

RESEARCH ARTICLE | DECEMBER 27 2023

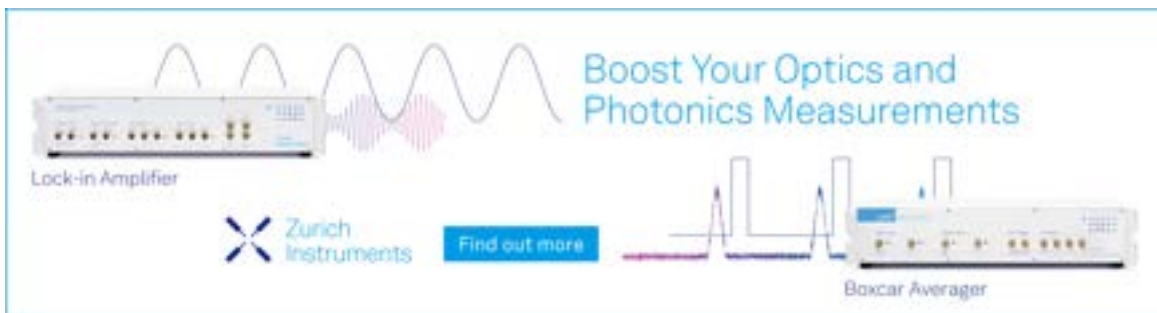
Molecularly informed field theory for estimating critical micelle concentrations of intrinsically disordered protein surfactants

My. V. T. Nguyen ✉ ; Kate Dolph ; Kris T. Delaney ; Kevin Shen ; Nicholas Sherck ; Stephan Köhler ; Rohini Gupta ; Matthew B. Francis ; M. Scott Shell ; Glenn H. Fredrickson 



J. Chem. Phys. 159, 244904 (2023)

<https://doi.org/10.1063/5.0178910>

Boost Your Optics and Photonics Measurements

Lock-in Amplifier

 Zurich Instruments

[Find out more](#)

Boxcar Averager

Molecularly informed field theory for estimating critical micelle concentrations of intrinsically disordered protein surfactants

Cite as: J. Chem. Phys. 159, 244904 (2023); doi: 10.1063/5.0178910

Submitted: 28 September 2023 • Accepted: 30 November 2023 •

Published Online: 27 December 2023



My. V. T. Nguyen,^{1,a)} Kate Dolph,² Kris T. Delaney,³ Kevin Shen,^{1,3} Nicholas Sherck,⁴ Stephan Köhler,⁵ Rohini Gupta,⁶ Matthew B. Francis,^{2,7} M. Scott Shell,^{1,b)} and Glenn H. Fredrickson^{1,3,8,c)}

AFFILIATIONS

¹ Department of Chemical Engineering, University of California, Santa Barbara, California 93106, USA

² Department of Chemistry, University of California, Berkeley, California 94720, USA

³ Materials Research Laboratory, University of California, Santa Barbara, California 93106, USA

⁴ BASF Corporation, Iselin, New Jersey 08830, USA

⁵ BASF SE, Ludwigshafen am Rhein 67056, Germany

⁶ California Research Alliance (CARA) by BASF, Berkeley, California 94720, USA

⁷ Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

⁸ Department of Materials, University of California, Santa Barbara, California 93106, USA

^{a)} Author to whom correspondence should be addressed: my@ucsb.edu

^{b)} E-mail: shell@ucsb.edu

^{c)} E-mail: ghf@ucsb.edu

ABSTRACT

The critical micelle concentration (CMC) is a crucial parameter in understanding the self-assembly behavior of surfactants. In this study, we combine simulation and experiment to demonstrate the predictive capability of molecularly informed field theories in estimating the CMC of biologically based protein surfactants. Our simulation approach combines the relative entropy coarse-graining of small-scale atomistic simulations with large-scale field-theoretic simulations, allowing us to efficiently compute the free energy of micelle formation necessary for the CMC calculation while preserving chemistry-specific information about the underlying surfactant building blocks. We apply this methodology to a unique intrinsically disordered protein platform capable of a wide variety of tailored sequences that enable tunable micelle self-assembly. The computational predictions of the CMC closely match experimental measurements, demonstrating the potential of molecularly informed field theories as a valuable tool to investigate self-assembly in bio-based macromolecules systematically.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0178910>

I. INTRODUCTION

The self-assembly of amphiphilic molecules, such as block copolymers, surfactants, and biomolecules, plays a critical role in many natural and industrial processes. Examples include the formation of biological cell membranes through the assembly of lipid molecules, the micellization of surfactants in detergents, and drug encapsulation.^{1–4} In these systems, the critical micelle concentration (CMC), which is the concentration of surfactants at which

micelles begin to form, is a key design parameter that quantifies the propensity for self-assembly. The CMC is also important for understanding the solution phase behavior of surfactants, offering insights into their interfacial activity, solubility, and emulsification properties.^{5–7} This has been highlighted by recent simulation efforts to predict the CMC, providing a systematic route for evaluating self-assembly in a vast design space (chemistry, molecule architecture, molecular weight, pH, temperature, etc.).^{8–15} Moreover, with the rising emphasis on sustainable and environmentally friendly prac-

tices in chemistry, there is a growing demand for the development of high-throughput screening methods as chemical feedstocks shift from petroleum- to bio-based sources. A predictive computational approach that is suitable to explore these new chemistries can offer an efficient means to screen and assess potential surfactant candidates, facilitating the exploration of greener and more sustainable alternatives to conventional commercial surfactants.¹⁶

Computational tools such as coarse-grained molecular dynamics⁸ and dissipative particle dynamics^{9,10} are commonly used to calculate the CMC in particle-based simulations. Most studies are performed in the NPT or NVT ensemble and track the concentration of free surfactants to estimate the CMC. To mitigate the need for large simulation boxes required near the CMC (typically on the order of mM for commonly studied surfactants),⁹ most particle-based simulation work is conducted at concentrations much higher than the CMC and relies on the assumption that the free surfactant concentration remains constant above the CMC. This assumption has been proven to be inaccurate, especially for ionic surfactant systems.^{11,17–19} Studies have shown that employing empirical corrections accounting for crowding effects due to aggregate formation can provide more accurate predictions of the CMC.^{11,12,20} These corrections, however, are system-dependent, not known *a priori*, and thus require careful investigation. Consequently, particle-based simulations remain limited in their ability to accurately calculate the CMC, especially for strongly micellizing systems with CMC values in the μM range. This is due to both the inaccuracy of extrapolating the free surfactant concentration from the higher concentration regime and the high computational cost of the large simulation boxes required near the CMC. In addition to the length-scale challenge, self-assembly involves inherently long-time-scale processes related to diffusion and micelle fission and fusion, which occur on the order of microseconds.^{21–24} While atomistic simulations are intractable for capturing the time-scales of such phenomena, even particle-based simulations of coarse-grained models face challenges¹¹ in sufficiently sampling the free surfactant concentration and equilibrium distribution of aggregate sizes required for accurate estimation of the CMC.

