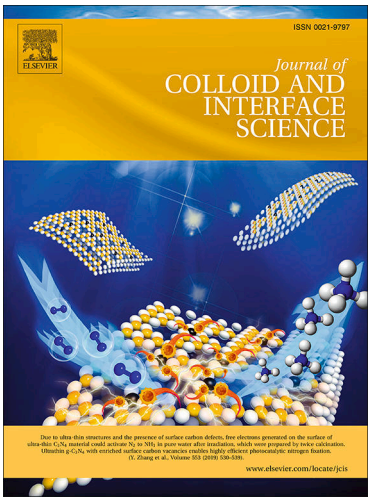
Journal Pre-proofs




# Predicting Surfactant Phase Behavior with a Molecularly Informed Field Theory

Kevin Shen, My Nguyen, Nicholas Sherck, Brian Yoo, Stephan Köhler, Joshua Speros, Kris T. Delaney, M. Scott Shell, Glenn H. Fredrickson

PII: S0021-9797(23)00012-7
DOI: https://doi.org/10.1016/j.jcis.2023.01.015
Reference: YJCIS 31606

To appear in: *Journal of Colloid and Interface Science*

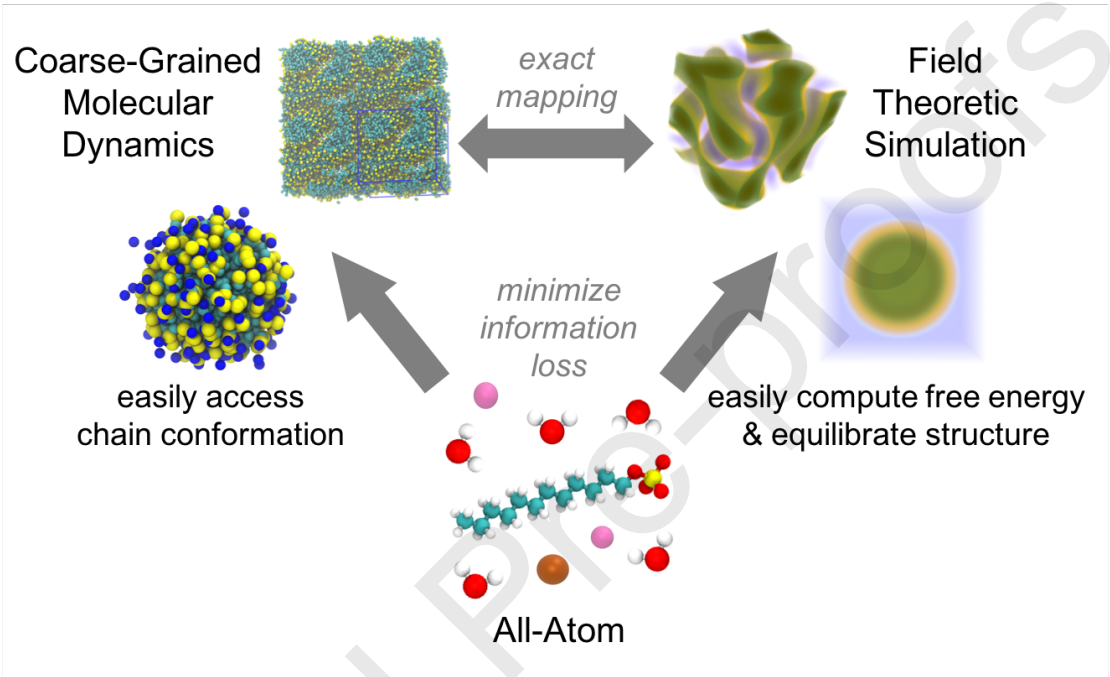
Received Date: 24 September 2022
Revised Date: 24 December 2022
Accepted Date: 3 January 2023

Please cite this article as: K. Shen, M. Nguyen, N. Sherck, B. Yoo, S. Köhler, J. Speros, K.T. Delaney, M.S. Shell, G.H. Fredrickson, Predicting Surfactant Phase Behavior with a Molecularly Informed Field Theory, *Journal of Colloid and Interface Science* (2023), doi: https://doi.org/10.1016/j.jcis.2023.01.015

# Graphical Abstract

**Predicting Surfactant Phase Behavior with a Molecularly Informed Field Theory**

Kevin Shen, My Nguyen, Nicholas Sherck, Brian Yoo, Stephan Köhler, Joshua Speros, Kris T. Delaney, M. Scott Shell, Glenn H. Fredrickson



# Highlights

**Predicting Surfactant Phase Behavior with a Molecularly Informed Field Theory**

Kevin Shen, My Nguyen, Nicholas Sherck, Brian Yoo, Stephan Köhler, Joshua Speros, Kris T. Delaney, M. Scott Shell, Glenn H. Fredrickson

- Chemically faithful simulations of self-assembly and phase behavior are possible with a molecularly informed field theory.
- We apply this workflow to sodium dodecylsulfate, show that it reproduces micellar properties and trends with added salt, and investigate self-assembled structures and mesophases. This workflow is readily extendable to complex systems with more components and larger molecules.
- Relative entropy coarse graining efficiently learns effective molecular interactions from all atom simulations.
- Field theoretic transformation of suitably coarse-grained models enables efficient equilibration and calculation of thermodynamic properties, such as phase boundaries, that are otherwise difficult to access by particle-based simulations.

J. Colloid Interface Sci.

Journal of Colloid and Interface Science 00 (2022) 1–29

# Predicting Surfactant Phase Behavior with a Molecularly Informed Field Theory

Kevin Shen[a,b,*], My Nguyen[a], Nicholas Sherck[a,1], Brian Yoo[c,2], Stephan Köhler[d], Joshua Speros[e], Kris T. Delaney[b], M. Scott Shell[a], Glenn H. Fredrickson[a,b,f]

[a]*Department of Chemical Engineering, University of California, Santa Barbara, Santa Barbara, 93106, California, United States*
[b]*Materials Research Laboratory, University of California, Santa Barbara, Santa Barbara, 93106, California, United States*
[c]*BASF Corporation, Tarrytown, 10591, New York, United States*
[d]*BASF SE, Ludwigshafen am Rhein, 67056, Germany*
[e]*California Research Alliance (CARA) by BASF, Berkeley, 94720, California, United States*
[f]*Department of Materials Engineering, University of California, Santa Barbara, Santa Barbara, 93106, California, United States*

## Abstract

**Hypothesis**

The computational study of surfactants and self-assembly is challenging because 1) models need to reflect chemistry-specific interactions, and 2) self-assembled structures are difficult to equilibrate with conventional molecular dynamics. We propose to overcome these challenges with a multiscale simulation approach where relative entropy minimization transfers chemically-detailed information from all-atom (AA) simulations to coarse-grained (CG) models that can be simulated using field-theoretic methods. Field-theoretic simulations are not limited by intrinsic physical time scales like diffusion and allow for rigorous equilibration *via* free energy minimization. This approach should enable the study of properties that are difficult to obtain by particle-based simulations.

**Simulation Work**

We apply this workflow to sodium dodecylsulfate. To ensure chemical fidelity we present an AA force field calibrated against interfacial tension experiments. We generate CG models from AA simulation trajectories and show that particle-based and field-theoretic simulations of the CG model reproduce AA simulations and experimental measurements.

**Findings**

The workflow captures the complex balance of interactions in a multicomponent system ultimately described by an atomistic model. The resulting CG models can study complex 3D phases like double or alternating gyroids, and reproduce salt effects on properties like aggregation number and shape transitions.

*Keywords:* coarse-graining, field theory, multiscale, formulations, self-assembly, simulation, microphase, micelle, surfactant, SDS

## 1. Introduction

Industrial solution formulations are complex, multicomponent mixtures consisting of a range of solvents, ions, small molecule additives, surfactants, and polymers. The formation of mesoscale structures in these formulations can endow them with unique physical properties. Recent advances in automated experimental screening provide an interesting opportunity to rapidly iterate on the vast formulation design space (*e.g.*, molecular weight, composition, architecture, chemistry, solution composition, and temperature) in order to achieve desired performance characteristics [1]. This approach has, for instance, been used to design lipid nanoparticles [2] and target yield stress in lamellar gels [3]. Predictive computer models of formulation properties provide a complementary tool for understanding and guiding the design process [4]. However, the long time scales associated with the relaxation and equilibration of molecular assemblies mean that many structural properties are out of the reach of traditional ab intio (AI) or classical all-atom (AA) molecular dynamics (MD) simulations. For example, the exchange time of some small molecule surfactants between aggregates has been estimated to be on the order of microseconds [5], which for moderately-sized systems has only recently come within reach of routine AAMD simulations.

To this end, many coarse-grained (CG) particle-based models have been developed for small molecule surfactants and polymers, typically sampled using particle-based MD approaches [6, 7, 8, 9, 10, 11, 12, 13, 14]. By mapping high-resolution AA models to coarser effective beads, CG models reduce the particle number, trading chemical detail for MD computational savings. Furthermore, the CG energetic landscape is usually softer, bringing about additional computational speed-ups relative to AAMD [15]. When working with CG surfactant models, great attention needs to be paid to equilibrating the systems; some CG surfactant models have been demonstrated to not be fully equilibrated even after 5 microseconds of simulation time [16, 17], reflecting the general difficulty of equilibrating large assembled structures *via* CGMD.

Despite these challenges, many CGMD studies have been fruitfully conducted on surfactants. The Shinoda-DeVane-Klein model of surfactants was one of the first models to be simulated on large scale using GPU's [10, 17]. A number of studies employed dissipative particle dynamics models and thermodynamic theory to develop accurate protocols using large simulation boxes to capture distributions of aggregation numbers [9, 13, 18, 19]. Another study similarly used large simulations (1080 surfactant molecules) and iterative Boltzmann inversion to develop long-ranged (12 nm) potentials that reliably reproduce aggregation numbers. MARTINI, a coarse-grained model originally designed to study lipids [20], has been extended to study micellar and phase behaviors of many surfactants [14]. For example, one study examined the LCST behavior when mixing oppositely charged surfactants [21]. Several more studies have used and tuned MARTINI models to study aggregation numbers [22] and morphologies, including hexagonal, bicontinuous,

*Corresponding author at University of California, Santa Barbara, 93106 California, United States.
*Email addresses:* kevinshen@ucsb.edu (Kevin Shen), shell@ucsb.edu (M. Scott Shell), ghf@ucsb.edu (Glenn H. Fredrickson)
[1]Present Address: BASF, Iselin, New Jersey 08830, United States.
[2]Present Address: Fero Labs, 520 Broadway, New York, New York 10012, United States.
[3]Abbreviations: molecular dynamics (MD), all atom (AA) coarse-grain (CG), self-consistent field theory (SCFT)

and lamellar phases, of alkyl trimethylammonium bromide ($C_n$TAB) surfactants [23, 24]. With careful construction of initial conditions, cubic diamond and gyroid phases have also been stabilized [25, 26, 27]. Several free energy methods for evaluating scission and transfer free energies have also been developed on MARTINI models of surfactant micelles [28, 29].

This work seeks to develop a complementary approach to the study of surfactant phase behavior. In general, the rigorous equilibration of microstructures requires that both the periodic cell and particle numbers are commensurate with the specified phase [30, 31]. To tackle this challenge, we highlight that field-theoretic simulations are uniquely poised to study self-assembly processes. Particle-based representations of CG models can in fact be exactly recast in terms of continuum fields, after which field-theoretic simulations provide *nearly-free* access to thermodynamic quantities like the cell stress, chemical potential, and free energy without requiring thermodynamic integration or related alchemical methods [30, 31]. This allows for facile and rigorous equilibration of self-assembled structures, including the evaluation of equilibrium cells for periodic phases and aggregation numbers for micelles [32, 33]. Such readily accessible free energy calculations in the field representation augments an already powerful toolbox of free energy methods in the particle representation [28, 29]. Additional advantages of the field-theoretic description include computational costs that are nearly independent of particle density, and the rapid equilibration of long chains.

To date, a practical limitation of field-theoretic models is their reliance on emergent parameters such as Flory $\chi$ interactions, typically measured or approximated *via* empirical means. In this work we overcome this practical limitation by following an approach we previously proposed to construct molecularly informed field theories [34]: we use AA simulations as a chemically-detailed source of raw data on effective molecular interactions, and use this information to parameterize field-theoretic CG models. This approach effectively leverages the strengths of both AA simulations, which are chemically detailed, with field-theoretic simulations, which excel at evaluating thermodynamic properties and equilibrating complex mesostructures. The workflow is illustrated in Figure 1a.

There exist several systematic bottom-up coarse-graining frameworks, such as force matching [35] and iterative Boltzmann inversion [36, 37]. In this work, we employ relative entropy minimization, a systematic formulation of coarse-graining as the mathematical problem of minimizing the information loss incurred when moving from a chemically-detailed AA model to a CG description [38]. Using AA models to generate molecular trajectories, we tune the CG interactions to minimize the mismatch between CG and AA configurational probability distributions, thus obtaining CG models that are faithful to the AA description.

After building the CG model, we use exact mathematical transformations to recast the CG particle model into a field theory [30], enabling direct access to thermodynamic properties of interest. Additionally, the field theories rapidly equilibrate self-assembled structures by directly propagating or relaxing collective field variables [30]. Our approach thus bridges chemical detail, accounted for by AA models, with large-scale self-assembled structures, which are characterized *via* field-theoretic techniques. Taken together, this is a powerful new approach for systematically modeling complex, multicomponent formulations and their intricate self-assembly behaviors.

In this work, we demonstrate that the approach can be effectively applied to a common anionic surfactant, sodium dodecylsulfate (SDS, $CH_3(CH_2)_{11}OSO_3^-Na^+$ , see Fig. 1b). SDS is used in

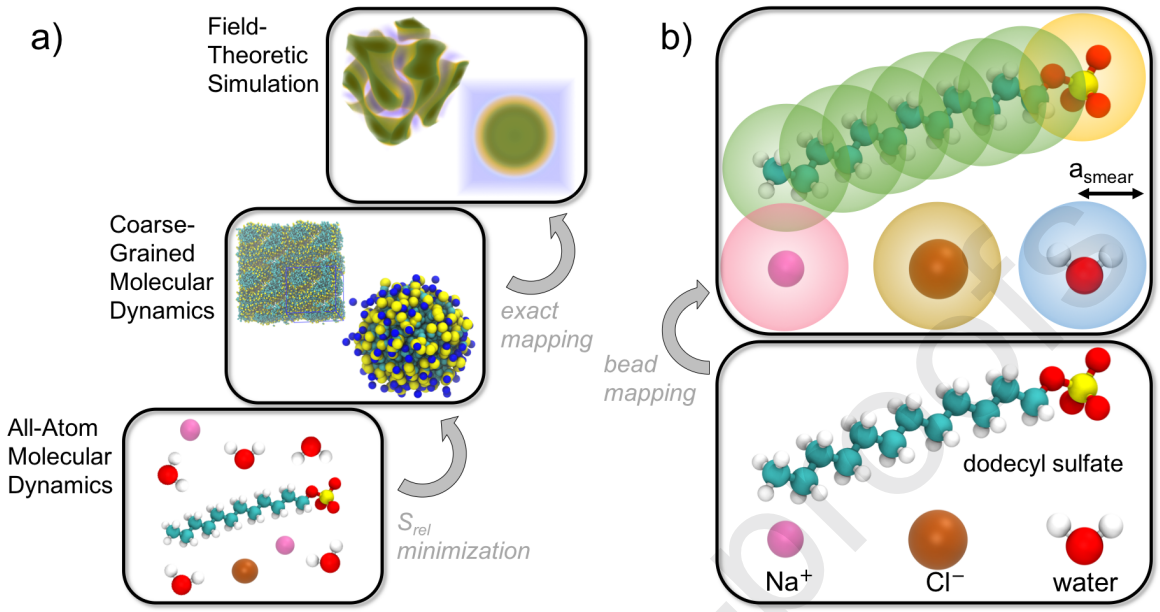



Figure 1: (a) This study employs all-atom molecular dynamics, coarse-grained molecular dynamics, and field-theoretic simulations, which are respectively well-suited for describing chemical detail, simulating small-scale assemblies, and equilibrating large self-assembled structures. (b) The scheme used for mapping all-atom coordinates to coarse-grained bead coordinates. All beads share the same smearing Gaussian radii, but have independent Gaussian prefactors that measure the strengths of nonbonded interactions in the CG model.

many formulations [39], and an accurate SDS model can be challenging to build: it must capture the sensitive balance of electrostatics, ion-water, water-water, and hydrophobic-hydrophilic interactions. In addition, even a "simple" system of SDS and salt is already a four-component system—water, SDS, cation, and anion—and has many interactions to parameterize. Using a field-theoretic representation of SDS informed by AA simulations, we explore the self-assembly behaviors of SDS. At low concentrations, we study a range of self-assembled micellar and interfacial structures, while at higher concentrations we study the morphological transitions and stabilities of different ordered mesophases.

