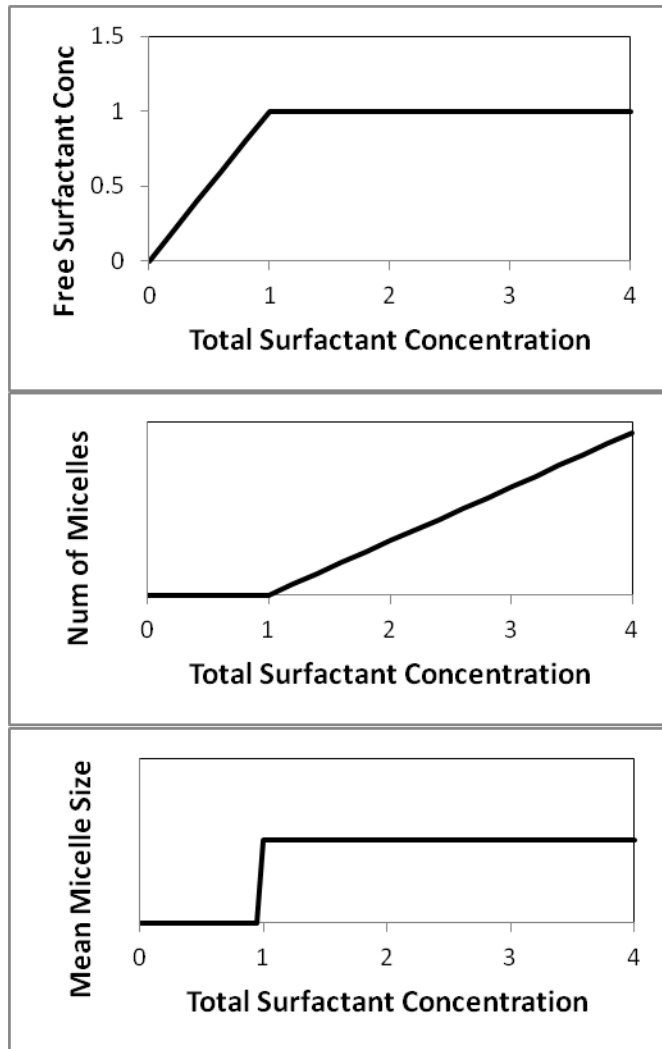


**Figure 3** Approach to equilibrium and detection of equilibration for H4T4 at 2% concentration. The observable being monitored in this case is the number of submicellar

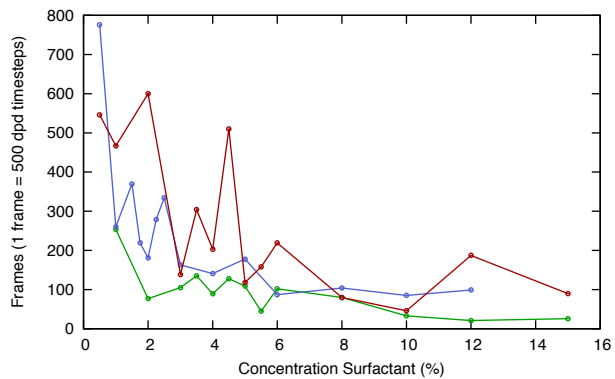
surfactant molecules (green, right vertical axis) plotted as a function of DPD simulation time. Block averages over 10 analysis frames are shown by green circles. The slope of the linear fits to sliding windows of 20 blocks is shown with black circles (left vertical axis) with the uncertainty in the slope indicated with blue error bars. Times at which this slope is within its uncertainty of zero are indicated in yellow. The time at which the slope of the fit *first* becomes zero (i.e., within its uncertainty of zero), and remains so for 20 consecutive frames is considered the end of the equilibration phase and sampling commences from this time forward for the computation of equilibrium averages. In this example, this happens approximately at frame 384.