In principle, a better approach is to calculate the free energy of micelle formation in the grand canonical ensemble. This approach directly determines the stability of the micellar state by comparing its grand free energy with the homogeneous (non-aggregated) state at the same chemical potential. One advantage of the grand canonical ensemble is the reduced simulation box size required for studying micelle formation. Unlike methods that necessitate large simulation boxes to accommodate multiple micelles, the grand canonical ensemble allows for simulations in smaller boxes containing a single micelle. A second advantage is the ability of the grand canonical ensemble to handle fluctuations in the number of particles at a constant chemical potential, which is beneficial when studying micelle formation because prior knowledge of the aggregation number is not required. In contrast to traditional approaches that use the free surfactant concentration as a proxy for the CMC, the grand canonical ensemble directly provides the composition at which micelle formation begins—precisely the definition of the CMC.

While the grand canonical approach is in principle exact and direct, it requires the matching of chemical potentials between the two states (aggregated and homogeneous) by allowing the particle

number to fluctuate. This step is computationally expensive or even intractable in particle-based simulations, particularly for systems that involve macromolecules, due to the need to evaluate chemical potentials, which requires molecular insertion and relaxation. In contrast, field theory has been successfully employed to calculate CMCs for block copolymer and homopolymer mixtures in the grand canonical ensemble,¹³ as chemical potentials and free energies can be directly evaluated through analytical approximations, including mean-field and Gaussian approximations^{25–27} or numerically computed without approximation through field-theoretic simulations via complex Langevin sampling.^{28,29} Furthermore, one can determine equilibrium sizes and aggregation numbers of micelles in the field theory by minimizing the free energy at constant concentration with respect to the simulation cell size, a task that is known to be challenging in particle-based approaches.

In this paper, we utilize a simulation framework enabling chemistry-specific estimation of the CMC, and we validate the accuracy of our method through experimental comparisons. Our demonstration focuses on a model system based on a bio-based surfactant class inspired by intrinsically disordered protein (IDP) sequences found in human neurons and previously studied by Klass and co-workers.^{30,31} This class of bio-inspired, protein-based surfactants possesses a remarkable degree of tunability, stemming from the diverse selection of the 20 naturally occurring amino acids. These amino acids offer a wide range of characteristics, including hydrophobicity, charge, polarity, and aromaticity. Such a rich chemical diversity enables precise engineering of the surfactant's properties, making them versatile and adaptable for various applications. In addition, IDP surfactants offer more precise control over chain length and the individual building block sequence than their synthetic counterparts. Importantly, prior studies demonstrated that these surfactants possess encapsulating properties similar to commonly used synthetic equivalents with the CMC $\sim 10\ \mu\text{M}$.^{30,31} This suggests that IDP surfactants are promising candidates as sustainable replacements for petroleum-based components in many industrial applications including care formulations, coatings, and drug delivery vehicles.

In the field theory literature, studies of bio-based (macro)molecules are relatively limited. In many regards, this arises from the challenges associated with obtaining accurate chemistry-specific interaction parameters that adequately capture the diverse amino acid compositions inherent to bio-based macromolecules. Previous studies often circumvented this by reducing the complex interactions to hydrophilic and hydrophobic interactions in simplified heteropolymer systems.^{32,33} Here, we employ a simulation strategy utilizing a recent development in *molecularly informed* field theory described in our previous publications.^{25,34–36} The method uses relative entropy coarse-graining³⁷ to derive chemistry-specific coarse-grained (CG) interaction parameters from small-scale, reference all-atom (AA) simulations. Subsequently, the coarse particle-based model is exactly transformed into a field-theoretic representation.^{38,39} Because free energies and chemical potentials are readily calculated by operators in the field-theoretic representation, this approach allows for the direct determination of the grand canonical free energy and chemical potential needed for CMC calculation while preserving important information about the underlying chemical components.

The paper is organized as follows: In Sec. II, we describe the computational details and methods of the entire workflow, including the CMC calculation in the field theory. Sec. III provides details of the experimental procedures. The CMC predictions and analysis of various surfactant models are presented in Sec. IV. Finally, we summarize our findings and discuss future directions in Sec. V.

II. COMPUTATIONAL DETAILS AND METHODS

A. All-atom simulations

To parameterize the CG model, we employ two sets of reference AA simulations: pure water (3305 water molecules) and aqueous solutions of the IDP surfactant fragments. The IDP surfactant is comprised of a hydrophilic head and hydrophobic tail, as shown in Fig. 1(a). The hydrophilic head sequence is inspired by the neurofilament heavy arm side-chain protein found in human neurons, which is comprised of n_h repeats of the amino acid

sequence (SPAEAKSPVEVK). The self-assembly of this surfactant is driven by the hydrophobic tail appended to the head domain at its C-terminus. To circumvent the long equilibration time of large IDPs in the AA simulations, we use a short sequence of the hydrophilic domain with $n_h = 2$ and split the surfactant molecule into the head and tail segments. Subsequently, at the connection point of the two domains in the full sequence, we attach neutral C-terminal amide (NME) and N-terminal acetyl (ACE) capping groups to the head and tail, respectively, according to Fig. 1(a). The purpose of these capping groups is to mimic the interaction that would occur between the amino acid at the connection and its neighboring amino acids in the full sequence. We note that these capping groups are not considered in the AA-to-CG mapping process. We consider two choices of reference systems for the IDP surfactant as shown in Fig. 2: a simulation of the head and tail fragments, and an extended ensemble^{40,41} of three simulations, each containing two fragments from the full sequence (head-head, tail-tail, and head-tail). In each of these simulations, we solvate the two surfactant fragments with