One advantage of studying SDS is that several properties (*e.g.* surfactant headgroup area, micelle morphology) of its self-assembled structures can be simultaneously studied by AAMD, CGMD, and field-theoretic simulations. Consequently, using a combination of AAMD, CGMD, field-theoretic simulations, and experimental data, we assess to what extent the field-theoretic simulations and CG models are faithful to the AA models, and to what extent the AA models are faithful to experimental results. In Section 2, we describe the MD, coarse-graining, and field-theoretic methods. In Section 3, we present results on the quantitative calibration of the AA model, and the performance of the resulting CG models in predicting micelle structures and morphological transitions.

## 2. Methods

### 2.1. All-Atom Simulations

We perform AA simulations with the OpenMM software [40]. Nonbonded interactions use a cutoff of 1.2 nm and electrostatics are treated using Particle Mesh Ewald with a tolerance of $10^{-4}$. We conduct AA simulations at a pressure of 1 bar using a Monte Carlo barostat and use a Langevin thermostat to control temperature. Overall, we perform a variety of both bulk and interfacial simulations to collect trajectories that are used to construct the CG models. The exact barostat conditions depend on the particular system and are detailed in Section 2.2. We constrain hydrogen bonds, allowing for time steps of $dt$ = 2 fs with a Langevin thermostat relaxation time of $\tau_{AA}$= 1ps (*i.e.* $dt = 0.002\tau_{AA}$). We also use the OpenMM software to conduct CGMD simulations, using custom nonbonded potentials to implement the interactions. In CGMD simulations the Langevin relaxation time $\tau_{CG}$ is taken as the unit of time. Due to the soft nature of the CG interactions and bonds, larger time steps of $0.1\tau_{CG}$ are feasible. In terms of the Langevin relaxation time, this already represents a 50× speedup over the AA time step size. Additional acceleration in system relaxation times are expected from the reduced number of particles in the coarsened representation, as well as the smoothed energy landscape, although we have not attempted to quantify the time rescaling in this work.

For the AA force field, we use the OPC 4-point water model [41] along with the Joung-Cheatham [42] ion model. These models have been demonstrated to well-reproduce water properties over a wide temperature range [41] and well-characterize ion solubilities [42, 43]. For dodecylsulfate $CH_3(CH_2)_{11}OSO_3^-$ , we use a GAFF-based force field previously proposed by qualitatively reproducing wormlike micelle morphologies [44], which combines the headgroup sulfate force field from Yan *et al.* 2010 [45] and uses the more modern Lipid 14 force field for the alkyl tail [46], which suppresses unphysical tail crystallization. Partial charges are also adopted from this prior work.

Prior simulations assessing combinations of water and ion models for SDS have demonstrated the importance of sulfate-ion interactions in producing qualitatively correct micelle morphologies, and these morphological structures have been used to select between ion force fields [5, 44, 47, 48]. While radial distribution functions can identify particularly problematic force fields, there are no comparable experimental data with which to evaluate against [5]. Similarly, the counterion binding degree is another closely related quantity that has been used to evaluate forcefield quality but its experimental determination was found to yield too large a range of possible values (0.45-0.86) to use as a precise target for optimizing a force field [44].

Instead of further modifying the water and ion models, which have already been calibrated to reproduce a variety of water and ion properties [41, 42], we instead focus on adjusting the nonbonded Lennard Jones interaction between sodium ions and oxygens that are bonded only to the sulfate headgroup (the oxygen joining the alkyl tail and the sulfur atom has a different atom type). In our parameterization strategy the cross-interaction between the chosen oxygen atom type and sodium ions, which is usually determined by the Lorentz-Berthelot mixing rule, is directly modified, leaving all other interactions involving the same atom types untouched. This targeted approach is sometimes termed a "nonbonded fix" or NBFIX [49, 50], and contrasts with an alternative approach of directly changing the sodium radius, which would change sodium interactions

with all species. We find that increasing the Lennard Jones interaction radius between these two atom types by just ~4% allows us to quantitatively reproduce the surface tension of SDS deposited at a water-vacuum interface. Surface tensions are calculated using the test area method by perturbatively calculating the free energy associated with a change in interfacial area at constant volume [51, 52]. Results are presented in Section 3.1, where we find that quantitatively matching surface tension leads to marked improvements in the micelle morphology and is in agreement with experiment and prior simulations [5, 44, 47].

### 2.2. Coarse-Graining Workflow

We use the relative entropy minimization technique[38] to minimize the information loss incurred when constructing a CG model from an AA simulation. The CG model uses Gaussian pair interactions between all beads:

$$\beta U_{ij}^{Gaussian} = \frac{u_{ij}}{(2\pi(a_i^2 + a_j^2))^{3/2}} e^{-r_{ij}^2/2(a_i^2+a_j^2)} \tag{1}$$

$$= \beta\gamma_{ij} e^{-r_{ij}^2/2(a_i^2+a_j^2)} \tag{2}$$

where $i, j$ index the beads, $r_{ij}$ the distance between beads $i$ and $j$, and $a_i$ denotes the "smearing length" of bead $i$. The interaction strength can either be given in terms of the dimensionless parameter $\beta\gamma_{ij}$ or volume $u_{ij}$. An interaction function of this form was used previously to quantitatively reproduce activity coefficients in simple binary mixtures after suitable relative entropy parameterization [53], and also applied to the prediction of PEO cloud points [34]. For simplicity, all beads share a common smearing length $a_0 = \rho_{water}^{-1/3} = 0.31317$ nm, where $\rho_{water}$ is measured from an AA simulation at $T = 348$ K and pressure 1 bar. The nonbonded cutoff is taken to be $r_{cut} = 5a_0$, beyond which the interaction energy $\beta U_{ij}^{Gaussian} < 4.3 \cdot 10^{-5} u_{ij}/a_0^3 \approx 0.002\beta\gamma_{ij}$.

Bonded interactions are treated with harmonic bonds, as in the discrete Gaussian chain model [30]:

$$\beta U_{ij}^{bond} = \frac{3}{2b_{ij}^2} r_{ij}^2. \tag{3}$$

where $b_{ij}$ can be interpreted as the intrinsic Gaussian bond length between beads $i$ and $j$ when no nonbonded interactions are present. Finally, electrostatic interactions are treated with smeared Coulomb interactions:

$$\beta U_{ij}^{elec} = \frac{l_b z_i z_j}{r_{ij}} \mathrm{erf}\left(\frac{r_{ij}}{\sqrt{2(a_i^2 + a_j^2)}}\right) \tag{4}$$

where $z_i$ and $z_j$ are the bead charges (in units of $e$), and $l_b = e^2/4\pi\varepsilon_r\varepsilon_0 k_B T$ is the Bjerrum length. The dielectric constant $\varepsilon_r(T)$ is obtained from bulk water simulations (at $T = 348$ K, $\varepsilon_r = 62.6$ and $l_b = 0.7676$ nm). The smearing lengths $a_i$ and $a_j$ use the same values as in the nonbonded Gaussian interactions.

These interactions are deliberately chosen to be soft interaction forms that avoid resolving angstrom-level structural detail. This facilitates rapid sampling (*i.e.*, no stiff interactions such as

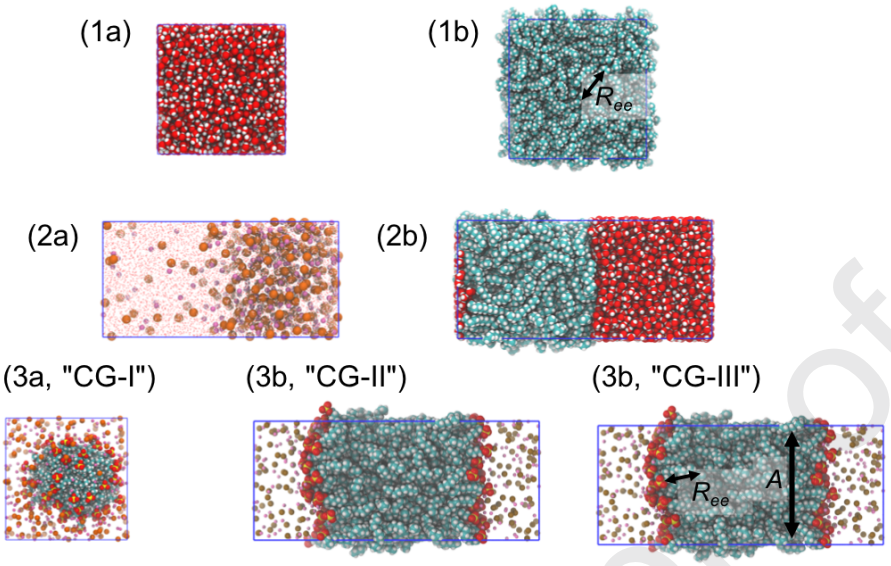



Figure 2: Simulation parameters are obtained *via* 3 stages and several reference all-atom simulations. Stage 1 considers obtaining bulk water and dodecane parameters. Stage 2 obtains water-ion and water-dodecane cross interactions, holding the water-water and alkane-alkane interactions fixed. Finally, in stage 3, three separate reference systems and/or coarse-graining schemes are considered. This leads to three different coarse-grained parameter sets: CG-I (coarse-grained from 60 DS molecules assembled in a micelle at 2M NaCl), CG-II (coarse-grained from DS molecules deposited at an alkane water interface), and CG-III (same as CG-II, except with additional $r_{ee}$ and surface area $A$ constraints applied.)

near-rigid bonds or hard core interactions). This approach is in the spirit of dissipative particle dynamics models (DPD) [9, 54, 12], and also proves advantageous in lowering the spatial resolution required to conduct the field-theoretic simulations.

Our mapping scheme and definitions of bead types are presented in Figure 1b. Each water molecule is mapped to a single neutral bead, while the small ions $Na^+$ and $Cl^-$ are mapped to single beads with $+1$ and $-1$ charge, respectively. Dodecane is also considered to facilitate parameterization of the dodecylsulfate (DS) alkyl tail. Both dodecane and the 12-carbon tail of the DS molecule are each divided into 6 beads of 2 carbons each and use the same $C_2$ bead type. Lastly, the sulfate $SO_4^-$ headgroup of DS is modeled by a single bead with charge of $-1$.

We coarse-grain the model in successive stages to manage the complexity of the multicomponent mixture. Interaction parameters are determined by the stage in which they are first introduced. The three stages of reference simulations and coarse-graining are presented in Figure 2. In the first stage, we determine CG parameters for pure, bulk systems from AA references simulated using an isotropic barostat. For a given temperature, we choose the water repulsion $\gamma_{WW} = 0.2862\ k_BT$ such that it approximately reproduces water's (in)compressibility [53]. When barostats are required, the CGMD simulations are conducted at a coarse-grained pressure $P_{cg}$, determined by simulating the water with the chosen water repulsive interaction $\beta\gamma_{WW}$ at the atomistically-determined water density. This approach is standard in dissipative particle dynamics simulations (DPD) [12] and has also been previously applied in the context of relative entropy coarse-graining—the main observation is that for liquid systems the exact pressure is of secondary importance for study-

ing mixture thermodynamics as long as the systems are sufficiently incompressible [53, 34]. We coarse-grain at 348 K, where SDS is known to form a variety of mesophases without crystallization of the alkyl tails (at room temperature SDS crystallizes above ~40 wt%) [55]; crystallization phenomena cannot be captured by the CG model. Additional temperatures are not considered in this work because the experimental mesophase morphology transitions we are interested in are essentially constant at temperatures where crystals are absent [55]. At 348 K and a pressure of 1 bar, the OPC force field for water predicts a density of $\rho_w = 0.9745$ g/mL, or a molecular volume of $v_w = \rho_w^{-1} = 0.03107$ nm$^3$. The CG pressure for this density and the selected $\beta\gamma_{WW}$ is $P_{cg} = 239.282\ k_BT$/nm$^3$. Next, the bonded and nonbonded interactions between $C_2$ beads of the alkyl tail of the SDS are parameterized by considering a dodecane solution. CGMD simulations during the relative entropy minimization of dodecane parameters are conducted at $P_{cg}$, and a target volume augmented Lagrangian constraint is applied to ensure the reproduction of the dodecane density. Further, since there are no angle potentials in our CG model, which significantly affect the chain conformation of short molecules like dodecane, an additional augmented Lagrangian constraint is applied to match the dodecane end-to-end distance [56].

In the second stage, we simulate and coarse-grain from mixtures. The dodecane-water interaction $\gamma_{C_2W}$ is obtained from a dodecane-water simulation, such that the simulation trajectory provides information about the immiscibility of the alkane with water. These simulations are conducted with the interfacial area held constant, and a barostat applied in the normal direction (1 bar in the AA simulations, $P_{cg}$ in the CG simulations). The salt-salt ($\gamma_{\mathrm{NaNa}}, \gamma_{\mathrm{NaCl}}, \gamma_{\mathrm{ClCl}}$) and salt-water ($\gamma_{\mathrm{Na/Cl\text{-}W}}$) parameters are coarse-grained using an external potential ensemble technique to increase force field transferability across ion concentration [53]. This entails simulating an elongated box with a sinusoidal potential applied to the salt ions; relative entropy optimization from such an inhomogeneous system improves predictions of the activity coefficents by enhancing the observed variance of local concentrations within a system [53]. The equilibrium box size is initially determined from a homogeneous simulation box with an isotropic barostat. Subsequently, the simulation box is held fixed and the external sinusoidal potential is turned on. The strength of the sinuosoidal potential is determined by choosing an external potential strength that maximizes the Fisher information. In both stages 1 and 2, AA simulations are equilibrated for 50 ns and production data collected over an additional 50 ns.