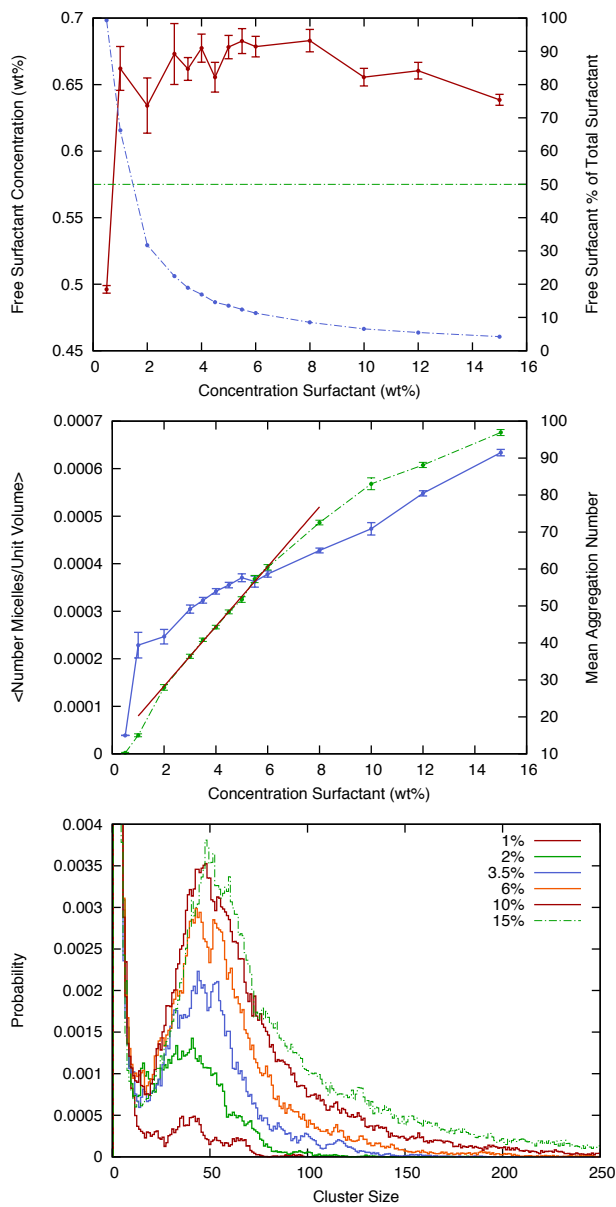
**Figure 4** Number of free surfactant molecules observed during a simulation of H4T4 with 2% concentration. Red region of the curve represents the equilibration phase; the green region represents results that are considered to be in equilibrium. The estimated correlation time in sample periods for this observable is shown with blue dots (right vertical axis). The cumulative average is shown in orange, with statistical uncertainty estimates based on the current estimates of the average, standard deviation and correlation time. As the simulation proceeds, the estimate of the correlation time improves along with the reliability of the uncertainty estimates.