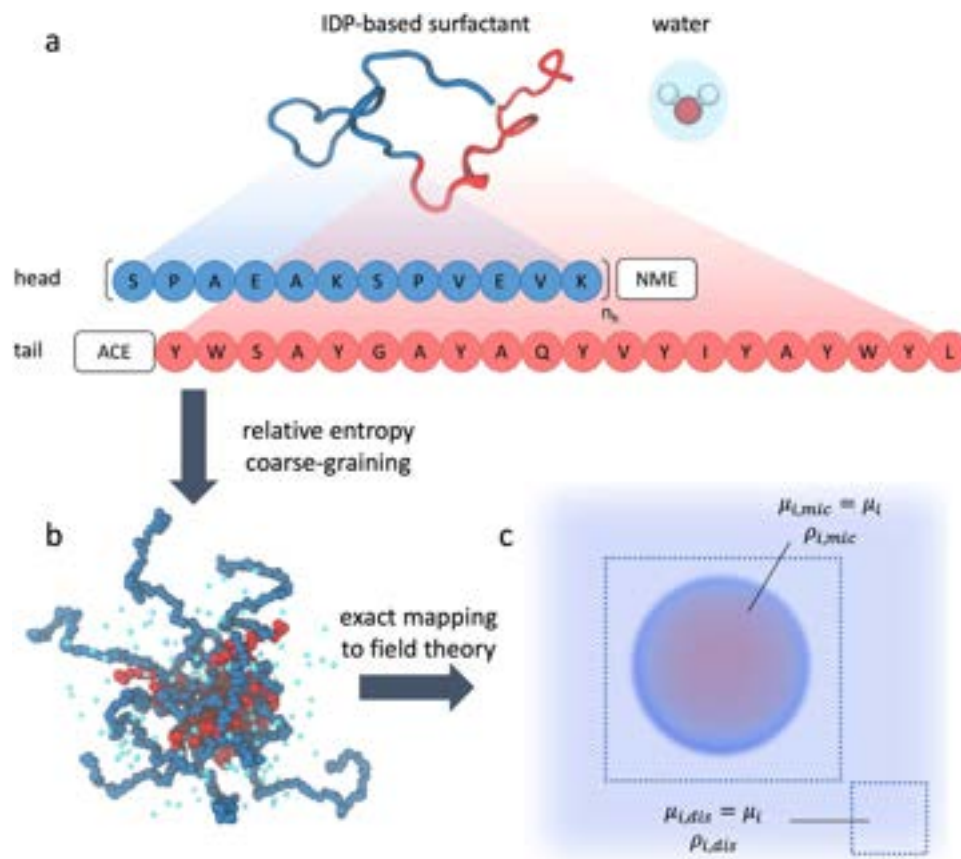


FIG. 1. Schematic of the multi-scale simulation workflow to construct a molecularly informed field-theoretic model of IDP surfactants. (a) Species involved in the all-atom system, which include the IDP surfactant and water. Instead of simulating the full surfactant sequence, we split the surfactant into the head (blue), composed of n_h repeats of the sequence (SPAEAKSPVEVK), and the tail (red) domains. At the connection point of the two domains in the full sequence, we attach neutral C-terminal amide (NME) and N-terminal acetyl (ACE) capping groups to the head and tail, respectively. (b) A coarse-grained particle-based model parameterized by relative entropy minimization. (c) An exact mapping from the coarse-grained particle-based description of the micelle to a field-theoretic model. This schematic also illustrates the CMC calculation approach, which involves matching the chemical potentials in the micellar, $\mu_{l,mic}$, and disordered, $\mu_{l,dis}$, states of compositions $\rho_{l,mic}$ and $\rho_{l,dis}$, respectively.

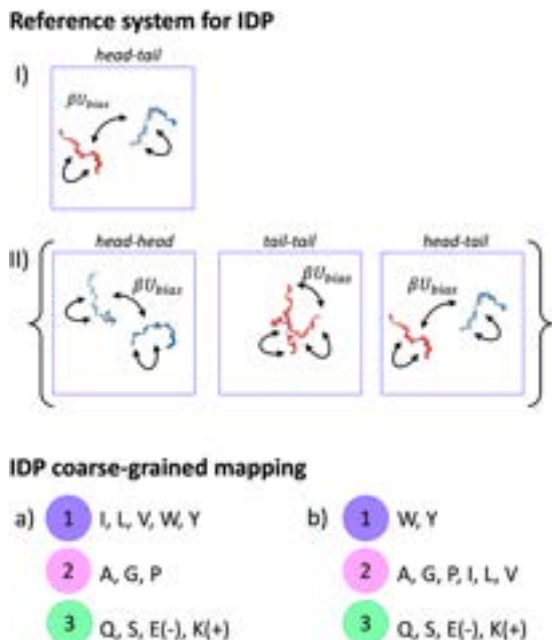


FIG. 2. Reference AA systems and CG mapping schemes considered in constructing the IDP surfactant model. Black arrows depict the biasing potential between the centers of mass of any two amino acids (see the main text for a detailed discussion).

25 500 water molecules and do not include explicit counterions since the hydrophilic head, while carrying charges, is overall neutral.

Although we only simulate two short fragments of the protein, it can still be difficult to obtain accurate distributions at room temperature with explicit solvent by conventional simulation methods because it is easy for the system to become trapped in local minimum-energy states at low temperatures. To circumvent this, we add a Gaussian repulsion between the centers of mass of any two amino acids, including the bonded pairs, in all simulations involving IDP fragments. The added repulsion has the following form:

$$\beta U_{bias}(r) = v_{bias} e^{-r^2/4a^2}, \quad (1)$$

where v_{bias} is the strength of the repulsion and $a = 0.5$ nm defines the interaction range. The functional form of this repulsion is identical to the excluded volume interaction of the CG model, which will be discussed in Sec. II B. This allows us to simply subtract the bias term later to recover the unbiased interaction. Conceptually, the added repulsion makes the IDP fragments less hydrophobic, reducing local minima that correspond to collapsed configurations in the unbiased system and smoothing out the energy landscape. Since we only consider short IDP fragments, this method serves as a practical alternative to more computationally intensive advanced sampling techniques like replica exchange molecular dynamics. In this work, we select $v_{bias} = 0.25 k_B T$, a value sufficient to reduce the probability of collapsed configurations, as evidenced by the reduction in the intensity of lower peaks in the radius-of-gyration and end-to-end distributions shown in Fig. S1 for the tail–tail simulation.

We employ the a99SB-disp force field, which was developed by Robustelli *et al.* to accurately describe both folded and disordered proteins in tandem with the modified TIP4P-D water model.⁴² We conduct reference AA simulations with the OpenMM simulation package.⁴³ A 1 nm cutoff is employed for the direct-space non-bonded interactions, and we use the Particle Mesh Ewald method to compute long-range Coulomb and Lennard-Jones interactions (LJPM method in OpenMM). In addition, we constrain the length of all bonds that involve a hydrogen atom and employ a time step of $dt = 0.002$ ps. The temperature is set to 298.15 K using the Langevin thermostat with a friction coefficient of 5 ps^{-1} , while the pressure is set to 1 atm using a Monte Carlo barostat with an update frequency of $1/(25 \text{ dt})$. We generate the initial configurations for the simulations with the Packmol package.⁴⁴

B. Bottom-up coarse-graining procedure

Although we provide details about the choice of CG potential in previous publications,^{25,34–36} we briefly discuss them here. After performing AA simulations as described in Sec. II A, we use relative entropy coarse-graining³⁷ with these as references to parameterize CG interaction potentials that are amenable to direct conversion to a field theory. In the CG model, bonded interactions in the surfactant molecule are described using a harmonic bond potential,

$$\beta U_{b,\alpha\gamma}(r) = \frac{3}{2b_{\alpha\gamma}^2} r^2, \quad (2)$$

where $\beta = 1/k_B T$ and $b_{\alpha\gamma}$ is interpreted as the root-mean-square length of a bond between bead species α and γ . For simplicity, we assume the same bond length for all amino acid pairs, $b_{\alpha\gamma} \equiv b$. The excluded volume and Coulomb interactions between all site pairs, including bonded pairs, are described by non-bonded pairwise terms involving repulsive Gaussian and regularized Coulomb potentials, respectively,

$$\beta U_{ev,\alpha\gamma} = v_{\alpha\gamma} e^{-r^2/2(a_\alpha^2 + a_\gamma^2)}, \quad (3)$$

$$\beta U_{el,\alpha\gamma} = \frac{l_B \sigma_\alpha \sigma_\gamma}{r} \text{erf}\left(\frac{r}{2\sqrt{a_\alpha^2/2 + a_\gamma^2/2}}\right), \quad (4)$$

where $v_{\alpha\gamma}$ is the excluded volume strength between bead species α and γ , and a_α and σ_α are the Gaussian regularization length and charge of bead species α , respectively. Here, we set the Bjerrum length, l_B , to 0.74 nm, which is approximately the Bjerrum length of OPC water at 298 K and 1 atm.⁴⁵ The regularized Coulomb potential of Eq. (4) reduces to the conventional unscreened Coulomb potential at large separations r but is finite at contact due to the error function. Such regularization is necessary because of the soft-core repulsions adopted in Eq. (3). This choice of regularized, soft potentials is physically motivated by the desire to retain long-length-scale physics while coarse-graining over sharp, short-length-scale features.