Finally, in the third stage, we obtain the remaining nonbonded interaction parameters (alkyl $C_2$ and $SO_4^-$ interactions with every other bead type) by evaluating three different approaches leading to three CG parameter sets, with reference systems and constraints as follows (also see Fig. 2):

- CG-I: A lone SDS micelle with an aggregation number of 60 and 2M of added NaCl salt.
- CG-II: SDS deposited at a dodecane-water interface with 2M NaCl in the water phase. Axial and lateral dimensions of the simulation box are allowed to fluctuate independently in response to the same bulk pressure, allowing interfacial area to fluctuate and equilibrate.
- CG-III: Same interfacial reference system as CG-II, but coarse-graining is done with an additional augmented Lagrangian constraint [56, 57] to match the interfacial area and the end-to-end distance of alkyl tails in the AA reference. The bonded interaction of the hydrocarbon tails is allowed to readjust.

All AA reference simulations are equilibrated for 50 ns before collecting production data for 250 ns, while CGMD simulations in each inner-loop of the relative entropy optimization are conducted for $1.5\times10^6$ time steps ($1.5\times10^5\tau_{CG}$ of simulation time). The aggregation number of model CG-I is commonly used in simulations of SDS [58]. Model CG-III is physically motivated by the desire to reproduce two important characteristics of surfactants that dictate their self-assembled morphology: surfactant length and headgroup area. Details of the augmented Lagrangian constraint used in deriving model CG-III are given in the SI. Models CG-I and CG-II serve as controls to understand the efficacy of direct relative entropy minimization.

Finally, AAMD simulations are also conducted at 30 wt% for a wormlike micelle in water that extends periodically along the long axis of an elongated box. An anisotropic barostat is applied to allow the wormlike micelle length and cross-sectional area to equilibrate. Although these simulations are not used for coarse-graining, they are used for validation of CG model transferability. The AA simulations of wormlike micelles are conducted for 275 ns. The simulation cell and micelle dimensions equlibrate over the first 60 ns, and the remaining trajectory is used for evaluating micelle statistics. CGMD simulations of the wormlike micelle are equilibrated for $1\times10^6$ time steps and production data collected over an additional $1\times10^6$ time steps.

### *2.3. Field-Theoretic Calculations*

The resulting CG models can be *exactly* transformed into field-theoretic models. Specifically, we use an auxiliary-field formulation, where we switch from a particle-based description of a system in terms of particle coordinates to a mathematically equivalent representation of the system using auxiliary $\{w\}$ fields, with one field per bead type:

$$\int\prod_i d\mathbf{r}_i e^{-U(\{\mathbf{r}\})}\rightarrow\int\prod_j^{species}\mathcal{D}w_j e^{-H[\{w\}]} \tag{5}$$

where $H[\{w\}]$ is now a field-theoretic action determining the statistical weight of a given field configuration $\{w\}$. Full field-theoretic sampling of field configurations is *equivalent* to running CGMD simulations. Details of this transformation are described elsewhere.[30] The $w$ fields are conjugated to density $\rho$-fields, which allows easy determination of density fields *via* appropriate thermodynamic derivatives with respect to the $w$-fields (for numerical efficiency the $w$-fields are linearly transformed into exchange fields) [32]. One notable aspect of the field theory is that the auxiliary fields are excellent collective variables that enable rapid equilibration over configurations. Another important feature of the field theory is that it provides ready access to quantities like the chemical potential and free energy that require significant and admirable work to access with particle representations of complex molecules [31, 32]. In particular, the free energy can be readily minimized over particle numbers *and* cell size to determine equilibrium self-assembled structures. To date, no analogous calculations can be performed with particle representations because of long equilibration times and cost of minimizing free energies over particle numbers.

While the free energy can be calculated with full fluctuations *via* complex Langevin sampling [31], for further gains in computational speed we invoke a mean field, self-consistent field theory (SCFT) approximation [30]. The SCFT approximation consists of evaluating the field-theoretic

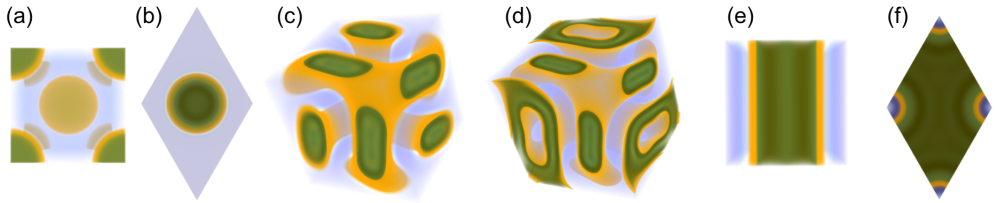



Figure 3: Self-assembled structures considered in this study: (a) spherical, (b) cylindrical (shown: planar cross-section), (c) double gyroidal, (d) alternating gyroidal, (e) lamellar, (f) inverse cylindrical (shown: planar cross section). Blue denotes the aqueous phase, green the surfactant alkyl tails, and gold the surfactant headgroups. For clarity, small ion concentrations are not shown.

partition function at a single dominant field configuration by solving for the saddle-point value of each auxiliary field $w_i^*$:

$$\left.\frac{\delta H}{\delta w_i}\right|_{w_i^*} = 0 \tag{6}$$

$$Z = \int \mathcal{D}w e^{-H[\{w\}]} \approx e^{-H[\{w^*\}]} \tag{7}$$

$$F = -\ln Z = H[\{w^*\}] \tag{8}$$

where the last line demonstrates that converged SCFT results for the fields $w^*$ provide direct access to the free energy. Thus the mathematical difference between SCFT and the fully fluctuation field theory (or its equivalent CGMD particle representation) is that the SCFT ignores the fluctuation of fields about the saddle point $\{w_i^*\}$, which is often a good quantitative approximation but sometimes can have qualitative consequences (e.g., in the close vicinity of critical points or unbinding transitions) [32]. SCFT has been used with great success to study coarse-grained polymeric and surfactant systems [30, 59]. The volumetric free energy density can be further minimized with respect to both the periodic cell size as well as the geometry of the self-assembled structure [30]. An often under-appreciated property of field-theoretic representations is that by restricting the eigenfunctions used to represent fields to those compatible with a desired space group, SCFT simulations also allow one to enforce microphase symmetries [60]. This allows one to rigorously study the free energies of different (even metastable) mesophases and assess their free energy distance from stability. We consider spherical, cylindrical, double gyroidal, alternating gyroidal, lamellar, and inverse cylindrical structures (see Figure 3 for examples of converged structures using model CG-III). For a prescribed composition (wt% of each species), the overall concentrations are determined by assuming incompressibility and using the molecular volumes estimated from AA simulations (see SI). Relaxing cell size at fixed concentration allows for the determination of free energy density of each candidate microstructure; the microstructure with the lowest free energy density is then the putative equilibrium structure.

## 3. Results and Discussion

### *3.1. Quantitative Calibration of SDS Force Field*

To quantitatively parameterize an AA SDS force field, we use the surface tension as a well-defined simulation target. We deposit SDS molecules at a water-vacuum interface at the experimentally determined, room-temperature saturation interfacial density of $2.27/\text{nm}^2$, or a headgroup area of $a_s = 0.44$ nm$^2$ [61, 62]. Specifically, we perform canonical ensemble simulations of a (5.3 nm)$^3$ cube of water flanked by 5.35 nm of vacuum on either side (with periodic boundary conditions, there is effectively 10.7 nm of vacuum between the two interfaces), and 64 SDS molecules deposited at each interface, with alkyl tails pointing towards the vacuum and the headgroup pointing into the water phase (Fig. S1-a). A commensurate number of sodium ions is included in the water phase to neutralize the sulfate headgroup charges.

With the default force field parameters, the SDS molecules are seen to form patchy aggregates at the water-vacuum interface (Fig. S1-Ia), resulting in a surface tension near that of water alone at the water-vacuum interface: 72 mN/m [63]. The pair distribution function also indicates unphysically strong sodium-sulfate binding (Fig. S2), which has also been observed in other studies on SDS force fields. To remedy this, we increase the Lennard-Jones diameter of the interaction between sodium ions and the three oxygen atoms on the sulfate headgroup (the oxygen connecting the dodecyl tail to the sulfur atom has a different atom type and its interactions with sodium are unchanged). By systematically increasing this parameter to $\sigma_{Na-oP} = 0.2679$ nm, we are able to reduce the surface tension to the experimental saturation value of 40 mN/m [61, 62, 64]. Upon matching the surface tension, the sodium-sulfate pair distribution function also becomes much more reasonable (Fig. S2), and examination of the water-vacuum interface reveals more even surfactant coverage (Fig. S1-IIa).

Further, following common practice in the AA simulation of SDS, we perform qualitative morphological studies of micelles in the newly calibrated force field [5]. The standard in the literature is to initialize the SDS molecules in a bilayer (Fig. S1-b). Force fields that overemphasize the bonding of sodium to the sulfate headgroups usually remain trapped in unphysical, disc-like bilayer structures termed "bicelles" [5] (Fig. S1-Ib). In contrast, the newly calibrated force field correctly relaxes the initial bilayer structure into elongated micelles (Fig. S1-IIb). This agrees with experiments and other AA studies that investigated and proposed different force field parameterizations that produce micelles instead of bicelles [5, 44, 47, 55]. Furthermore, because the time scales required to obtain equilibrated micelle morphologies are inaccessible by AA simulations, this protocol is restricted to making qualitative discriminations of micelle structures. Instead of relying only on a qualitative assessment of micelle morphologies, our proposed nonbonded correction for the force field quantitatively reproduces experimental surface tensions. The morphological agreement of micelle structures is an emergent behavior from matching this simple, well-characterized thermodynamic property. It should be noted that if given accurate AA force fields, no experimental input would be necessary, and the coarse-graining and field-theoretic simulation workflow described in the following sections would give truly *a priori* predictions.

### *3.2. Coarse-Grained Model Development and Transferability*

We first assess the faithfulness of the proposed workflow in preserving chemical information and describing surfactant self-assembly by simultaneously comparing AA, CGMD, and SCFT

| Model | $r_{ee}$ (nm) | $a_s$(nm$^2$), MD | $a_s$, SCFT |
|---|---|---|---|
| AA | 1.05 | 0.41(7) | – |
| CG-I | 1.18 | N/A | N/A |
| CG-II | 1.32 | 0.33(0) | 0.311 |
| CG-III | 1.04 | 0.41(4) | 0.435 |

Table 1: SDS interfacial characterization. CG-I: standard $S_{rel}$ coarse-graining from N=60 micelle at 2M NaCl. CG-II: standard $S_{rel}$ coarse-graining from SDS at dodecane-water interface. CG-III: coarse-grained from SDS at dodecane-water interface, with additional $r_{ee}$ and $a_s$ constraints.

results. In order to compare to the AA force field upon which the models are based, we need well-controlled systems that are reproducible and accessible by AA simulations. The first system we consider is the same interfacial configuration used for coarse-graining some of the models. Representative AA and CG structures, and SCFT density profiles, are presented in Figure 4. We report both the surfactant tail end-to-end distance $r_{ee}$ and the headgroup area $a_s$, given their known importance in determining the packing parameter [65, 66] and micelle shapes. The results are presented in Table 1.

We first report that CGMD simulations of model CG-I, coarse-grained from an AA reference with 60 surfactants at 2M NaCl, correctly recapitulate the AA reference structure, with only a slight discrepancy in the tail length (AA: 1.04 nm, CGMD: 1.07 nm). However, Table 1 shows that despite well-reproducing its AA reference, model CG-I is not transferrable to the interfacial system. In addition to over-estimating $r_{ee}$ by 12.4%, it is unable to stabilize the dodecane-water interface. In fact, the surfactant molecules desorb and partition predominantly as unimers into the dodecane phase, continually shrinking the interface until it is less than the simulation cutoff range. Model CG-I was coarse-grained from a very different system, a spherical micelle, so errors may be attributable to a severe lack of model transferability when coarse-graining from an isolated micelle. In fact, referring to the CG parameters in the SI Table S3, one can see that model CG-I has much less repulsive sulfate headgroup interactions than models CG-II and CG-III. Additionally, the hydrocarbon tail of model CG-I uses parameters that reproduce dodecane's end-to-end distance (1.04 nm) in the bulk, which is not far from the AA end-to-end distance at the interface (1.05 nm). However, due to the flexibility of the CG model, the 12 carbon alkyl tail undergoes additional stretching when adsorbed to the dodecane-water interface, leading to an overestimate of the end-to-end distance in simulations of model CG-I.

The lack of transferability of model CG-I to the interfacial system underscores the sensitivity of CG parameters to the references they are coarse-grained from. It is likely that some other micellar reference, or even an extended ensemble of micelles may be a more appropriate reference [67]. While some emerging work has explored the possibility of optimizing the reference ensemble [53], in this work we avoid the complexities of exploring and adequately equilibrating an ensemble of micellar references, and instead investigate whether a simpler, interfacial reference can produce reasonable CG models (*i.e.* models CG-II and CG-III).

Model CG-II also shares the same alkyl tail bonding parameters as model CG-I, and similarly exhibits overstretching of the alkyl tail. However, even with a poor tail model, coarse-graining from the alkyl-water interface leads to a significant *qualitative* improvement in model CG-II over coarse-graining from a spherical micelle (model CG-I). An important distinction is the fact that model CG-II can stabilize an interface and surfactants are correctly retained at the interface (this explains why model CG-II shows more chain stretching than CG-I, because more chains are adsorbed to the interface in model CG-II). Models CG-I and CG-II differ only in their nonbonded

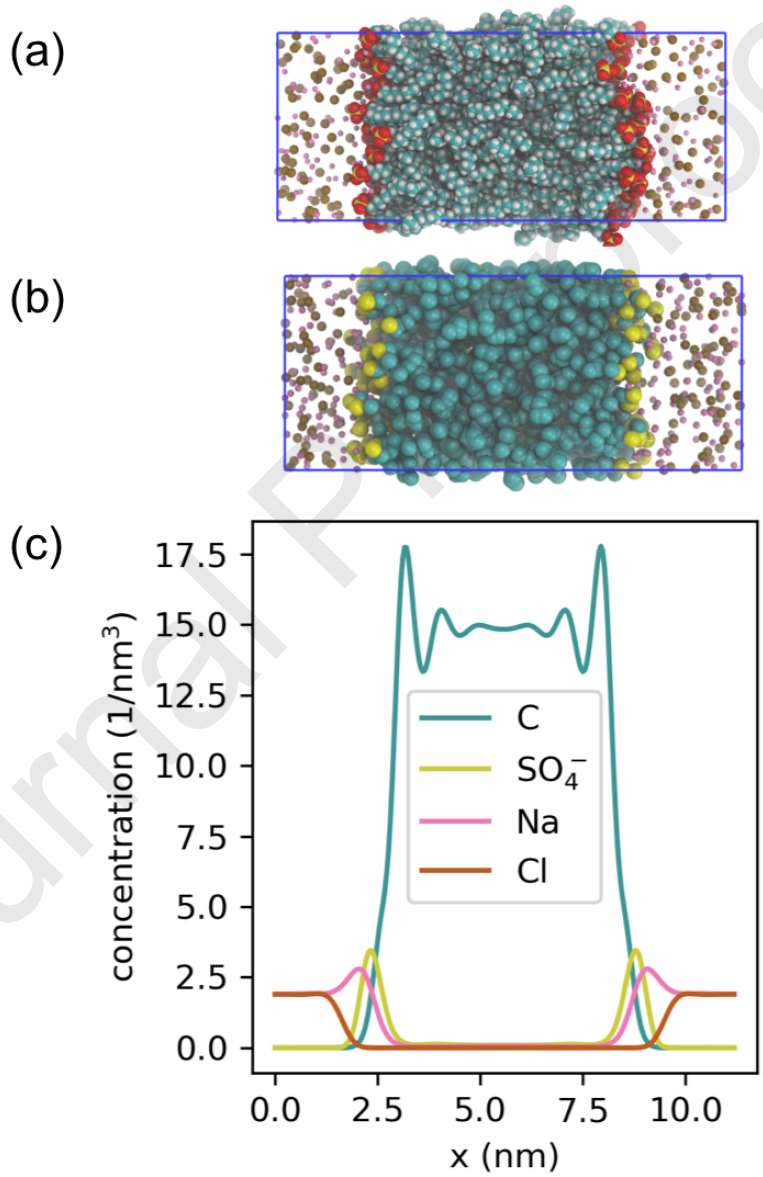



Figure 4: (a) All-atom reference simulation. (b) Coarse-grained molecular dynamics using model CG-III. (c) SCFT density profiles using model CG-III.

parameters, demonstrating that tuning nonbonded interactions alone can correct for the unphysical desorption of surfactants in model CG-I.