**Figure 5** Idealized behavior of micellar systems as a function of total surfactant concentration.

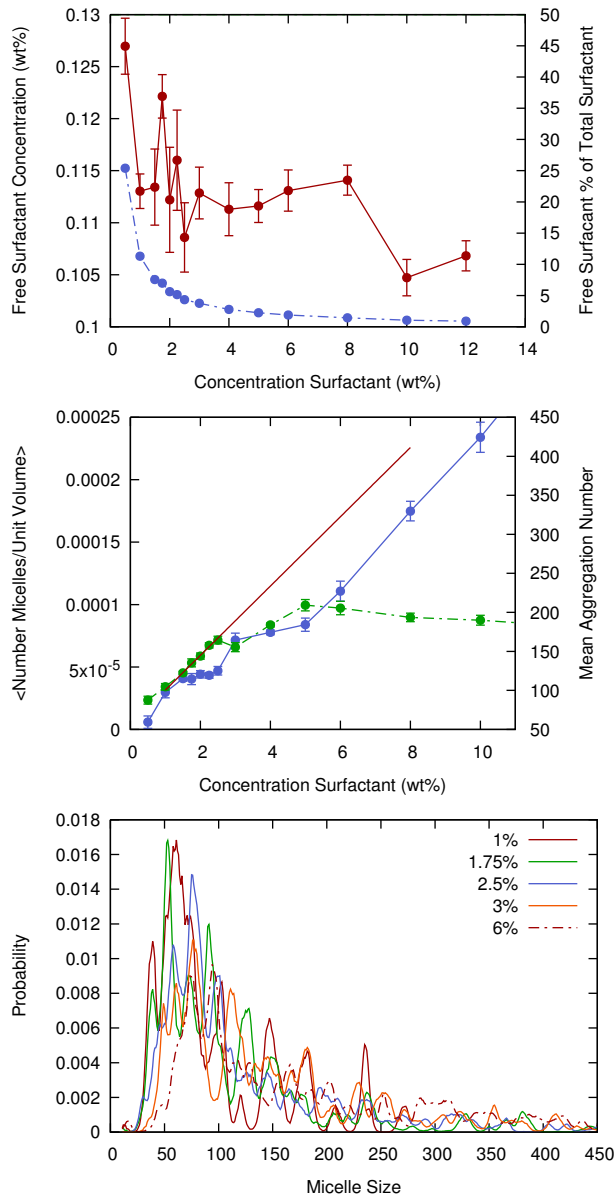


**Figure 6** Equilibration times based on obtaining stable values for the number of surfactant molecules in submicellar clusters. The green, blue and red lines are for H4T3, H4T4 and H6T6 respectively.



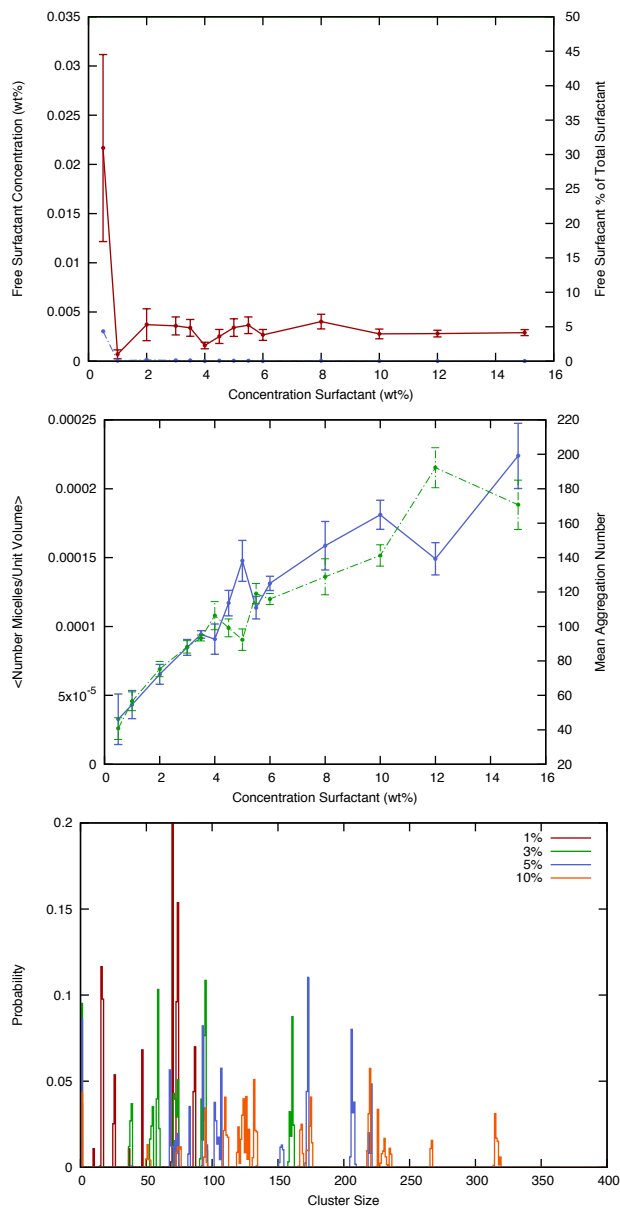
**Figure 7** a) (top) Concentration of surfactant molecules in submicellar clusters as a function of total surfactant concentration for H4T3 (red - number of free surfactant, blue - free surfactant %). b) (middle) Number of micelles and mean aggregation number as a function of

total surfactant concentration for H4T3 (blue - mean aggregation number, green - average number of micelles per unit volume, red - linear fit). c) (bottom) Size distributions for H4T3 at various total surfactant concentrations. The distributions are the fraction of the total number of clusters seen of various sizes. The area under each curve is normalized to unity.