We obtain mapped AA reference trajectories for coarse-graining by mapping the center-of-mass coordinates of groups of atoms in the AA representation to CG sites. Specifically, we map each water molecule to a single neutral bead, and each amino acid

is mapped to one neutral bead, with the exception of glutamic acid (E) and lysine (K), which bear a -1 and $+1$ charge, respectively. To reduce the parameter space of the CG model, we categorize the amino acids into three CG bead types based on their hydrophobicity and using two different trial mapping schemes, as shown in Fig. 2. Both schemes share the same definition of bead species 3, which includes polar [serine (S) and glutamine (Q)] and charged amino acids [glutamic acid (E) and lysine (K)]. We note that charges of glutamic acid and lysine are described explicitly via the electrostatic interaction of Eq. (4). In scheme **a**, the small neutral amino acids are grouped into bead species 2, while the larger amino acids are lumped into bead species 1. On the other hand, in scheme **b**, CG bead 1 only includes tyrosine (Y) and tryptophan (W), which have bulky aromatic side chains, while the rest of the hydrophobic and neutral amino acids are mapped to CG bead 2. In principle, one can further subdivide the amino acids into more CG bead types to achieve greater chemical specificity. For instance, each individual amino acid could be mapped to its own dedicated CG bead. In this scenario, the method would follow a similar process as outlined here.

We fix the Gaussian regularization range, a_α , of each CG bead to approximately the cube root of its molecular volume. By this convention, the water interaction range, a_w , is set to 0.31 nm, and those of the amino acids, a_1 , a_2 , and a_3 , are set to 0.5 nm in this work. The water–water repulsion parameter is obtained from a pure water AA simulation following the same procedure as in our previous publication.³⁶ In this step, we derive v_{ww} in the NPT ensemble at the CG pressure P_{CG} of $3.218 k_B T/a_w^3$. This determines the CG pressure that we use in the subsequent coarse-graining steps of the IDP surfactant.

We derive the remaining CG parameters ($v_{11}, v_{22}, v_{33}, v_{12}, v_{13}, v_{1w}, v_{23}, v_{2w}, v_{3w}$, and b) for the surfactant from two choices of reference systems: a single simulation and an extended ensemble of three simulations, as illustrated in Fig. 2. With the two choices of reference systems (I and II) and two mapping schemes for the surfactant (**a** and **b**), we have four candidate surfactant models in this work: Ia, Ib, IIa, and IIb. In each of these cases, the coarse graining is performed by running S_{rel} minimization multiple times to obtain replicates of the CG force field. This allows us to perform error analysis and sensitivity assessments of the CMC that will be discussed in Sec. IV. We tabulate the parameters in Tables SI–SV.

C. Calculating the critical micelle concentration with field theory

The CG model defined in Sec. II B can be exactly represented and simulated in a field theoretic representation via the Hubbard–Stratonovich–Edwards transformation.^{38,39} This transformation decouples the non-bonded pair interactions such that particles interact only via the bonded potential and with auxiliary fields introduced by the transform. The result is a partition function in terms of integrals over field configurations instead of particle coordinates,

$$\mathcal{Z} = \int d\mathbf{r}^n e^{-\beta U(\mathbf{r}^n)} \rightarrow \int \mathcal{D}\mathbf{w} e^{-H[\mathbf{w}]}, \quad (5)$$

where H is an effective Hamiltonian describing the statistical weight of the auxiliary field configuration $\mathbf{w}(\mathbf{r})$ and is systematically described elsewhere.^{38,39} It should be emphasized that $\mathbf{w}$ represents

a set of auxiliary fields that is sufficient to decouple all pairwise interactions of the functional forms defined in Sec. II B.

In the mean-field approximation, also referred to as self-consistent field theory (SCFT), the canonical partition function takes the form

$$\mathcal{Z} \approx e^{-H[\mathbf{w}^*]} \equiv e^{-H^*}, \quad (6)$$

where $\mathbf{w}^*$ is the saddle-point value of each auxiliary field, representing the dominant field configuration contributing to the partition function, and H^* is the mean-field effective Hamiltonian. SCFT reduces computational costs compared to sampling $\mathbf{w}$ field configurations. The grand free energy Ω can be determined from the Legendre transform of the Helmholtz free energy $A \approx H^*$ as follows:

$$\beta\Omega = \beta A - \sum_{i \in [w, idp]} \beta\mu_i n_i, \quad (7)$$

where n_i is the number of molecule i . Therefore, SCFT provides a readily accessible approximation of the grand free energy, which is necessary for CMC calculations.

To determine the CMC, we compute the grand free energy difference between a micelle and the homogeneous phase in chemical equilibrium with that micelle, following the procedure outlined by Zhou and Shi.¹³ We initiate SCFT simulations of a micelle in the *canonical ensemble* with a spherical configuration for various values of the IDP mole fraction, $\phi_{idp, mic}$. We opt for spherical geometry as an initial seed for the micelle in accordance with experimental observations.^{30,31} It is worth noting that there are no constraints imposed on the micelle geometry, and the final spherical configuration of the micelles [Fig. 1(c)] naturally emerges as a result of solving the SCFT equations. This results in a series of spherical micelles at different IDP and water chemical potentials, $\mu_{w, mic}$ and $\mu_{idp, mic}$, respectively. The homogeneous disordered state that is in chemical equilibrium with each of these micelle states is modeled by setting the number of mesh points to 1, and its composition is determined from a *grand canonical* simulation at $\mu_{i, dis} = \mu_{i, mic} \equiv \mu_i$. This process is illustrated in Fig. 1(c), which depicts a micelle in chemical equilibrium with the homogeneous phase.

The grand free energy difference $\beta\Delta\Omega = \beta\Omega_{mic} - \beta\Omega_{dis}$ is obtained for a series of $\phi_{idp, mic}$ values. Figures 3(a) and 3(b) show examples of free-energy-difference curves as a function of the IDP surfactant concentration in the disordered homogeneous solution and the surfactant chemical potential, respectively. For large surfactant densities $\rho_{idp, dis}$, the negative free energy difference $\beta\Delta\Omega$ indicates the micelles are more stable than the homogeneous state. As $\rho_{idp, dis}$ decreases, $\beta\Delta\Omega$ increases and eventually becomes positive. The CMC, $\rho_{idp, CMC}$, is defined as the surfactant concentration in the homogeneous state at which the free energy of micelle formation is 0.

To investigate finite-size effects, we conduct micelle simulations using different box sizes, ranging from ~ 10 to 22 nm in side length. We then extrapolate the CMC values against the inverse box side length, $1/L$ [Fig. 3(c)], and extract the CMC at $1/L \rightarrow 0$ or as the micelle simulation box size approaches infinity ($L \rightarrow \infty$). We repeat these steps for each surfactant model and report the mean and median values of the CMC.