It is surprising that although model CG-II is coarse-grained from an interfacial simulation and leads to a large qualitative improvement over model CG-I, it *does not* reproduce the headgroup interfacial area or tail length. Apparently the relative entropy minimization determines that the penalty associated with getting the interfacial area wrong is outweighed by other contributions to the information loss. This rationalizes the necessity of applying additional constraints to the relative entropy optimization process to reproduce both the alkyl tail $r_{ee}$ and interfacial area in the derivation of model CG-III (details in Sec. 2.2 and SI Sec. 4). Naturally, once the constraints are applied, both properties are reproduced. In our experience, applying constraints on $r_{ee}$ or interfacial area alone was insufficient to reproduce the other property; both constraints need to be applied. Interestingly, whereas the $r_{ee}$ predicted for the interfacial system by model CG-I (coarse-grained from micelle) was off by 12.4%, model CG-III (coarse-grained from the dodecane-water interface) exhibits improved transferability in the reverse direction and only mis-estimates the micelle $r_{ee}$ by 4%. It may seem surprising that simply matching the average surfactant head group area and $r_{ee}$ at the interface (instead of the full distribution of end-to-end vectors of the whole chain) is sufficient to achieve reasonable self-assembly behaviors. We believe that there are physical grounds for this agreement—the surfactant packing parameter $p = v/(a \cdot l)$ [66] characterizing self-assembled morphologies is dependent on just the surfactant area $a$ and length $l$.

Additionally, 1D SCFT simulations are also conducted to evaluate the equilibrated interfacial morphologies. By integrating the density profile, one obtains the surfactant area density in each interface, and the headgroup area $a_s$ is taken to be the inverse of the area density. The SCFT approximation successfully reproduces the CGMD value for the surfactant headgroup area to within 0.02 nm$^2$ (5%), which is remarkable given the short nature of the surfactants and the expectation for significant fluctuation effects. This gives us increased confidence that SCFT can be used to quickly study properties that would otherwise be expensive *via* MD simulations.

For the next test, we consider a system different from any of the reference states (spherical micelle and alkane-water interface). We consider SDS at 30 wt%, where it is experimentally known to form wormlike micelles [68]. Comparing to wormlike micelles provides an additional test of both the faithfulness of the CG models to the underlying AA reference, as well as transferability of the CG models to system states that are not included in the AA references. To simulate wormlike micelles, we preassemble SDS surfactants into straight cylinders and apply an anisotropic barostat where the box length along which the micelle is aligned is allowed to fluctuate independently of the lateral dimensions of the simulation cell. This allows the wormlike micelles to equilibrate their lengths, and we are able to reliably equilibrate wormlike micelles in both AA and CG molecular dynamics. Over the simulation times considered, the AA model does not show any signs of aggregate breakup and the micelles remain elongated. This is consistent with experimental observations at this condition, and further validates that matching the surface tension is a viable way of obtaining a reasonable AA parameterization of SDS.

Figure S3 presents snapshots of the resulting wormlike micellar structures. Most notably, model CG-I, coarse-grained from a spherical micelle, actually formed flat structures in CGMD instead of wormlike micelles. This model performs the worst in reproducing SDS behavior at an interface, and it continues to perform poorly for the wormlike micelle as well. Secondly, we

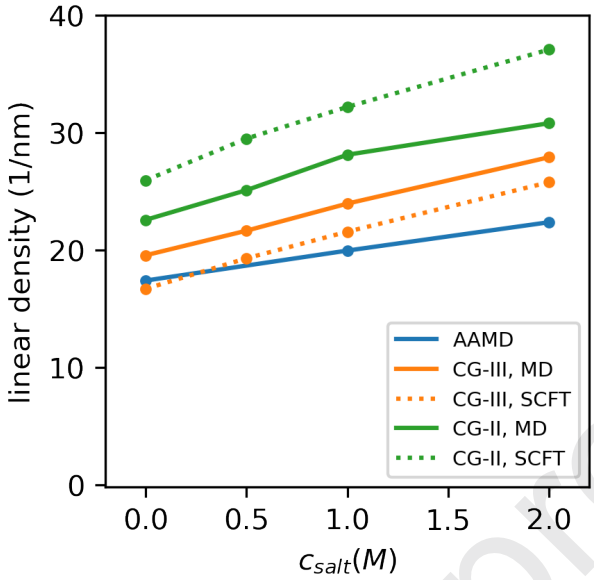



Figure 5: Micelle linear densities at 30 wt% SDS. Solid lines are linear densities along the box axial direction, while dotted lines are the same quantity evaluated using SCFT.

observe that both models CG-II (not shown) and CG-III correctly form cylindrical micelles. To quantify the wormlike micelle structure, we consider the linear density of surfactants along the z-axis of micelle orientation. Given the fairly straight configurations of the wormlike micelles as seen in Figure S3 (the micelles do not fold back on themselves), this is a reasonable leading estimate of the surfactant linear density.

In Figure 5 we quantitatively characterize the surfactant linear densities at 30 wt% SDS as a function of added salt NaCl. We observe that qualitatively, both model CG-II and model CG-III correctly predict that the surfactant linear density increases with increasing salt. Quantitatively, we observe that model CG-III is in reasonable agreement with the AAMD results, overpredicting surfactant linear densities (relative to the box length) in the range of 10-25%. Surprisingly, the upper end of the deviations is at 2M salt conditions where the CG model parameters are derived, and the model is in fact *more* accurate when transferred down to 0M salt conditions. Meanwhile, model CG-II is less accurate than model CG-III, overestimating the surfactant linear density by 25-40% over the salt concentration range.

Finally, we also validate the SCFT field theory simulation by equilibrating 2D simulations at the same concentrations as those in the CGMD simulations—circular structures in 2D represent idealized cylindrical structures in 3D. The unit cell sizes are relaxed with concentrations fixed to those obtained from CGMD simulations. These 2D SCFT simulations do not include fluctuations of the micelle contour, but should nevertheless give a good approximation of the free energies involved in assembling the cylindrical morphology of wormlike micelles. Integrating the concentration profiles over the unit cell cross-sectional area yields equilibrium linear densities of the surfactant along the length of the cylindrical micelles. The SCFT results for model CG-III are in good agreement with both the CGMD results and the AA results, reproducing the CGMD values

to roughly 10%. Meanwhile, the SCFT predictions for model CG-II actually overestimate the surfactant linear density obtained in CGMD by 15-25%. As detailed in Sec. 2.3, the main difference between the SCFT-approximated field theory and CGMD is that the former ignores field fluctuations [30]; we thus attribute discrepancies between SCFT and CGMD in models II and III to these fluctuations. Lastly, although model CG-I does not produce a stable wormlike micelle, SCFT is still able to trap metastable cylindrical structures to estimate an expected surfactant linear density. The resulting values exceed the axes of Figure 5, and range from 43/nm at 0M NaCl to 260/nm at 2M NaCl.

Overall, model CG-III, which is parameterized to reproduce surfactant tail length and headgroup area at the alkane-water interface, performs the best when considering wormlike micelles. This again suggests that with an appropriate coarse-graining scheme, quantitative CG models that capture and describe the self-assembly physics of the underlying AA model can be built.

### *3.3. Micelle Aggregation Numbers*

Given the reasonable agreement of the AAMD and CGMD simulations, and SCFT with the CGMD simulations, we now explore field-theoretic predictions of quantities that take significant work to obtain by MD, whether AA or CG; we first focus our attention on micelle aggregation numbers ($N_{agg}$). Typically, in MD attention needs to be paid to simulation scales (in both time and number of surfactants) to equilibrate the slow processes of diffusion, surfactant exchange, and micelle fission and fusion [29, 69, 70, 71, 72]. The characteristic time for these processes are easily on the order of $\mu$s for common surfactants [16, 17].

As previously discussed, one of the unique features of the field-theoretic representation is the facile access to free energies and its direct minimization over cell size *and* particle number. At lower concentrations where discrete micelles are most stable, minimizing the free energy with respect to the periodic cell size at fixed molecular concentrations (thus varying the total number of molecules) allows for the straightforward determination of equilibrium micelle aggregation numbers and micelle number densities. This procedure complements an equivalent particle-based approach of repeatedly removing surfactants from the micelle and evaluating the free energy difference upon each removal [28, 29].

In Figure 6a, we present SCFT predictions of micelle aggregation numbers of SDS under salt-free conditions. We see that model CG-I, parameterized from a spherical micelle greatly overestimates the aggregation number. Model CG-II similarly also overpredicts the micelle sizes. It is likely that the significantly larger micelle sizes in model CG-II are a function of both a surfactant tail that stretches too much as well as smaller headgroup areas that allow for much tighter packing of surfactants (see Table 1). In contrast, model CG-III performs quite admirably for the aggregation number when comparing to inferences from scattering data [73]. SCFT evaluations of model CG-III appear to provide a good estimate of the balance of the energetic driving force for surfactants to aggregate into discrete clusters versus the entropic preference for surfactant molecules to be free.

Next, in Figure 6b, we demonstrate that the SCFT of model CG-III correctly captures the salt-dependence on micelle size, and is in qualitative agreement with experiment [74]. Adding salt is expected to increase the screening of headgroup charges and reduce the repulsions, thus increasing the surfactant density and allowing for more surfactants to pack together. Correspondingly, this

leads to an increase in micelle size with added salt, and is closely related to the salting out effect of ionic surfactants [75]. We conclude that the derived CG models, when combined with SCFT, can give reasonable predictions of micelle formation. Interestingly, for nonionic surfactants adding salt may yield either salting-in *or* salting-out behaviors [76]. With neutral surfactants, electrostatic

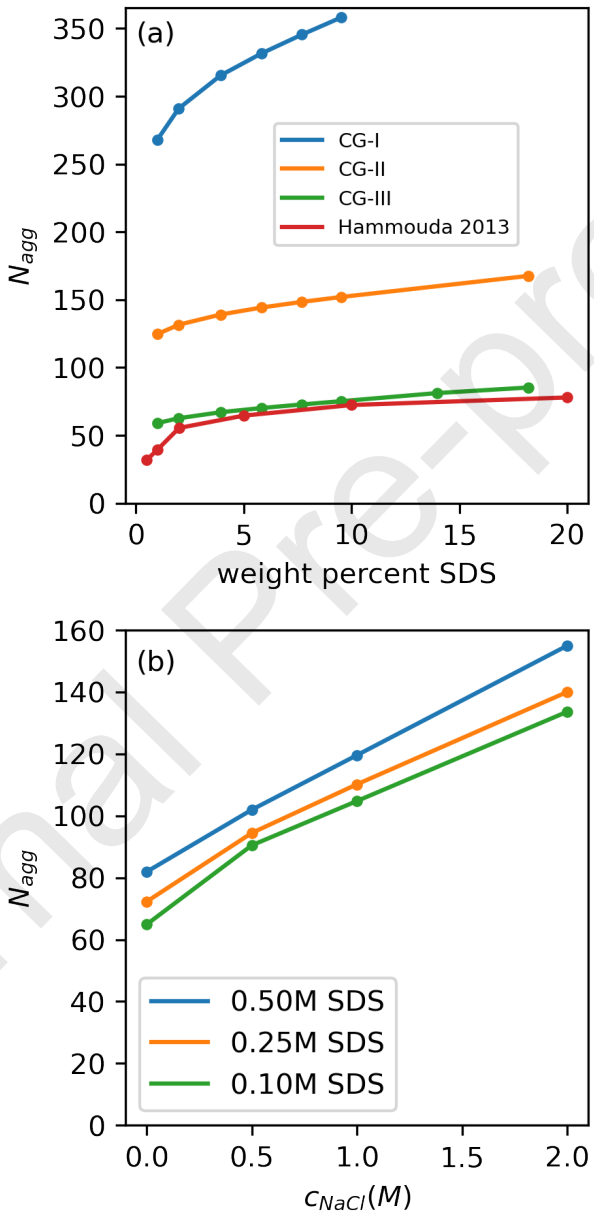



Figure 6: (a) SCFT-predicted micelle aggregation numbers as a function of surfactant concentration for the various models considered. The red line is experimental data from Hammouda 2013 [73] at T=343 K. (b) SCFT-predicted micelle aggregation numbers of model CG-III as a function of NaCl concentration. At fixed SDS molar concentration the weight percentage changes slightly with changing salt content, with 0.10M, 0.25M, 0.50M SDS corresponding to 2.7 ~ 2.9 wt%, 6.8 ~ 7.3 wt%, 13.5 ~ 14.5 wt% SDS, respectively.

screening is no longer the dominant effect, and whether salt promotes or suppresses aggregation instead depends on more chemistry-specific effects of salt-induced hydration and dehydration of micelles. In future work it would be interesting to see if systematic coarse-graining can capture this rich phenomenon in nonionic surfactants.

We remark that many CGMD studies have found similarly good agreement between predicted and experimental aggregation numbers, albeit at room temperature where most CGMD studies have been conducted [9, 13, 21, 22, 77]. One important distinction should be highlighted. Most CGMD studies of the aggregation number use large cell calculations that yield information regarding the distribution of micelle sizes. In contrast, in the SCFT calculations of this work we perform free energy minimization over the size of a single micelle at constant overall surfactant concentration; this is analogous to a maximum term approximation of the most probable micelle size. If a full micellar distribution were required of the SCFT calculations, one would instead have to evaluate the free energy of different sized micelles at constant chemical potential, which is beyond the scope of this work.

### *3.4. SCFT Studies of Morphological Transitions*

Having established the fidelity of the CG model and SCFT to realistic AA models and experimental results, we conclude by leveraging the field theory to evaluate free energies across the full composition range to probe SDS self-assembled surfactant mesostructures.