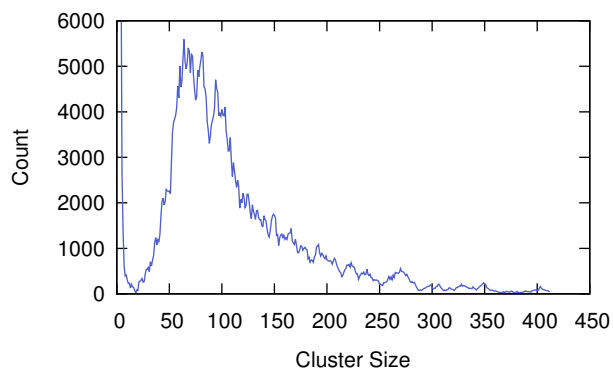
**Figure 8** a) (top) Concentration of surfactant molecules in submicellar clusters as a function of total surfactant concentration for H4T4 (red - number of free surfactant, blue - free surfactant %). b) (middle) Number of micelles and mean aggregation number as a function of total surfactant concentration for H4T4 (blue - mean aggregation number, green - average number of micelles per unit volume, red - linear fit). c) (bottom) Size distributions for H4T4 at various total surfactant concentrations. The distributions are the fraction of the total number of clusters seen of various sizes. The area under each curve is normalized to unity. The bottom image has been truncated at maximum aggregate size of 450 the distribution for larger cluster sizes. H4T4 exhibits a long tail due to the presence of extremely large clusters (possibly supra-micellar aggregates) observed at higher concentrations (See Supporting Information).



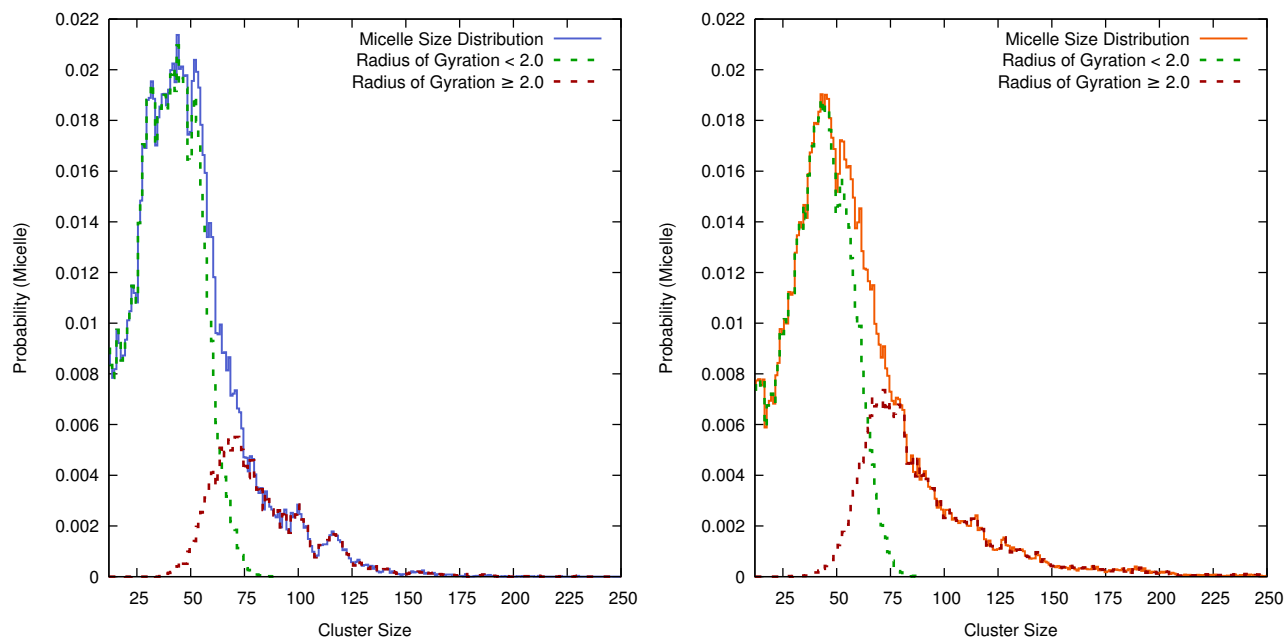


**Figure 9** a) (top) Concentration of surfactant molecules in submicellar clusters as a function of total surfactant concentration for H6T6 (red - number of free surfactant, blue - free surfactant %). b) (middle) Number of micelles and mean aggregation number as a function of total surfactant concentration for H6T6 (blue - mean aggregation number, green - average number of micelles per unit volume, red - linear fit). c) (bottom) Size distributions for H6T6