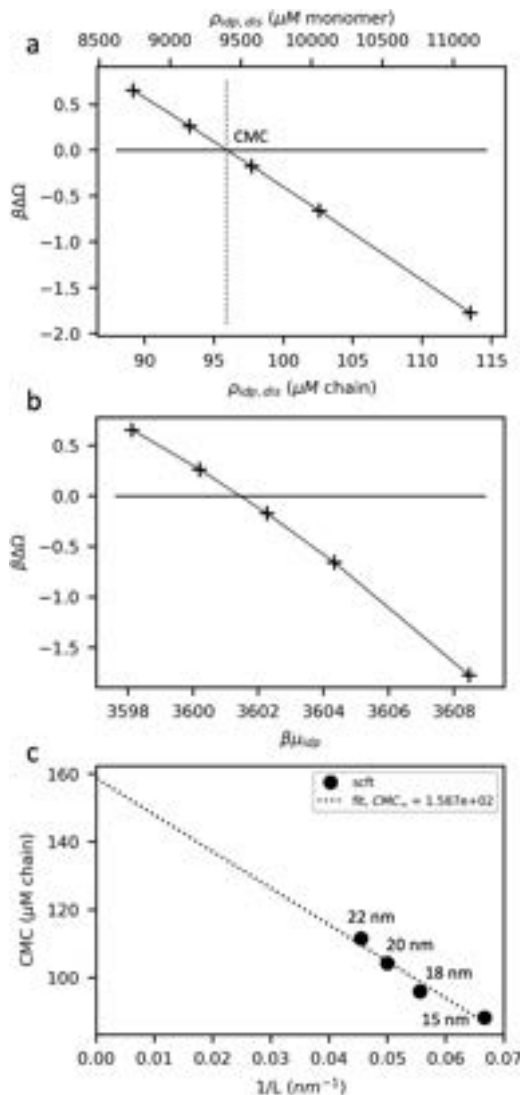


FIG. 3. Example grand free energy difference $\beta\Delta\Omega$ between a spherical micelle and the homogeneous phase as a function of (a) surfactant concentration in the homogeneous phase $\rho_{idp,dis}$ (chain and monomer basis) and (b) surfactant chemical potential $\beta\mu_{idp}$. (c) To account for finite-size errors, we extrapolate the CMC linearly with respect to the inverse of the box size length.

III. DETERMINATION OF THE CRITICAL MICELLE CONCENTRATION VIA PYRENE-BASED FLUORESCENCE ASSAY

We determine the CMC of the surfactant with a hydrophilic block length of $n_h = 6.5$ experimentally by a solvatochromic pyrene-based fluorescence assay, as described previously.^{30,31} In short, fluorescence emission intensities from the first (I_I) and third (I_{III}) vibronic bands of pyrene are dependent on the polarity of its local environment. For a $2 \mu\text{M}$ pyrene solution in 10 mM phosphate buffer, the I_I/I_{III} ratio is ~ 1.3 and lowers to ~ 0.8 when pyrene is

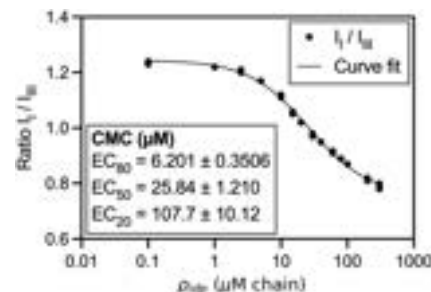


FIG. 4. Pyrene I_I/I_{III} fluorescence emission ratio across concentrations of the IDP surfactant with $n_h = 6.5$. Solutions containing 0.1–300 μM surfactant in 2 μM pyrene and 10 mM phosphate buffer, pH 6.5, were excited at 330 nm, and the emission was recorded at 373 nm (I_I) and 384 nm (I_{III}).

encapsulated in the less polar hydrophobic core. The CMC was determined by plotting triplicate measurements of the I_I/I_{III} ratio across a range of surfactant concentrations. We fit a nonlinear least squares regression to the following equations:

$$I_I/I_{III} = y_0 + \frac{c}{1 + (EC_{50}/\rho_{idp})^n}, \quad (8)$$

$$EC_{50} = EC_F \left(\frac{100 - F}{F} \right)^{1/n}. \quad (9)$$

The fitted parameters include the Hill coefficient, n , the IDP surfactant concentration at the inflection point, EC_{50} , the vertical shift, y_0 , and the scaling coefficient, c . In Fig. 4, we report the CMC and its standard error as the surfactant concentration at $F\%$ of the maximum signal, EC_F , of this nonlinear fit, where F is 80, 50, or 20. Importantly, prior research suggests that for surfactants with CMC values below 1 mM, the inflection point EC_{50} is a more suitable approximation.^{46,47}

IV. RESULTS AND DISCUSSION

A. Coarse-grained model evaluation

To calculate the CMC, it is necessary to have a well-defined homogeneous phase in coexistence with a micelle. In other words, the system should not undergo macrophase separation at compositions near the expected CMC values. To identify the two-phase boundary, we employ the Gibbs ensemble method and invoke the mean-field approximation for the free energy and chemical potential calculations. A detailed discussion of this procedure can be found in our previous publication.²⁵ Figure 5(a) shows the binodals using representative parameter sets for the four different IDP models. As the number of repeating hydrophilic unit n_h increases, the fraction of hydrophilic beads also increases, resulting in a reduction in the tendency for macrophase separation. Consequently, the two-phase region becomes narrower.

At the hydrophilic block length of $n_h = 6.5$, which corresponds to the experimental system, the dilute branch of models IIa and IIb extends down to $\sim 10^{-10} \mu\text{M}$. This value is orders of magnitude smaller than the experimentally determined CMC range (6.201–107.7 μM). This indicates that these models are likely too hydrophobic and will undergo macrophase separation instead of

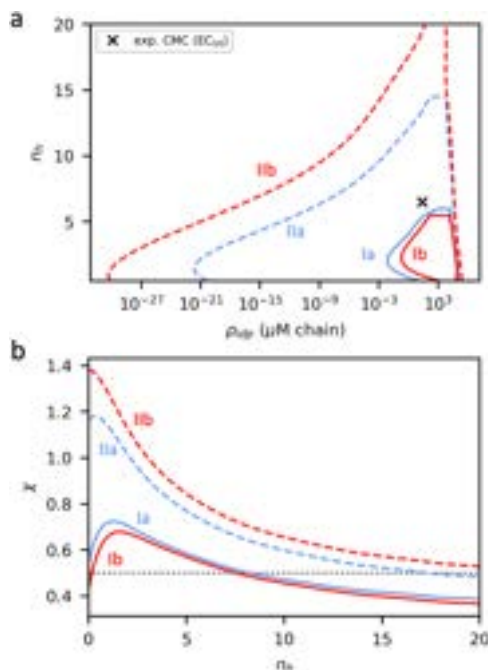


FIG. 5. (a) Binodals calculated from representative parameter sets for four IDP models at varying numbers of hydrophilic repeating units, n_h . The symbol denotes the experimentally determined CMC (EC_{50} value) at $n_h = 6.5$. (b) Corresponding χ parameters against n_h . The dotted line denotes $\chi = 0.5$.

micelle formation at low IDP concentrations. In contrast, models Ia and Ib exhibit smaller two-phase regions that disappear at $n_h \sim 6$. This indicates that the homogenous state remains stable over a larger composition range when using models Ia and Ib, allowing for the formation of micelles before reaching concentrations where macrophase separation occurs.