The experimental phase diagram observes a rich "intermediate phases" regime between 60 wt% and 70 wt%, with several unidentified but partially-characterized and even coexisting phases [55]. To this end, in addition to considering spherical (sph), cylindrical (cyl), inverse cylindrical (cyl-inv), and lamellar (lam) phases, we also consider the Ia$\bar{3}$d double gyroid phase (gyr), which is commonly observed in diblock copolymer systems, and the $I4_132$ alternating gyroid (alt-gyr) cubic phase, which is consistent with some of the partial scattering data observed in experiments on SDS [78]. By comparing free energy densities, we predict the equilibrium, stable self-assembled morphologies. Further, the free energy calculations elucidate the energy difference between different morphologies, and quantify the relative stability of metastable morphologies that in practice may become kinetically trapped.

Given the previously observed discrepancy in self-assembled structures for model CG-I, it is instructive to consider the energetics of self-assembly of this model, as predicted by SCFT (results presented in Figure S4). SCFT predicts that bilayer structures are the globally stable morphology, down to around 2 wt% SDS. This is in accord with earlier observations that model CG-I does not maintain wormlike micelles at zero salt, and instead forms bilayer structures at 30 wt%. In fact, at 30 wt% the SCFT free energy difference between bilayer/lamellar and cylindrical structures is around 0.1 $k_BT$/nm$^3$. For the simulation volumes considered in the wormlike micelle MD simulations, this would translate to around a 40 $k_BT$ difference between the two morphologies.

Next, we consider the zero salt phase diagram of model CG-III, thus far the best-performing and most transferable of the considered CG models. In Figure 7a, we show the SCFT free energy densities for each morphology relative to the structureless disordered phase under salt free conditions. A free energy density less than zero indicates that there is a driving force for self-assembly. Representative morphologies obtained from the field theory simulations are shown in Figure 3.

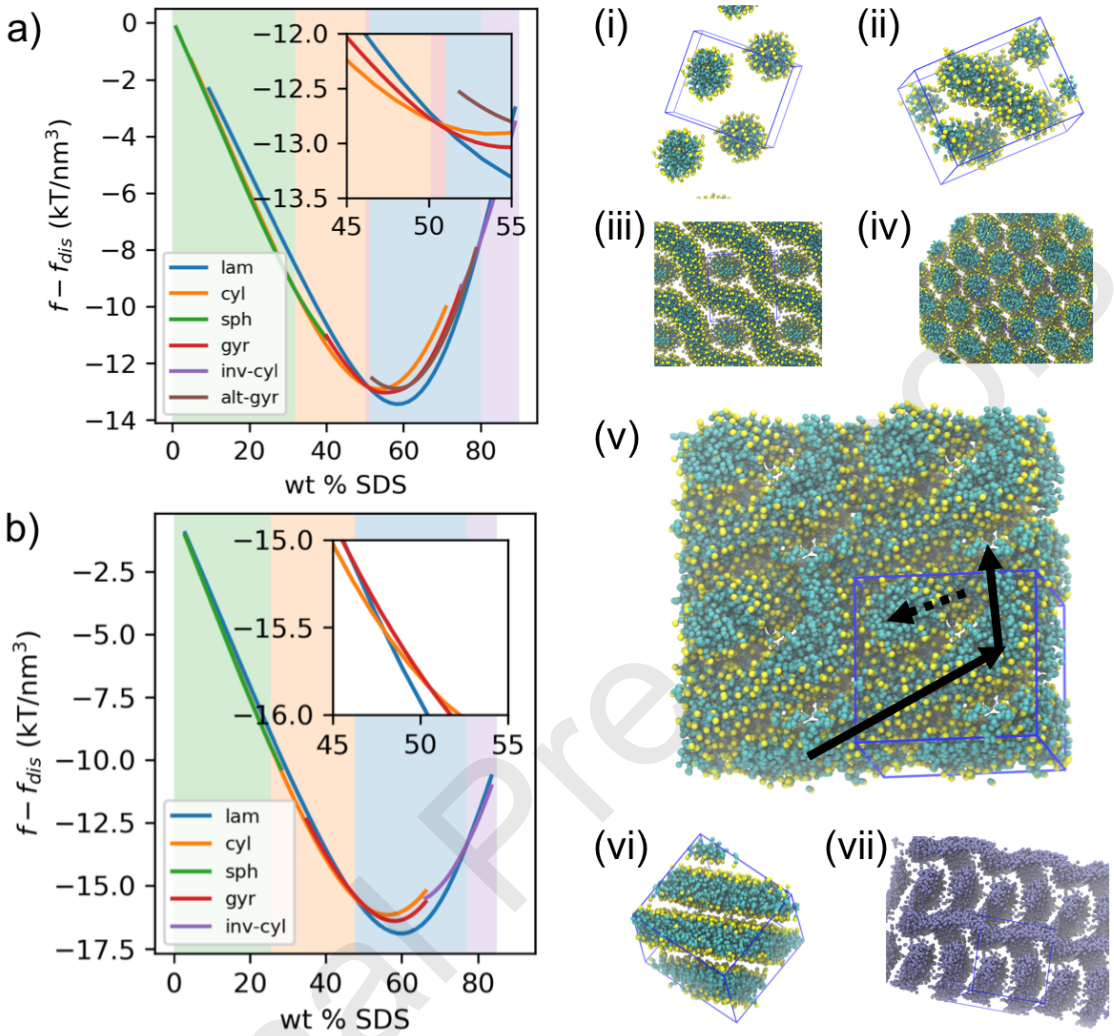



Figure 7: SCFT-calculated free energy densities at zero salt for model CG-III under (a) salt-free conditions and (b) 0.50 M NaCl added salt. The disordered (unstructured) free energy density is subtracted as a reference value. Insets are the same free energy densities, but with the domain chosen to highlight the transition from cylindrical to lamellar structures. On the right (i-vii) are CGMD snapshots of model CG-III with zero added salt at (i) 10 wt%, (ii) 20 wt%, (iii) 30 wt%, (iv) 45 wt%, (v) 50 wt%, (vi) 60 wt%, (vii) 80 wt%. The blue bounding boxes in each snapshot indicate the periodic cell. The black arrows on the 50 wt% snapshot are guides to the eye meant to indicate two cylindrical networks weaving by one another. At 80 wt% it becomes clearer to visualize the water channels (here represented by the blue beads) instead of visualizing the surrounding SDS tubes.

In tandem, in Figures 7i-vii, we present CGMD snapshots of the same model (relaxed from initially random configurations) at various weight percentages to judge the correspondence between CGMD and SCFT morphological predictions.

First, a transition from spherical to cylindrical micelles is estimated by SCFT to occur at around 30 wt% (Fig. 7a). For comparison, experimentally, at room temperature the 2nd CMC for SDS (characterizing the elongation of micelles) is taken to be around 13 wt% [68, 79], while other

simulations models have predicted it to occur somewhere between 15 and 20 wt% [9]. At the higher temperatures (T=348 K) considered in this work, aggregation numbers are expected to be smaller than at room temperature and correspondingly the 2nd CMC ought to be higher. Our observed transition is thus in qualitative agreement with prior work and expectations, but likely overestimates the transition. Although free energy analysis with SCFT predicts a transition from discrete, spherical micelles to infinite, cylindrical micelles, we are unable to converge ellipsoidal or elongated micellar shapes around the transition composition—non-spherical shapes may require fully fluctuating field-theoretic simulations.

CGMD simulations of model CG-III observe the emergence of discrete, ellipsoidal micelles at 20 wt% (Fig. 7ii), suggesting that the 2nd CMC of this system indeed occurs at a lower wt% closer to values observed at 25$^o$C. At 30 wt%, around where the SCFT-predicted transition occurs, infinitely long wormlike micelles (*via* periodic images) appear in the CGMD (Fig. 7iii); these wormlike micelles do not align within the simulation time and remain transverse to one another. Thus, the SCFT-predicted shape transition potentially correlates to the transition from discrete to infinitely long micelles, instead of when discrete ellipsoidal micelles appear. Additionally, between 40 and 45 wt% (Fig. 7iv) the CGMD simulations begin exhibiting hexagonal packing of cylinders, which is still consistent with SCFT (Fig. 7a). Promisingly, experimentally the transition from wormlike to hexagonally packed cylinders occurs at around 40 wt% [55], which is in good agreement with the CGMD simulations.

Next, we investigate the crossover from cylindrical to intermediate structures. There is a very narrow window around 50-51 wt% where the double gyroid is predicted to be stable by SCFT, and very close in free energy to both cylindrical and lamellar structures (Fig. 7a). Again, CGMD simulations show approximately the same shape transitions. At 50 wt% (Fig. 7v) the CGMD simulations appear to exhibit an interpenetrating network of cylinders. Although the CGMD simulation does not relax into a double gyroid structure, this may be because the simulation cell is incommensurate with the equilibrium cell size for double gyroids. The SCFT-predicted free energetic driving force for forming a double gyroid network relative to other structures is also very small, suggesting that it may be very hard to equilibrate and observe in MD. The SCFT-predicted window of stability is also very small and probably shifted by fluctuation effects intrinsically included in CGMD, making the double gyroid window even harder to pinpoint in CGMD. Nevertheless, the observed CGMD morphology is consistent with the SCFT prediction that hexagonally packed cylinders are no longer the favored morphology, and the CGMD morphology also more closely resembles the double gyroid structure than well-aligned, hexagonally packed cylinders.

Following the gyroid phase, SCFT predicts that the next globally stable structure is the lamellar phase (Fig. 7a). Correspondingly, at 55 wt% (not shown), the CGMD simulations show signs of adjacent cylindrical structures fusing and forming flatter, strut-like structures. It is not until 60 wt% (Fig. 7vi) that the CGMD simulation unambiguously forms lamellar structures. This is also around the weight fraction that maximizes the SCFT-predicted free energy stability of lamellar structures relative to the double gyroid (Fig. 7a). Experimentally, the lamellar transition is observed around 70 wt% [55]. The lamellar transition is thus underpredicted by somewhere between 10 and 20 wt% by CGMD and SCFT, respectively. There are several potential reasons for this underestimation and lack of transferability, which will be discussed in Section 4. Finally, at higher weight percentage of SDS, the predicted stable SCFT morphology features inverse cylindrical structures, where it

is water that forms the channels while SDS bilayers fill the interstices between them (Fig. 7a). CGMD at 80 wt% (Fig. 7vii) confirms this SCFT prediction.

In SCFT we are also able to trap metastable alternating gyroid structures (Fig. 3d) of two networks (one formed by cylindrical structures with an alkyl core and a second network formed by water channels surrounded by a SDS bilayer). Although SCFT predicts this phase is less stable than lamellar structures at all concentrations considered, it does outcompete double gyroid structures above 60 wt% (Fig. 7a), consistent with experimental observations of unidentified intermediate phases in the window from 60-70 wt% [78]. The reasonable qualitative agreement of the location of this alternating gyroid phase with experiment suggests that it could be one of the intermediate phases. It is possible that future refinement of the coarse-grained model could globally stabilize the alternating gyroid phase over the lamellar phase.

The structures and transitions observed in our CGMD simulations of SDS with zero added salt are also largely in agreement with those observed in CGMD studies using other CG models of surfactants, following the same sequence of structures as the concentration of surfactant increases [23, 24]. Furthermore, the transitions happen at roughly the same concentrations as well. For instance, when comparing to recent simulations of several dissipative particle dynamics models of SDS at room temperature, the transitions for our model (their model) occur at: micellar → ellipsoidal at 20 wt% (not reported) → cylindrical at 30 wt% (35 wt%) → hexagonal at 40-45 wt% (40-45 wt%) → bicontinuous at 50 wt% (50-60 wt%) → lamellar at 60 wt% (75-80 wt%) as the concentration of surfactant increases [7]. The largest discrepancy in our model is for the lamellar transition, as discussed previously when comparing to experiment.

Next, the same trends are largely preserved upon the addition of 0.5M NaCl salt (Figure 7b). Relative to the salt-free case (Fig. 7a), all the transitions are decreased by around 3-5 wt%. The transition from discrete micelles to infinitely long cylindrical micelles now happens at around 25 wt%, and the cylinder-lamellar transition occurs at around 48 wt%. While the crossover in free energy densities between cylindrical and gyroidal structures remains at around 51 wt%, the lamellar phase is stabilized sufficiently such that the double gyroidal phase is never globally stable. Again, these shifts in the phase boundary are in qualitative agreement with physical expectations. In Figure 6b, we previously demonstrated that increasing salt concentration increases the aggregation number of micelles. This increased micelle size is expected to lead to an earlier transition to wormlike and lamellar micelles. Alternatively, at constant surfactant volume $v$ and tail length $l$, decreasing the headgroup area (*via* increased screening from added salt) increases the packing parameter $p = v/(a \cdot l)$ [66]. Since the packing parameter theory expects the morphology sequence to follow sphere → cylinder → lamellae → inverse structures as the packing parameter increases, increasing salt pushes the transitions in the same (and observed) direction.

These observations demonstrate that our coarse-grained models yield reasonable morphological transitions, and that CGMD and SCFT can be used in a complementary way on the same underlying microscopic model. We observe that in regions where the SCFT-predicted driving force for forming a particular morphology is small, the CGMD simulations also exhibit intermediate morphologies, perhaps because the driving force towards the global equilibrium structure is small or that there are many metastable states of comparable energy. Due to equilibration challenges, self-assembled structures in MD are usually full of defects, possibly stressed in an incommensurate cell, subject to hysteresis, and hence phase transitions can be difficult to resolve or interpret.

SCFT serves as an efficient way to understand and even quantify the competitive landscape of self-assembled structures.

Finally, to better understand the morphologies, in Figure S5 we examine the characteristic spacing between SDS structures. For discrete, spherical structures, we define the characteristic spacing to be $d = V^{1/3}$, where V is the volume of the periodic cell, or roughly the available translational volume per micelle. For cylindrical structures, this spacing $d$ is taken to be the primitive cell length of the hexagonal unit cell $d = l$. The characteristic spacing for lamellar structures and inverse-cylindrical structures are similarly defined. For the double gyroid structure, because there are two networks per primitive cell, we define $d = l/2$ where $l$ is the side length of the double gyroid primitive cell. For the alternating gyroid, there are effectively 3 SDS "layers" as one moves along a primitive cell vector: one from the SDS-filled cylinder, and two bilayer structures from SDS wrapping around the water channel. We hence define $d = l/3$ for the alternating gyroid. Because the geometry of each phase is different, we expect these definitions to only be a rough guide for comparing spacings between phases.

Nevertheless, we make the important *qualitative* observation that the characteristic spacing for each morphology is non-monotonic with respect to the SDS wt%, just as the free energy curves for each morphology in Figure 7 appear parabolic with increasing SDS content. This non-monotonic behavior indicates that upon adding SDS, the characteristic spacing of a given morphology cannot decrease indefinitely, and the domains start swelling beyond a certain concentration of SDS. To continue decreasing the characteristic spacing as additional surfactant is added, morphological transitions are necessary. Correspondingly, the results suggest that morphological transitions roughly correlate with whichever morphology is best able to minimize the characteristic spacing, and the sequence of phases that minimize spacing follows the same ordering as predicted by the free energy calculations (albeit quantitatively at slightly lower concentrations).