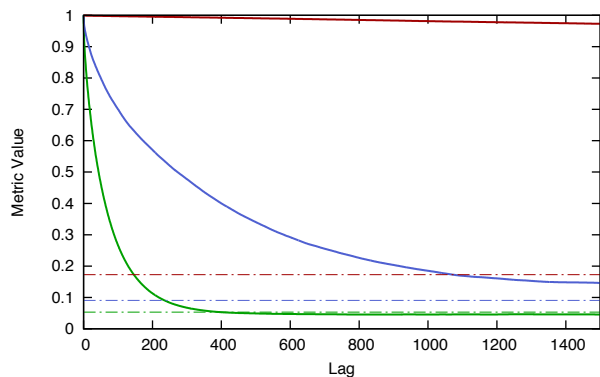
at various total surfactant concentrations. The distributions are the fraction of the total number of clusters seen of various sizes. The area under each curve is normalized to unity.



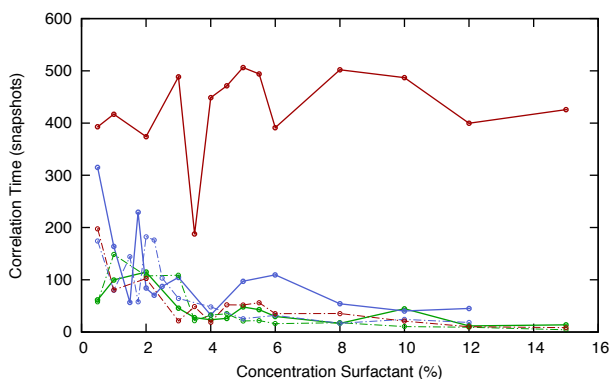
**Figure 10** Cluster size distribution for H4T4 at 1.75% total surfactant. The distributions are the fraction of the total number of clusters seen of various sizes. This distribution was produced with significantly more data than was used to produce Figure 8(c), in order to better resolve structure in the distribution due to supra-micellar aggregates, such as in the peaks near sizes of 150, 220, 270 and 350.



**Figure 11** Size distributions for micelle sized clusters for H4T3 at 3.5% (left) and 6.0% (right) concentrations resolved into contributions from micellar sizes with a radius of gyration ( $R_g$ ) less than or greater than or equal to 2 distance units. Micelles with  $R_g < 2$  are generally spherical or oblate ellipsoids. Those with  $R_g \geq 2$  are generally prolate ellipsoids, suggesting the onset of rod shaped micelles.

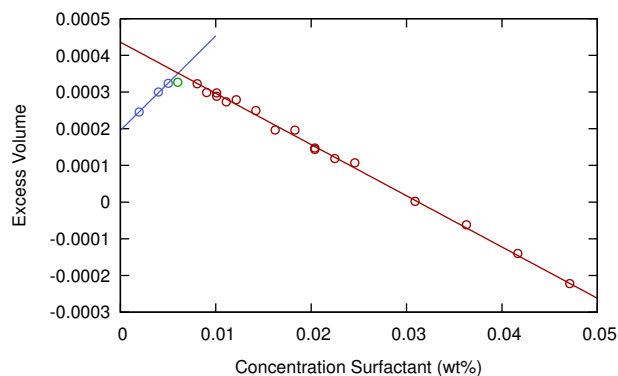


**Figure 12** The cohabitation correlation function,  $q(t)$  (equation 6), for the three molecules, shown as a function of lag time, expressed in units of the sampling period. The red, blue and green decaying lines represent H6T6 (5%), H4T4 (2%) and H4T3 (4.5%) respectively. The function  $q(t)$  gives the conditional probability that two surfactant molecules are in the same cluster at time  $t$  given that they were in the same cluster at time 0. The asymptotic values for each curve are shown with the same color dot-dashed lines.