To better understand the differences between the four CG models, we simplify the parameter space by reducing the ten pairwise interactions to a single effective Flory–Huggins parameter, χ .^{48,49} The Flory–Huggins binary interaction χ approximates the overall affinity between IDP chains in solution and is correlated with the mixture phase behavior; a higher value of χ indicates a greater tendency toward phase separation. In this work, χ is defined as

$$\chi = v_{ref} \left(\rho_w^* \rho_{idp}^* U_{idp\ w} - \frac{1}{2} (\rho_{idp}^{*2} U_{idp\ idp} + \rho_w^{*2} U_{w\ w}) \right), \quad (10)$$

where the reference volume v_{ref} is taken to be the molecular volume of water, $v_{ref} = (\rho_w^*)^{-1}$. The neat chain density of species i , ρ_i^* , is estimated using the mean-field approximation (detailed in the supplementary material of Ref. 25) as follows:

$$\rho_i^* = \frac{-1 + \sqrt{1 + 2U_{ii}P_{CG}}}{U_{ii}}. \quad (11)$$

Equations (10) and (11) involve the excluded volume parameter U_{ij} between molecules i and j , which is defined by summations over bead and molecule species,

$$U_{ij} = \sum_{\alpha, \gamma \in [w, 1, 2, 3]} \sum_{i, j \in [w, idp]} u_{\alpha\gamma} f_{i,\alpha} f_{j,\gamma} N_i N_j, \quad (12)$$

where $u_{\alpha\gamma}$ is the integrated value of the excluded volume interaction $\beta U_{ev,\alpha\gamma}$ between beads α and γ , i.e., $u_{\alpha\gamma} = v_{\alpha\gamma} (2\pi(a_\alpha^2 + a_\gamma^2))^{3/2}$. The number fraction of bead α on chain i is denoted as $f_{i,\alpha}$, and the chain lengths of water and surfactant are $N_w = 1$ and $N_{idp} = 12n_h + 20$, respectively. According to this definition, the excluded volume strength between water molecules is the same in both the bead-basis and molecule-basis definitions, i.e., $U_{ww} = u_{ww}$.

The previous observation of wider two-phase regions in models IIa and IIb is supported by the fact that they consistently have larger χ values than those of models Ia and Ib at all values of n_h , as illustrated in Fig. 5(b). As a first approximation, phase separation typically occurs at $\chi \gtrsim 0.5$,⁵⁰ indicating that models with χ values exceeding the critical threshold are more likely to undergo macrophase separation. In such systems, the CMC is either very small and lies to the left of the dilute branch or does not exist at all. The distinct behavior of models IIa and IIb compared to Ia and Ib can be attributed to the choice of the reference simulation. Reference system II is an extended ensemble that includes the tail–tail simulation. As evidenced by the small center-of-mass distance between the tail fragments (~ 1 nm, Fig. S2), the aggregated state is the dominant conformation, with a little sampling of the dissociated state, which overemphasizes hydrophobic interactions that occur in the tail regions. In comparison to reference system I, which is composed of the head–tail simulation, system II has a higher number of contacts between IDP residues (Table SVI). Consequently, this results in IDP models that are more hydrophobic in models IIa and IIb, which promotes macrophase separation, as reflected in the large χ parameter and wide binodal region. This observation suggests that IDP models (models Ia and Ib) derived from reference system I, a solution comprising a hydrophilic head and a hydrophobic tail fragment, are more suitable for determining the CMCs; we only proceed with these models for further analysis in Sec. IV B.

B. Critical micelle concentration

While the CMC calculation procedure with SCFT is deterministic, the derivation of CG parameters via relative entropy minimization involves stochastic sampling from short CGMD simulations, leading to CG parameter variation. Therefore, we repeated the parameterization process 20 times and performed the necessary calculations for each of these replicates to obtain statistically meaningful values. In Table I, we present Flory–Huggins χ , as well as the mean and median values of the CMC for the hydrophilic block length $n_h = 6.5$ using surfactant models Ia and Ib. The CMC distributions from the two models are right-skewed, with the mean values larger than the medians. It is noteworthy that model Ia yields smaller mean and median CMC values that are in better agreement with the experimental CMC at EC_{50} as compared to model Ib. The CMC value calculated from model Ia shows relatively good agreement with the experimental result, particularly when considering the median values, which are within a factor of 2 of the experimental

TABLE I. SCFT-predicted results for IDP surfactant at $n_h = 6.5$ and experimental data.

Model	n_h	χ	CMC (μM chain)		Diameter (nm)	n_{agg}
			Mean	Median		
Ia	6.5	0.523 ± 0.003	90 ± 25	49	17.37 ± 0.21	7.58 ± 0.58
Ib	6.5	0.503 ± 0.004	260 ± 44	237	17.08 ± 0.42	6.75 ± 0.41
Experiment	6.5	$\dots$	6.20 ± 0.35 (EC ₈₀)		19.6 ± 4.9^a	$\dots$
			25.84 ± 1.21 (EC ₅₀)			
			107.7 ± 10.1 (EC ₂₀)			

^aReported in Ref. 31.

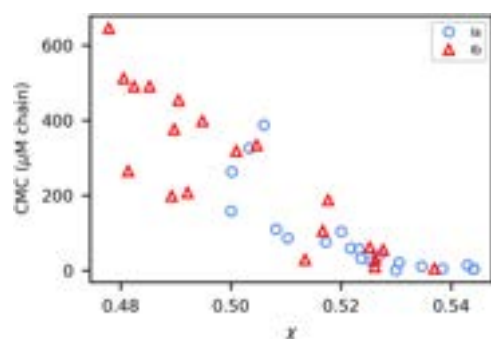
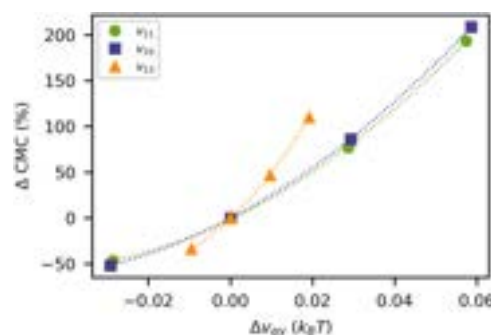
EC₅₀ value. It is important to note that micellization is not a true thermodynamic phase transition such that properties, e.g., free surfactant concentration, osmotic pressure, and volume occupied by micelles, exhibit rapid *continuous* changes through the CMC instead of a sharp transition.⁵¹ Therefore, different methods of inferring the CMC can yield slightly different values.^{12,52–55} Specifically, in this case, it is reasonable to expect micelles to form at any concentration between the EC₈₀ and EC₂₀ values, 6.201–107.7 μM . Taking this into account, both the median and mean values from model Ia fall within the experimental CMC range and are both reasonable proxies for the CMC.

The difference in the CMCs between models Ia and Ib stems from the use of different CG mapping schemes. Grouping the smaller amino acids (I, L, and V) with the larger aromatic amino acids (Y and W) in CG bead 1 results in a smaller effective size of the bead. As a consequence, model Ia's bead 1 exhibits a smaller excluded volume interaction, as evidenced by the ~ 2 times smaller integrated excluded volume interaction u_{11} compared to that of model Ib (Fig. S3). The increase in the excluded volume of bead 1 and other bead types in model Ib leads to a less hydrophobic IDP, thereby resulting in a higher CMC compared to model Ia.