## 4. Conclusion

Surfactant self-assembly is highly sensitive to the chemical structure of the surfactants: for example, the addition of a couple atoms can significantly change phase behavior [80, 81]. Geometric and thermodynamic theories have been developed to understand the general behavior of surfactants in terms of a few key quantities like the headgroup area, tail length, and the surfactant transfer free energy between free solution and micelle [66, 82]. However, it has always been a challenge to connect underlying chemistry to physically interpretable parameters, and approaches range from using group contribution methods to extensive experimental calibration [82, 83]. We show that without resorting to increasingly complicated potentials or coarse-graining strategies [84], our multiscale modeling approach can provide a systematic path towards quantifying how underlying chemical detail manifests in terms of the effective interactions of a bead spring model. Although simple, the model can be predictive, much in the spirit of how surfactant theories identified transparent and interpretable quantities governing surfactant behavior.

Our modeling approach's ability to make *de novo* predictions of surfactant self-assembled morphologies stems from its use of all-atom (AA) simulations as the only input for chemical detail. The approach leverages AA trajectories as chemically-detailed reference systems, and uses the relative entropy to incorporate this chemical detail into coarse-grained (CG) models that can be

simulated by field theory. The resulting molecularly informed field theory provides efficient access to thermodynamic quantities relevant to the calculation of equilibrium structures. We showed that coarse-graining from a micellar structure does not yield a reasonable CG model. One potential way to improve upon this model is to study an extended ensemble of micellar references, although this would introduce many additional considerations for what constitutes an appropriate selection of micelle structures. Instead, we showed that a straightforward procedure of coarse-graining from SDS at a water-dodecane interface by minimizing the relative entropy with constraints on the surfactant headgroup area and tail length can result in reasonable surfactant models that are transferable across compositions.

The best CG model (CG-III) reproduces surfactant headgroup area and tail lengths at an interface to within 1%, while wormlike micelle linear densities are reproduced to within 10-25%. The resulting surfactant model at 348 K quantitatively matches experimental aggregation numbers, and also reproduces expected experimental trends: the aggregation number increases with increasing surfactant and salt concentrations. This demonstrates the ability of the proposed coarse-graining approach to produce CG surfactant models that are reasonably predictive for micellar properties, *without* the use of explicit experimental data during the coarse-graining process. Part of this agreement stems from the fact that the AA force field quantitatively reproduces the SDS surface tension, but the performance of the CG models also relies on the coarse-graining procedure preserving the complex interplay of interactions present in the AA model. While the current workflow is able to make *a priori* predictions, in the future, experimentally available quantities (*e.g.* on aggregation numbers) can also be used to further constrain and refine the CG models.

At higher densities, the CG model reasonably describes the sequence of morphological transitions, and we show that although CGMD simulations are difficult to equilibrate, the observed morphologies nevertheless are broadly consistent with SCFT simulations of the same model. This demonstrates that despite the mathematical approximations inherent in SCFT calculations, it can still be profitably used to understand and predict the self-assembly of even small molecule surfactant systems. For example, we are able to isolate a novel alternating gyroid morphology by the field theory and locate it where experiments expect to see cubic structures [78]. However, the alternating gyroid was only metastable in the SCFT, and further work is needed to see if fluctuations or a more refined CG model would be able to globally stabilize this morphology.

Further comparing with experiment, the hexagonal ordering transition observed in CGMD is within 5% of experiment, whereas the lamellar transition is about 10-15 wt% lower than expected in experiment. There are several potential reasons why the lamellar phase may be overstabilized at higher concentrations in the CG models, and these reasons provide avenues of future research. First, our CG models are developed at lower concentrations. This may explain their good performance at reproducing micellar properties at lower concentrations, but they exhibit decreased transferability at concentrations above 50 wt% SDS. Secondly, consistent with a model developed at lower concentrations, the electrostatic interaction employed in our model uses a constant background dielectric set by pure water. This is correct in the low concentration limit when the solution is mostly water, but does not accurately describe microphase-forming SDS phases, where much of the mass comes from alkyl tails, which have a much lower dielectric constant (expected to be around 2) [85, 86, 87]. As an alternative to assuming a constant background dielectric, one could coarse-grain to a polarizable bead model that gives the correct effective dielectric constant in the

water and alkyl phases [88, 89, 90]. This would better describe the changing dielectric properties as SDS is added, and may lead to more accurate model predictions at higher SDS concentrations.

Additionally, the model may be further refined by introducing additional bead types, most likely a new bead type to describe the last bead of the SDS tail and associated chain-end effects, as well as an additional bead type to describe the junction between the aklyl tail and the surfactant headgroup. The introduction of just two additional bead types would introduce 13 additional nonbonded interactions to the system, which may give it additional representational power to more accurately describe self-assembly at higher concentrations [91].

Another rich direction for exploration is to make fuller use of experimental data in guiding model design. For example, SDS exhibits rich behavior in the intermediate 60-70 wt% composition range. To better resolve these intricate phases, one could attempt to coarse grain directly at the higher concentrations of interest. However, this approach is challenged by the difficulty of obtaining appropriate, equilibrated atomistic trajectories to coarse grain from. In this case, additional experimental data would be invaluable. For example, if space groups were unambiguously resolved for the phases in this intermediate window, the identified space groups can be applied as constraints to evaluate the stability of the phases and facilitate exploration of coarse-grained models to stabilize the target structures. Yet another approach would be to apply additional constraints on the coarse-grained model, such as requiring the coarse-grained surfactant to reproduce bending energies of self-assembled bilayers or wormlike micelles (obtained either by experiment or all-atom simulation). Lastly, another important area for connecting with experimental results is to re-coarse grain the model with different salts to assess salt-specific effects, or re-coarse grain at multiple temperatures to assess effects on micelle morphology at lower temperatures. This will be pursued in future work.

In summary, we have shown that systematic bottom-up coarse-graining can be used to derive, *de novo*, CG models that can be effectively simulated in field theory to study self-assembly behaviors. Our approach demonstrates how field theoretic simulations can complement CGMD simulations *of the same molecular model*, and enables new capabilities like easily evaluating free energies and equilibrating the preferred size and symmetry of ordered microstructures. Importantly, chemical details representing the complex interplay of ionic and hydrophobic interactions in the AA model are retained in the CG model, resulting in at times near-quantitative reproduction of surfactant properties, comparable to an extensive literature of CGMD studies of surfactants [7, 9, 21, 23, 77]. We believe this work is an important step towards the accurate and efficient simulation of the self-assembly behaviors of realistic formulations, which are usually highly multicomponent and depend on the delicate balance of a myriad of interactions. Future work will develop methodologies for calculating additional properties of interest like the critical micelle concentration, as well as combine the surfactant model developed in this work with additional components, such as PEO [34], polyelectrolytes [92], and other small molecule additives.

## CRediT authorship contribution statement

**Kevin Shen:** Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, project administration, writing - original draft, writing - review and editing, visualization. **My Nguyen:** Conceptualization, methodology, software, validation,

writing - review and editing. **Nick Sherck:** Conceptualization, methodology, software, validation, writing - review and editing, funding acquisition. **Brian Yoo, Stephan Kohler, Joshua Speros, Kris T. Delaney, M. Scott Shell, Glenn H. Fredrickson:** Conceptualization, resources, writing - review and editing, supervision, project administration, funding acquisition.

## Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by BASF Corporation through the California Research Alliance. BY, SK, JS from BASF helped conceptualize the research such that it would be relevant to engineering efforts at BASF. GHF and KTD also derived partial support from the National Science Foundation CMMT Program under grant number DMR-2104255. KS also received support from the BioPACIFIC Materials Innovation Platform (NSF DMR-1933487). Use was made of computational facilities purchased with funds from the National Science Foundation (OAC-1925717) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR-1720256) at UC Santa Barbara.

## Supplementary Information and Data Files

Additional figures and force field parameters are included in supplementary info.

## References

[1] L. Cao, D. Russo, A. A. Lapkin, Automated robotic platforms in design and development of formulations, AIChE Journal 67 (5) (2021) e17248.

[2] Y. Fan, C.-W. Yen, H.-C. Lin, W. Hou, A. Estevez, A. Sarode, A. Goyon, J. Bian, J. Lin, S. G. Koenig, et al., Automated high-throughput preparation and characterization of oligonucleotide-loaded lipid nanoparticles, International Journal of Pharmaceutics 599 (2021) 120392.

[3] A. Davies, S. Amin, Microstructure design of ctac: Fa and btac: Fa lamellar gels for optimized rheological performance utilizing automated formulation platform, International Journal of Cosmetic Science 42 (3) (2020) 259–269.

[4] T. E. Gartner III, A. Jayaraman, Modeling and simulations of polymers: a roadmap, Macromolecules 52 (3) (2019) 755–786.

[5] X. Tang, P. H. Koenig, R. G. Larson, Molecular dynamics simulations of sodium dodecyl sulfate micelles in water the effect of the force field, The Journal of Physical Chemistry B 118 (14) (2014) 3864–3880.

[6] S. D. Peroukidis, D. G. Tsalikis, M. G. Noro, I. P. Stott, V. G. Mavrantzas, Quantitative prediction of the structure and viscosity of aqueous micellar solutions of ionic surfactants: a combined approach based on coarse-grained martini simulations followed by reverse-mapped all-atom molecular dynamics simulations, Journal of Chemical Theory and Computation 16 (5) (2020) 3363–3372.

[7] M. Choudhary, S. Kamil, Phase diagram study of sodium dodecyl sulfate using dissipative particle dynamics, ACS omega 5 (36) (2020) 22891–22900.

[8] T. Taddese, R. L. Anderson, D. J. Bray, P. B. Warren, Recent advances in particle-based simulation of surfactants, Current Opinion in Colloid & Interface Science 48 (2020) 137–148.
[9] R. L. Anderson, D. J. Bray, A. Del Regno, M. A. Seaton, A. S. Ferrante, P. B. Warren, Micelle formation in alkyl sulfate surfactants using dissipative particle dynamics, Journal of Chemical Theory and Computation 14 (5) (2018) 2633–2643.
[10] W. Shinoda, R. DeVane, M. L. Klein, Coarse-grained force field for ionic surfactants, Soft Matter 7 (13) (2011) 6178–6186.
[11] J. Saathoff, Effectively parameterizing dissipative particle dynamics using cosmo-sac: A partition coefficient study, The Journal of Chemical Physics 148 (15) (2018) 154102.
[12] J. G. Fraaije, J. van Male, P. Becherer, R. Serral Gracià, Coarse-grained models for automated fragmentation and parametrization of molecular databases, Journal of Chemical Information and Modeling 56 (12) (2016) 2361–2377.
[13] R. Mao, M.-T. Lee, A. Vishnyakov, A. V. Neimark, Modeling aggregation of ionic surfactants using a smeared charge approximation in dissipative particle dynamics simulations, The Journal of Physical Chemistry B 119 (35) (2015) 11673–11683.
[14] S. Wang, R. G. Larson, Coarse-grained molecular dynamics simulation of self-assembly and surface adsorption of ionic surfactants using an implicit water model, Langmuir 31 (4) (2015) 1262–1271.
[15] T. Ohkuma, K. Kremer, A composition transferable and time-scale consistent coarse-grained model for cis-polyisoprene and vinyl-polybutadiene oligomeric blends, Journal of Physics: Materials 3 (3) (2020) 034007.
[16] B. G. Levine, D. N. LeBard, R. DeVane, W. Shinoda, A. Kohlmeyer, M. L. Klein, Micellization studied by gpu-accelerated coarse-grained molecular dynamics, Journal of Chemical Theory and Computation 7 (12) (2011) 4135–4145.
[17] D. N. LeBard, B. G. Levine, P. Mertmann, S. A. Barr, A. Jusufi, S. Sanders, M. L. Klein, A. Z. Panagiotopoulos, Self-assembly of coarse-grained ionic surfactants accelerated by graphics processing units, Soft Matter 8 (8) (2012) 2385–2397.
[18] E. Lavagnini, J. L. Cook, P. B. Warren, C. A. Hunter, Translation of chemical structure into dissipative particle dynamics parameters for simulation of surfactant self-assembly, The Journal of Physical Chemistry B 125 (15) (2021) 3942–3952.
[19] A. Del Regno, P. B. Warren, D. J. Bray, R. L. Anderson, Critical micelle concentrations in surfactant mixtures and blends by simulation, The Journal of Physical Chemistry B 125 (22) (2021) 5983–5990.
[20] S. J. Marrink, A. H. De Vries, A. E. Mark, Coarse grained model for semiquantitative lipid simulations, The Journal of Physical Chemistry B 108 (2) (2004) 750–760.
[21] G. Pérez-Sánchez, N. Schaeffer, T. L. Greaves, J. F. Pereira, J. A. Coutinho, Tuning the ionic character of sodium dodecyl sulphate via counter-ion binding: An experimental and computational study, Frontiers in Materials 9 (2022) 1011164.
[22] S. Illa-Tuset, D. C. Malaspina, J. Faraudo, Coarse-grained molecular dynamics simulation of the interface behaviour and self-assembly of ctab cationic surfactants, Physical Chemistry Chemical Physics 20 (41) (2018) 26422–26430.
[23] S.-C. Chien, G. Pérez-Sánchez, J. R. Gomes, M. N. D. Cordeiro, M. Jorge, S. M. Auerbach, P. A. Monson, Molecular simulations of the synthesis of periodic mesoporous silica phases at high surfactant concentrations, The Journal of Physical Chemistry C 121 (8) (2017) 4564–4575.
[24] S. D. Anogiannakis, P. C. Petris, D. N. Theodorou, Promising route for the development of a computational framework for self-assembly and phase behavior prediction of ionic surfactants using martini, The Journal of Physical Chemistry B 124 (3) (2019) 556–567.
[25] S.-J. Marrink, D. P. Tieleman, Molecular dynamics simulation of a lipid diamond cubic phase, Journal of the American Chemical Society 123 (49) (2001) 12383–12391.
[26] V. Kocherbitov, Molecular dynamics simulations of liquid crystalline phases of dodecyltrimethylammonium chloride, Journal of Molecular Liquids 210 (2015) 74–81.
[27] S. Sahu, N. S. Schwindt, B. J. Coscia, M. R. Shirts, Obtaining and characterizing stable bicontinuous cubic morphologies and their nanochannels in lyotropic liquid crystal membranes, The Journal of Physical Chemistry B (2022).