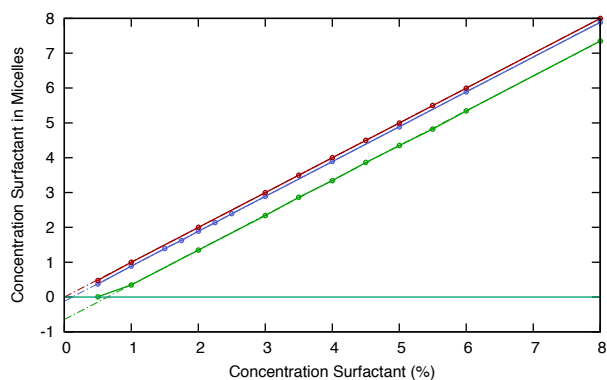


**Figure 13** Correlation times expressed in sampling periods for three molecules (green – H4T3, blue - H4T4, red - H6T6) computed for two different observables, number of micelles (solid lines) and concentration in the simulation volume due to surfactant in submicellar clusters (dashed lines). Unusually small correlation times, compared with those from nearby

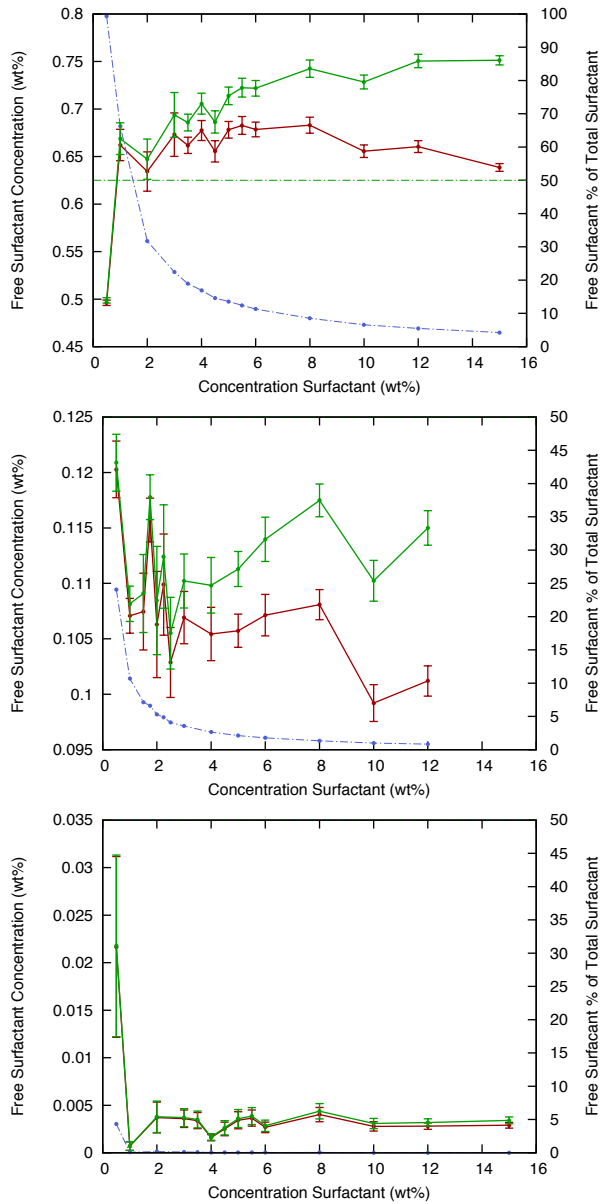
concentrations, are a sign that the simulation may not have been long enough to properly observe and characterize the timescale for variation of the relevant variable. This is the case for the mean micelle size for H6T6 at a concentration of 3.5%. When this happens, uncertainty estimates will be underestimated.



**Figure 14** Excess volume as a function of H4T3 surfactant concentration. The slope gives the average volume of a surfactant bead relative to that of a solvent bead in pure solvent. This volume depends on whether additional surfactant tends to be surrounded by solvent (below the CMC) or other surfactant (above the CMC). Linear fits (fitted gradients are 0.0258 and -0.014 in the positive and negative regions respectively) use average volumes from concentrations either below 0.6% (blue) or above 0.7% (red). The linear fits cross at a concentration of 0.60%, very near to the data point shown with green.



**Figure 15** The concentration of surfactant molecules in micellar sized clusters as a function of the total surfactant concentration for the three molecular systems (solid lines, green – H4T3, blue – H4T4, red – H6T6). A linear fit works very well for these data (dashed lines). The x-intercepts of the fits provide estimates for the CMC, since this is the total concentration where micellar material first appears.



**Figure 16** Concentration of surfactant in submicellar clusters for H4T3 (upper panel), H4T4 (middle), and H6T6 (lower) without (red) and with (green) a correction to account for volume occupied by surfactant in micellar clusters.

**Table of Contents Image**