Despite the comparable χ values at $n_h = 6.5$, the difference in CMC values is significant between the two models. The sensitivity of the CMC to χ is evident from Fig. 6, where higher χ values promote micellization, resulting in smaller CMCs. This is reflected in model

Ia, which has a higher average χ value and, therefore, exhibits a lower CMC compared to model Ib. To further evaluate this sensitivity, we analyze the impact of perturbing individual excluded volume parameters, v_{11} , v_{33} , and v_{13} , on the CMC of a replicate in model Ia. Figure 7 illustrates that even small perturbations in the interaction parameters can significantly affect the CMC. Specifically, a mere increase of $\sim 0.02 k_B T$ in the excluded volume strength between hydrophobic bead 1 and hydrophilic bead 3, v_{13} , can cause the CMC to vary by up to 100%. Considering that ~ 6 out of 9 excluded volume parameters involving IDP residues exhibit variations larger than $0.02 k_B T$ across 20 replicates for both models Ia and Ib (Fig. S4), it is expected that the CMC will exhibit substantial variation between different replicates. It is important to acknowledge that these observed uncertainties in the parameters are relatively small compared to the typical error in solvation energies of atomistic force fields, which can be up to 0.5 kcal/mol or $\sim 0.8 k_B T$.⁵⁶

The variation in the excluded volume parameters across the replicates arises from the inherent stochastic nature of finite-length CGMD simulations, which are used to evaluate derivatives for updating parameters during relative entropy minimization.³⁷ In this study, we use a simulation length that provides $\gtrsim 50$ independent samples of the fragment end-to-end distance, which we believe is adequate while maintaining a reasonable computational cost for the relative entropy minimization. However, it is worth noting that increasing the simulation time of the trial CGMD simulations could

**FIG. 6.** CMC at $n_h = 6.5$ calculated from 20 replicates for models Ia and Ib against χ . The higher average χ of model Ia suggests that this model is slightly more hydrophobic than model Ib, resulting in a lower average CMC value.**FIG. 7.** Sensitivity analysis of the CMC with respect to the change in excluded volume parameters v_{11} , v_{33} , and v_{13} . A plot that shows the percentage change in the CMC with respect to the change in the excluded volume parameter from the base value.

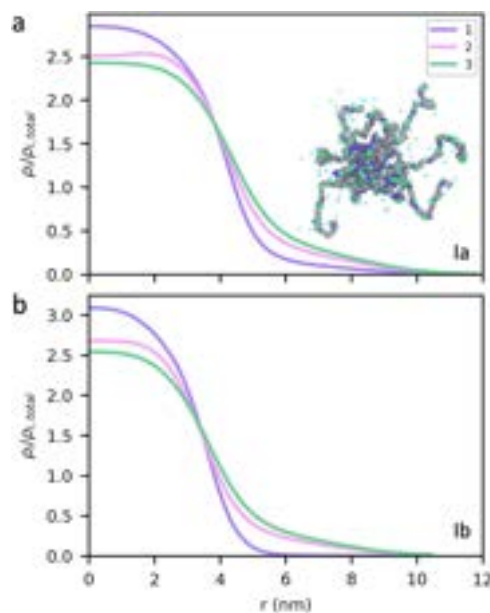


FIG. 8. Density profile of micelles at $n_h = 6.5$ from representative parameter sets for models (a) Ia and (b) Ib at the corresponding CMCs. The inset is a snapshot of the micelle from CGMD reconstructed based on the equilibrium aggregation number calculated in the field theory.

potentially further reduce parameter variations, and this will be carefully considered in future work.

Taking the high sensitivity of the CMC to the CG interaction parameters into account, the proposed CMC calculation workflow using molecularly informed field theories demonstrates good agreement with experimental data, particularly when using model Ia. Additionally, we calculate the equilibrium micelle size measured at the corresponding CMCs from SCFT density profiles (Fig. 8). Micelle diameter is defined as the distance where the local concentration of IDP is 2.5% of the peak value at the micelle center. Using this criterion, we find that both models produce micelles of similar diameter to those determined by dynamic light scattering experiments,³¹ as reported in Table I. SCFT simulations also reveal that the average aggregation number n_{agg} is 7.58 ± 0.58 and 6.75 ± 0.41 for models Ia and Ib, respectively. Remarkably, this quantitative agreement in CMC and micelle diameter is achieved using a simplified CG surfactant model that reduces the complexity of the 12 unique amino acids found in the actual IDP surfactant chemistry to only 3 CG bead types.

The ability to achieve such agreement with a reduced CG model underscores the potential of the approach. By incorporating more chemical detail into the CG model, such as specifying additional CG bead types, we anticipate that even higher accuracy can be obtained. This refinement would enhance the representation of the molecular interactions of the surfactant systems, thereby improving the predictive capability of the CMC calculation. Furthermore, leveraging the efficiency of the field theory in obtaining equilibrium micelle structures, one can readily use the predicted aggregation number to reconstruct micelles in particle-based CGMD

simulations, as shown in the inset of Fig. 8. This can be performed by pre-assembling micelles based on the SCFT-predicted aggregation numbers or by implementing a backmapping strategy proposed by Lequeu.⁵⁷ The flexibility to transform between particle and field-theoretic representations allows for a detailed examination of micelle conformations while overcoming the challenge of long time scales faced by particle-based approaches.

V. CONCLUSIONS AND OUTLOOK

In this study, we have presented a workflow for calculating the critical micelle concentration (CMC) of bio-based surfactants using molecularly informed field-theoretic models. Our approach incorporates chemistry-specific effects, which are often overlooked in field-theoretic studies, by employing relative entropy coarse-graining to systematically determine field theory parameters by coarse-graining from all-atom simulations. We have illustrated the effectiveness of the field-theoretic models in capturing the self-assembly behavior of a model intrinsically disordered protein surfactant. Despite using a simplified coarse-grained surfactant model with only three distinct chemical species to represent the complex chemistry of the surfactants composed of 12 unique amino acids, our simulations yielded a CMC that falls within the experimental CMC range and is within a factor of 2 of the experimental EC_{50} value. Notably, our approach is capable of tackling a chemical space characterized by significantly lower CMC values (in the μM range) compared to previous simulation studies in the literature.^{9,10} This highlights the potential of this approach, particularly in modeling bio-based molecules where complex interactions could arise from a diverse set of amino acids.