[28] G. Rossi, P. Fuchs, J. Barnoud, L. Monticelli, A coarse-grained martini model of polyethylene glycol and of polyoxyethylene alkyl ether surfactants, The Journal of Physical Chemistry B 116 (49) (2012) 14353–14362.
[29] B. Wen, B. Bai, R. G. Larson, Surfactant desorption and scission free energies for cylindrical and spherical micelles from umbrella-sampling molecular dynamics simulations, Journal of Colloid and Interface Science 599 (2021) 773–784.
[30] G. Fredrickson, et al., The equilibrium theory of inhomogeneous polymers, Oxford University Press on Demand, 2006.
[31] G. H. Fredrickson, K. T. Delaney, Direct free energy evaluation of classical and quantum many-body systems via field-theoretic simulation, Proceedings of the National Academy of Sciences 119 (18) (2022) e2201804119.
[32] K. T. Delaney, G. H. Fredrickson, Recent developments in fully fluctuating field-theoretic simulations of polymer melts and solutions, The Journal of Physical Chemistry B 120 (31) (2016) 7615–7634.
[33] M. Matsen, Field theoretic approach for block polymer melts: Scft and fts, The Journal of Chemical Physics 152 (11) (2020) 110901.
[34] N. Sherck, K. Shen, M. Nguyen, B. Yoo, S. Kohler, J. C. Speros, K. T. Delaney, M. S. Shell, G. H. Fredrickson, Molecularly informed field theories from bottom-up coarse-graining, ACS Macro Letters 10 (5) (2021) 576–583.
[35] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, H. C. Andersen, The multi-scale coarse-graining method. i. a rigorous bridge between atomistic and coarse-grained models, The Journal of Chemical Physics 128 (24) (2008) 244114.
[36] A. P. Lyubartsev, A. Laaksonen, Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach, Physical Review E 52 (4) (1995) 3730.
[37] F. Müller-Plathe, Coarse-graining in polymer simulation: from the atomistic to the mesoscopic scale and back, ChemPhysChem 3 (9) (2002) 754–769.
[38] M. S. Shell, Coarse-graining with the relative entropy, Advances in chemical physics (2016).
[39] G. Kume, M. Gallotti, G. Nunes, Review on anionic/cationic surfactant mixtures, Journal of Surfactants and Detergents 11 (1) (2008) 1–11.
[40] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, et al., Openmm 7: Rapid development of high performance algorithms for molecular dynamics, PLoS Computational Biology 13 (7) (2017) e1005659.
[41] S. Izadi, R. Anandakrishnan, A. V. Onufriev, Building water models: a different approach, The Journal of Physical Chemistry Letters 5 (21) (2014) 3863–3871.
[42] I. S. Joung, T. E. Cheatham III, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, The Journal of Physical Chemistry B 112 (30) (2008) 9020–9041.
[43] P. S. Bulutoglu, S. Wang, M. Boukerche, N. K. Nere, D. S. Corti, D. Ramkrishna, An investigation of the kinetics and thermodynamics of nacl nucleation through composite clusters, PNAS Nexus (2022).
[44] V. S. Farafonov, A. V. Lebed, Developing and validating a set of all-atom potential models for sodium dodecyl sulfate, Journal of Chemical Theory and Computation 13 (6) (2017) 2742–2750.
[45] H. Yan, S.-L. Yuan, G.-Y. Xu, C.-B. Liu, Effect of ca2+ and mg2+ ions on surfactant solutions investigated by molecular dynamics simulation, Langmuir 26 (13) (2010) 10448–10459.
[46] C. J. Dickson, B. D. Madej, Å. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould, R. C. Walker, Lipid14: the amber lipid force field, Journal of Chemical Theory and Computation 10 (2) (2014) 865–879.
[47] S. Abdel-Azeim, Revisiting opls-aa force field for the simulation of anionic surfactants in concentrated electrolyte solutions, Journal of Chemical Theory and Computation 16 (2) (2020) 1136–1145.
[48] J. Aqvist, Ion-water interaction potentials derived from free energy perturbation simulations, The Journal of Physical Chemistry 94 (21) (1990) 8021–8024.
[49] D. Tolmachev, O. Boyko, N. Lukasheva, H. Martinez-Seara, M. Karttunen, Overbinding and qualitative and quantitative changes caused by simple na+ and k+ ions in polyelectrolyte simulations: Comparison of force fields with and without nbfix and ecc corrections, Journal of Chemical Theory and Computation 16 (1) (2019) 677–687.
[50] J. Yoo, A. Aksimentiev, New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions, Physical Chemistry Chemical Physics 20 (13) (2018) 8432–8449.

[51] J. R. Errington, D. A. Kofke, Calculation of surface tension via area sampling, The Journal of Chemical Physics 127 (17) (2007) 174709.
[52] G. J. Gloor, G. Jackson, F. J. Blas, E. de Miguel, Test-area simulation method for the direct determination of the interfacial tension of systems with continuous or discontinuous potentials, The Journal of Chemical Physics 123 (13) (2005) 134703.
[53] K. Shen, N. Sherck, M. Nguyen, B. Yoo, S. Köhler, J. Speros, K. T. Delaney, G. H. Fredrickson, M. S. Shell, Learning composition-transferable coarse-grained models: Designing external potential ensembles to maximize thermodynamic information, The Journal of Chemical Physics 153 (15) (2020) 154116.
[54] P. Espanol, P. B. Warren, Perspective: Dissipative particle dynamics, The Journal of Chemical Physics 146 (15) (2017) 150901.
[55] P. Kékicheff, C. Grabielle-Madelmont, M. Ollivon, Phase diagram of sodium dodecyl sulfate-water system: 1. a calorimetric study, Journal of Colloid and Interface Science 131 (1) (1989) 112–132. doi:https://doi.org/10.1016/0021-9797(89)90151-3.
URL https://www.sciencedirect.com/science/article/pii/0021979789901513
[56] D. P. Bertsekas, Constrained optimization and Lagrange multiplier methods, Academic press, 2014.
[57] N. I. M. Gould, On the convergence of a sequential penalty function method for constrained minimization, SIAM Journal on Numerical Analysis 26 (1) (1989) 107–128.
[58] C. D. Bruce, M. L. Berkowitz, L. Perera, M. D. Forbes, Molecular dynamics simulation of sodium dodecyl sulfate micelle in water: micellar structural characteristics and counterion distribution, The Journal of Physical Chemistry B 106 (15) (2002) 3788–3793.
[59] J. Zhou, A.-C. Shi, Critical micelle concentration of micelles with different geometries in diblock copolymer/homopolymer blends, Macromolecular Theory and Simulations 20 (8) (2011) 690–699.
[60] M. W. Matsen, M. Schick, Stable and unstable phases of a linear multiblock copolymer melt, Macromolecules 27 (24) (1994) 7157–7163.
[61] J. Lu, R. Thomas, J. Penfold, Surfactant layers at the air/water interface: structure and composition, Advances in Colloid and Interface science 84 (1-3) (2000) 143–304.
[62] J. Hines, T. P. Garrett, G. Rennie, J. Penfold, Investigation of mixing in binary surfactant solutions by surface tension and neutron reflection: anionic/nonionic and zwitterionic/nonionic mixtures, Journal of Physical Chemistry B 101 (1997) 9215–9223.
[63] N. Vargaftik, B. Volkov, L. Voljak, International tables of the surface tension of water, Journal of Physical and Chemical Reference Data 12 (3) (1983) 817–820.
[64] J. Penfold, R. K. Thomas, Neutron reflectivity and small angle neutron scattering: An introduction and perspective on recent progress, Current Opinion in Colloid & Interface Science 19 (3) (2014) 198–206.
[65] C. Tanford, The hydrophobic effect: formation of micelles and biological membranes 2d ed, J. Wiley., 1980.
[66] J. N. Israelachvili, D. J. Mitchell, B. W. Ninham, Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers, Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics 72 (1976) 1525–1568.
[67] T. C. Moore, C. R. Iacovella, C. McCabe, Derivation of coarse-grained potentials via multistate iterative boltzmann inversion, The Journal of Chemical Physics 140 (22) (2014) 06B606_1.
[68] F. Reiss-Husson, V. Luzzati, The structure of the micellar solutions of some amphiphilic compounds in pure water as determined by absolute small-angle x-ray scattering techniques, The Journal of physical chemistry 68 (12) (1964) 3504–3511.
[69] R. Becker, W. Döring, Kinetische behandlung der keimbildung in übersättigten dämpfen, Annalen der physik 416 (8) (1935) 719–752.
[70] J. A. Mysona, A. V. McCormick, D. C. Morse, Mechanism of micelle birth and death, Physical Review Letters 123 (3) (2019) 038003.
[71] J. A. Mysona, A. V. McCormick, D. C. Morse, Nonlinear dynamics in micellar surfactant solutions. i. kinetics, Physical Review E 105 (3) (2022) 034602.
[72] E. Aniansson, S. N. Wall, Kinetics of step-wise micelle association, The Journal of Physical Chemistry 78 (10) (1974) 1024–1030.
[73] B. Hammouda, Temperature effect on the nanostructure of sds micelles in water, Journal of Research of the

National Institute of Standards and Technology 118 (2013) 151.
[74] R. Ranganathan, L. Tran, B. L. Bales, Surfactant-and salt-induced growth of normal sodium alkyl sulfate micelles well above their critical micelle concentrations, The Journal of Physical Chemistry B 104 (10) (2000) 2260–2264.
[75] P. Mukerjee, The nature of the association equilibria and hydrophobic bonding in aqueous solutions of association colloids, Advances in Colloid and Interface Science 1 (3) (1967) 242–275.
[76] S. Miyagishi, K. Okada, T. Asakawa, Salt effect on critical micelle concentrations of nonionic surfactants, n-acyl-n-methylglucamides (mega-n), Journal of Colloid and Interface Science 238 (1) (2001) 91–95.
[77] W. C. Swope, M. A. Johnston, A. I. Duff, J. L. McDonagh, R. L. Anderson, G. Alva, A. T. Tek, A. P. Maschino, Challenge to reconcile experimental micellar properties of the cnem nonionic surfactant family, The Journal of Physical Chemistry B 123 (7) (2019) 1696–1707.
[78] P. Kekicheff, B. Cabane, Between cylinders and bilayers: Structures of intermediate mesophases of the sds/water system, Journal de Physique 48 (9) (1987) 1571–1583.
[79] N. Christov, N. Denkov, P. Kralchevsky, K. Ananthapadmanabhan, A. Lips, Synergistic sphere-to-rod micelle transition in mixed solutions of sodium dodecyl sulfate and cocoamidopropyl betaine, Langmuir 20 (3) (2004) 565–571.
[80] F. Broome, C. Hoerr, H. Harwood, The binary systems of water with dodecylammonium chloride and its n-methyl derivatives, Journal of the American Chemical Society 73 (7) (1951) 3350–3352.
[81] A. V. Peresypkin, F. M. Menger, Zwitterionic geminis. coacervate formation from a single organic compound, Organic Letters 1 (9) (1999) 1347–1350.
[82] R. Nagarajan, E. Ruckenstein, Theory of surfactant self-assembly: a predictive molecular thermodynamic approach, Langmuir 7 (12) (1991) 2934–2969.
[83] P. Rehner, B. Bursik, J. Gross, Surfactant modeling using classical density functional theory and a group contribution pc-saft approach, Industrial & Engineering Chemistry Research 60 (19) (2021) 7111–7123.
[84] T. Sanyal, M. S. Shell, Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation, The Journal of chemical physics 145 (3) (2016) 034109.
[85] A. Sen, V. Anicich, T. Arakelian, Dielectric constant of liquid alkanes and hydrocarbon mixtures, Journal of Physics D: Applied Physics 25 (3) (1992) 516.
[86] T. Simonson, C. L. Brooks, Charge screening and the dielectric constant of proteins: insights from molecular dynamics, Journal of the American Chemical Society 118 (35) (1996) 8452–8458.
[87] B. J. Kirby, P. Jungwirth, Charge scaling manifesto: A way of reconciling the inherently macroscopic and microscopic natures of molecular simulations, The Journal of Physical Chemistry Letters 10 (23) (2019) 7531–7536.
[88] J. M. Martin, W. Li, K. T. Delaney, G. H. Fredrickson, Statistical field theory description of inhomogeneous polarizable soft matter, The Journal of Chemical Physics 145 (15) (2016) 154104.
[89] D. J. Grzetic, K. T. Delaney, G. H. Fredrickson, Field-theoretic study of salt-induced order and disorder in a polarizable diblock copolymer, ACS Macro Letters 8 (8) (2019) 962–967.
[90] D. J. Grzetic, K. T. Delaney, G. H. Fredrickson, Contrasting dielectric properties of electrolyte solutions with polar and polarizable solvents, Physical Review Letters 122 (12) (2019) 128007.
[91] M. E. Johnson, T. Head-Gordon, A. A. Louis, Representability problems for coarse-grained water potentials, The Journal of Chemical Physics 126 (14) (2007) 144509.
[92] M. Nguyen, N. Sherck, K. Shen, B. Yoo, S. Köhler, J. Speros, K. T. Delaney, G. H. Fredrickson, M. S. Shell, Predicting polyelectrolyte coacervation from amolecularly informed field-theoretic model, In Preparation (2022).

# Supporting Information:

# Predicting Surfactant Phase Behavior with a Molecularly Informed Field Theory

Kevin Shen,[*,†] My Nguyen,[†] Nicholas Sherck,[†] Brian Yoo,[¶] Stephan Köhler,[§] Joshua Speros,[‖] Kris T. Delaney,[*,‡] M. Scott Shell,[*,⊥] and Glenn H. Fredrickson[*,†,‡,#]

†*Department of Chemical Engineering, University of California, Santa Barbara, California 93106, United States*

‡*Materials Research Laboratory, University of California, Santa Barbara, California 93106, United States*

¶*Fero Labs, 520 Broadway, New York, New York 10012, United States*

§*BASF SE, Ludwigshafen am Rhein 67056, Germany*

‖*California Research Alliance (CARA) by BASF, Berkeley, California 94720, United States*

⊥*Department of Chemical Engineering, University of California, Santa Barbara, California, 93106, United States*

#*Department of Materials Engineering, University of California, Santa Barbara, California 93106, United States*

E-mail: kevinshen@ucsb.edu; kdelaney@ucsb.edu; shell@ucsb.edu; ghf@ucsb.edu

## Contents

# 1 All atom force field development

## 1.1 Surface tension and all atom morphologies

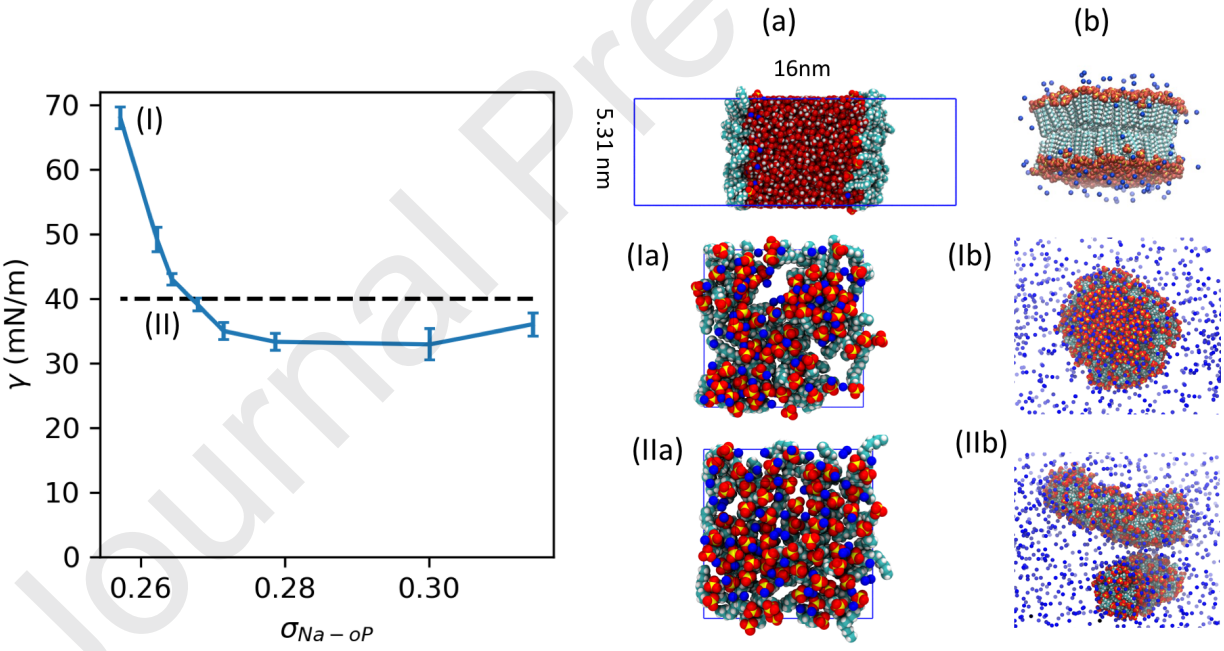



Figure S1: Room temperature water-air surface tension at a surfactant coverage of $0.44/\text{nm}^2$, as a function of the Lennard-Jones interaction diameter $\sigma$ between sodium ions and oxygens with atom type oP. Dashed line is the experimental surface tension of SDS. Representative morphologies are presented for (I) unmodified force field, (II) force field parameter used in this work ($\sigma = 0.2679$nm), (a) water-vacuum interface, (b) initialized from a bilayer structure, simulated at 2wt% NaCl. Figures Ia and IIa depict the water-vacuum interface, viewed from inside the water layer, while Figure Ib shows a frozen bilayer structure, viewed from above. Figure IIb shows two wormlike micelles.

## 1.2 Sodium-headgroup g(r)'s

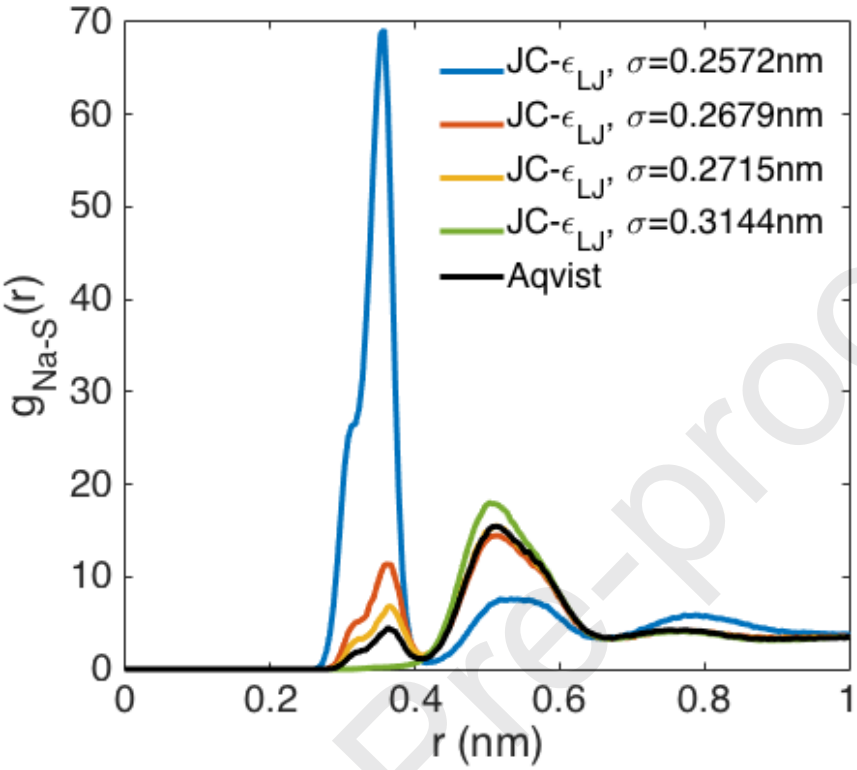



Figure S2: Pair distribution functions g(r) obtained with different Lennard-Jones radii used between sodium ions and the sulfur atom on the sulfate head group. The final model that reproduces the experimental surface tension is $\sigma = 0.2679$nm. The black line uses the Aqvist ion force field,[S1] a popular ion force field that fortuitiously produces reasonable radial distribution functions, but it is also known to prematurely crystallize around 2M.[S2]

## 2 Coarse-Grained Parameters

Table S1: Fixed Parameters

| Parameter | Value |
|---|---|
| $a_W$ | 0.3132 nm |
| $a_{C_2}$ | 0.3132 nm |
| $a_{SO_4^-}$ | 0.3132 nm |
| $a_{Na^+}$ | 0.3132 nm |
| $a_{Cl^-}$ | 0.3132 nm |
| $z_{SO_4^-}$ | -1e |
| $z_{Na^+}$ | 1e |
| $z_{Cl^-}$ | -1e |
| $l_b$ | 0.7676 nm |

Table S2: Common parameters determined by stages 1 and 2

| Parameter | Value | unit |
|---|---|---|
| $\gamma_{WW}$ | 0.2862 | $k_BT$ |
| $\gamma_{C_2C_2}$ | 1.6039 | $k_BT$ |
| $\gamma_{C_2W}$ | 0.9028 | $k_BT$ |
| $\gamma_{NaNa}$ | 1.1207 | $k_BT$ |
| $\gamma_{ClCl}$ | 2.2511 | $k_BT$ |
| $\gamma_{NaCl}$ | -0.7135 | $k_BT$ |
| $\gamma_{NaW}$ | -0.1867 | $k_BT$ |
| $\gamma_{ClW}$ | 0.6314 | $k_BT$ |

Table S3: Model-specific parameters

| Parameter | CG-I | CG-II | CG-III | unit |
|---|---|---|---|---|
| $\gamma_{NaC_2}$ | 0.0676 | 0.2114 | 0.2860 | $k_BT$ |
| $\gamma_{ClC_2}$ | 2.4583 | 2.5245 | 2.5137 | $k_BT$ |
| $\gamma_{SO_4^-SO_4^-}$ | 0.6967 | 2.9058 | 2.9214 | $k_BT$ |
| $\gamma_{SO_4^-W}$ | -0.2515 | 0.0180 | -0.0807 | $k_BT$ |
| $\gamma_{SO_4^-Na^+}$ | -0.8933 | -1.4236 | -0.0300 | $k_BT$ |
| $\gamma_{SO_4^-Cl^-}$ | 0.2970 | 0.3817 | 0.3689 | $k_BT$ |
| $\gamma_{SO_4^-C_2}$ | 0.0295 | 0.9269 | 1.0368 | $k_BT$ |
| $b_{C_2C_2}$ | 0.4106 | 0.4106 | 0.3182 | nm |
| $b_{C_2SO_4^-}$ | 0.2112 | 0.1844 | 0.1868 | nm |

# 3 Molecular Volumes

Table S4: Estimated molecular volumes.

| Molecule | volume ($nm^3$) |
|---|---|
| water | 0.0307 |
| dodecane | 0.4148 |
| NaCl pair | 0.0316 |
| Na-DS | 0.4042 |

# 4 Augmented Lagrangian Relative Entropy Minimization

For each target property $X_i$, an additional augmented Lagrangian penalty term is added to the relative entropy:[S3,S4]

$$S_{rel}(\boldsymbol{\gamma}) - \sum_i \lambda_{i,n}(\langle X_i\rangle_{CG}(\boldsymbol{\gamma}) - \langle X_i\rangle_{AA}) + \frac{c_{i,n}}{2}(\langle X_i\rangle_{CG}(\boldsymbol{\gamma}) - \langle X_i\rangle_{AA})^2 \quad (1)$$

where $\boldsymbol{\gamma}$ are the forcefield parameters (both bonded and non-bonded), $i$ indexes the property of interest, and $n$ indexes the stage. The coefficient $c_{i,n}$ is the strength of the quadratic penalty and is gradually increased in each stage until the constraint is satisfied. The inclusion of the Lagrange term with coefficient $\lambda_{i,n}$ enables the constraint to be satisfied without requiring the quadratic coefficient $c_{i,n}$ to go to infinity. In the first stage, $\lambda_{i,0}$ is set to 0. After each stage, the Lagrange coefficient is updated $\lambda_{i,n+1} = \lambda_{i,n} - c_{i,n}(\langle X_i\rangle_{CG} - \langle X_i\rangle_{AA})$.

Since the bonding model in this work does not feature angle potentials and only has one parameter, the bond length, it can not simultaneously reproduce root mean square bond lengths (which would be achieved if minimizing relative entropy alone) *and* end-to-end distances. In the spirit of coarse-grained polymer models where the statistical segment length is chosen to reproduce larger-scale properties like the chain end-to-end distance instead of monomer-monomer bond lengths,[S5] we similarly opt to reproduce the end-to-end distance as the physical observable relevant for self-assembly. To this end, we treat the bond strength as the only variable that directly responds to the end-to-end distance constraints.

Coarse graining then follows an iterative scheme:

1. Holding all parameters *except* for the $C_2$-$C_2$ bond length $b_{C_2C_2}$ parameter constant, we optimize the bond length with only the augmented Lagrangian term.

2. Holding $b_{C_2C_2}$ constant, we relax the relevant coarse-grained parameters for a given stage of coarse graining via relative entropy minimization, with augmented Lagrangians applied for either the box volume or surface area constraints, where applicable.

In practice, we find that iterating the above process twice is sufficient to reproduce bond lengths to within 0.01nm (i.e. after 2 iterations further relaxing the nonbonded interactions via the augmented relative entropy optimization does not significantly affect the observed end-to-end distances).

## 5 Wormlike Micelles

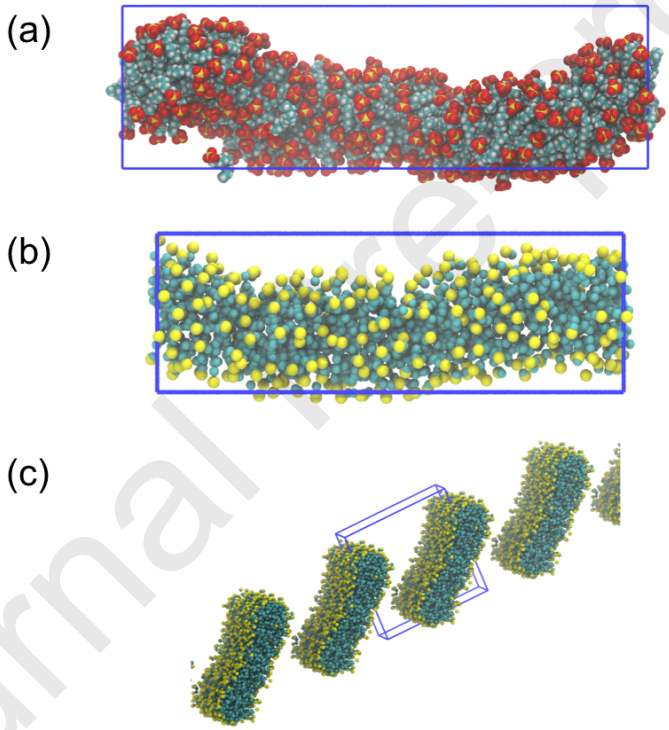



Figure S3: Representative wormlike micelle structures at zero salt for (a) all-atom, (b) using model CG-III, and (c) using model CG-I, which collapses into a bilayer, sheetlike structure.

## 6 Morphological free energies of model CG-I

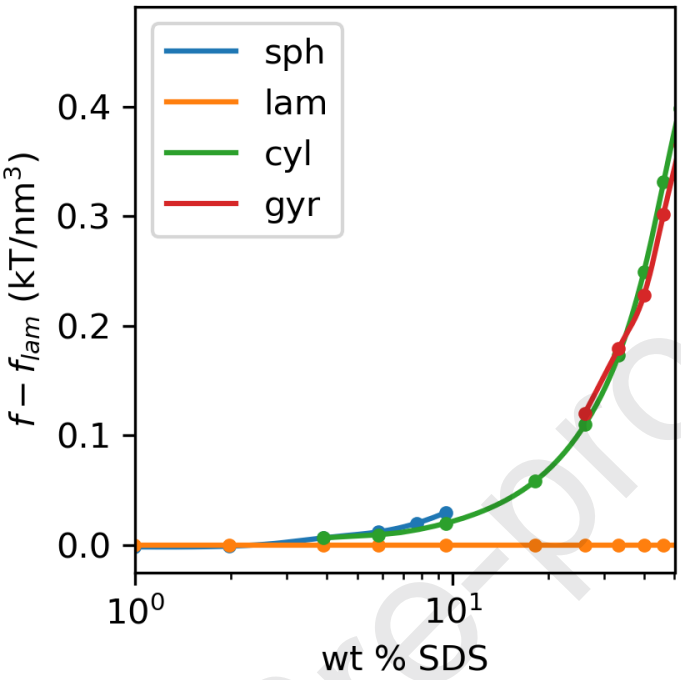



Figure S4: Predicted free energy densities of model CG-I for each considered morphology. Free energies are plotted relative to the value for the lamellar phase in order to emphasize that bilayer structures are stable down to around 2wt% SDS. This is consistent with CGMD simulations that were unable to stabilize wormlike micelles and instead found bilayer structures. At all concentrations considered, the disordered phase has higher free energy density than the self-assembled structures.

## 7 Characteristic spacing of morphologies in model CG-III

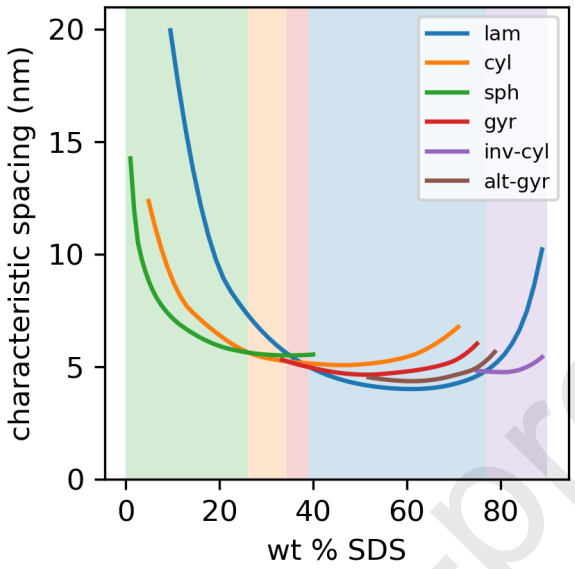



Figure S5: SCFT-predicted characteristic spacings of each morphology as a function of wt% SDS for model CG-III. The shadings highlight the phase with minimum characteristic spacing. See text for definition of the characteristic spacing for each morphology.

## References

(S1) Aqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry* **1990**, *94*, 8021–8024.

(S2) Joung, I. S.; Cheatham III, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The Journal of Physical Chemistry B* **2008**, *112*, 9020–9041.

(S3) Gould, N. I. M. On the convergence of a sequential penalty function method for constrained minimization. *SIAM Journal on Numerical Analysis* **1989**, *26*, 107–128.

(S4) Bertsekas, D. P. *Constrained optimization and Lagrange multiplier methods*; Academic press, 2014.

(S5) Fredrickson, G., et al. *The equilibrium theory of inhomogeneous polymers*; Oxford University Press on Demand, 2006.

Credit Author Statement



Kevin Shen: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, project administration, writing - original draft, writing - review and editing, visualization.

My Nguyen: Conceptualization, methodology, software, validation, writing - review and editing

Nick Sherck: Conceptualization, methodology, software, validation, writing - review and editing, funding acquisition

Brian Yoo, Stephan Kohler, Joshua Speros, Kris T. Delaney M. Scott Shell, Glenn H. Fredrickson: conceptualization, resources, writing - review and editing, supervision, project administration, funding acquisition

**Declaration of interests**

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.