We proposed factors that affect the accuracy of the CMC prediction, including the choice of the reference simulations for coarse-graining and the definition of coarse-grained bead types (Fig. 2). We have found that the coarse-grained models derived from an extended ensemble of three simulations, each composed of a pair of the hydrophilic head and/or hydrophobic tail domains in water, tend to overemphasize hydrophobic interactions due to aggregation and limited sampling in the tail regions. Consequently, this overly promotes macrophase separation at much lower surfactant concentrations than the experimentally determined CMC, suggesting that coarse-grained models parameterized by this type of reference system are not representative of the surfactant. In contrast, a reference system comprising a hydrophilic head and a hydrophobic tail fragment is a more suitable reference for coarse-graining. Coarse-grained models derived from this system exhibit less hydrophobicity compared to those obtained from the extended ensemble approach. This alternative reference system mitigates the tendency toward macrophase separation at low surfactant concentrations, which allows micelles to form at concentrations close to the experiments. Lastly, we have compared two schemes for defining the coarse-grained bead types; both schemes have the same number of bead types. We found that scheme a, which groups the aromatic and hydrophobic residues into one coarse-grained (CG) bead type, produces better agreement in the CMC with experimental data as compared to scheme b, in which the aromatic residues are grouped into one CG bead and the hydrophobic residues are grouped together with neutral residues into another CG bead. Overall, this highlights the influence of CG bead definitions on the

accuracy of the predicted CMC. In general, we can readily extend the workflow to include more chemical species in the CG model. Increasing the number of bead types is expected to enhance accuracy by providing greater chemical specificity, but at the cost of a more complex CG force field and potential challenges in sampling. The latter arises from the fact that, with more bead types, the pair interactions are now parsed into statistically smaller groups; therefore, longer simulations are needed to sufficiently sample the different contacts.

We note that the current work employs the mean-field approximation, which ignores fluctuation effects in the field-theoretic model. While this approximation is reasonable for studying surfactant self-assembly based on our previous findings,³⁴ incorporating w -field fluctuations using techniques such as complex Langevin sampling^{39,58} could improve the model's accuracy and change the quantitative prediction of the CMCs. The grand free energy and chemical potential can be directly calculated in such simulations from ensemble average operators.²⁹

In conclusion, our study showcases the capability of molecularly informed field theories to systematically predict the CMC of bio-based molecules. Our simulation framework offers an efficient route for calculating the CMC, particularly for strongly micellizing systems, where traditional particle-based simulations face challenges. This work opens up possibilities for employing molecularly informed field theories in the study and design of bio-based macromolecules, providing valuable insights into their self-assembly properties, and facilitating the optimization of their performance in various applications. Overall, our approach contributes to the design of sustainable formulations and advances our understanding of bio-inspired surfactant systems.

SUPPLEMENTARY MATERIAL

The force field parameters of all coarse-grained models and additional analysis of these models are available in the supplementary material.

ACKNOWLEDGMENTS

This work was supported by BASF Corporation through the California Research Alliance. K.D. is also supported by the Chemical Biology Graduate Program at UC Berkeley (Grant No. NIH T32GM066698). We acknowledge Paul Huang and Kueyoung Kim for their assistance with the surfactant synthesis and pyrene CMC assays. G.H.F. and K.T.D. derived partial support from the National Science Foundation CMMT Program under Grant No. DMR-2104255. M.S.S. acknowledges funding support from the National Science Foundation through Award No. CHEM-1800344. Use was made of the BioPACIFIC Materials Innovation Platform computing resources of the National Science Foundation Award No. DMR-1933487 and computational facilities purchased with funds from the National Science Foundation (Grant No. CNS-1725797) 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; Grant No. NSF DMR 2308708) at UC Santa Barbara.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

My Nguyen: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Methodology (lead); Validation (lead); Visualization (lead); Writing – original draft (lead). **Kate Dolph:** Data curation (supporting); Formal analysis (supporting); Writing – original draft (supporting). **Kris T. Delaney:** Conceptualization (equal); Funding acquisition (equal); Methodology (supporting); Software (supporting); Writing – review & editing (equal). **Kevin Shen:** Methodology (supporting); Writing – review & editing (equal). **Nicholas Sherck:** Conceptualization (equal); Funding acquisition (equal); Writing – review & editing (equal). **Stephan Köhler:** Conceptualization (equal); Funding acquisition (equal); Writing – review & editing (equal). **Rohini Gupta:** Conceptualization (equal); Funding acquisition (equal); Writing – review & editing (equal). **Matthew B. Francis:** Conceptualization (equal); Funding acquisition (equal); Writing – review & editing (equal). **M. Scott Shell:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Resources (equal); Supervision (equal). **Glenn H. Fredrickson:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Resources (equal); Supervision (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

- O. G. Mouritsen and K. Jørgensen, "A new look at lipid-membrane structure in relation to drug research," *Pharm. Res.* **15**, 1507–1519 (1998).
- E. Goddard, "Polymer/surfactant interaction—Its relevance to detergent systems," *J. Am. Oil Chem. Soc.* **71**, 1–16 (1994).
- M. Singh, M. Briones, G. Ott, and D. O'Hagan, "Cationic microparticles: A potent delivery system for DNA vaccines," *Proc. Natl. Acad. Sci. U. S. A.* **97**, 811–816 (2000).
- W. C. Blocher McTigue and S. L. Perry, "Design rules for encapsulating proteins into complex coacervates," *Soft Matter* **15**, 3089–3103 (2019).
- J. H. Clint, "Micellization of mixed nonionic surface active agents," *J. Chem. Soc., Faraday Trans. 1* **71**, 1327–1334 (1975).
- W. C. Presto and W. Preston, "Some correlating principles of detergent action," *J. Phys. Colloid Chem.* **52**, 84–97 (1948).
- M. J. Rosen and J. T. Kunjappu, *Surfactants and Interfacial Phenomena* (John Wiley & Sons, 2012).
- S. A. Sanders and A. Z. Panagiotopoulos, "Micellization behavior of coarse grained surfactant models," *J. Chem. Phys.* **132**, 114902 (2010).
- M.-T. Lee, A. Vishnyakov, and A. V. Neimark, "Calculations of critical micelle concentration by dissipative particle dynamics simulations: The role of chain rigidity," *J. Phys. Chem. B* **117**, 10304–10310 (2013).
- A. Vishnyakov, M.-T. Lee, and A. V. Neimark, "Prediction of the critical micelle concentration of nonionic surfactants by dissipative particle dynamics simulations," *J. Phys. Chem. Lett.* **4**, 797–802 (2013).
- A. Jusufi and A. Z. Panagiotopoulos, "Explicit- and implicit-solvent simulations of micellization in surfactant solutions," *Langmuir* **31**, 3283–3292 (2015).

- ¹²A. P. Santos and A. Z. Panagiotopoulos, "Determination of the critical micelle concentration in simulations of surfactant systems," *J. Chem. Phys.* **144**, 044709 (2016).
- ¹³J. Zhou and A.-C. Shi, "Critical micelle concentration of micelles with different geometries in diblock copolymer/homopolymer blends," *Macromol. Theory Simul.* **20**, 690–699 (2011).
- ¹⁴S. Qin, T. Jin, R. C. Van Lehn, and V. M. Zavala, "Predicting critical micelle concentrations for surfactants using graph convolutional neural networks," *J. Phys. Chem. B* **125**, 10610–10620 (2021).
- ¹⁵V. Sresht, E. P. Lewandowski, D. Blankschtein, and A. Jusufi, "Combined molecular dynamics simulation–molecular-thermodynamic theory framework for predicting surface tensions," *Langmuir* **33**, 8319–8329 (2017).
- ¹⁶M. Lechuga, M. Fernández-Serrano, E. Jurado, J. Núñez-Olea, and F. Ríos, "Acute toxicity of anionic and non-ionic surfactants to aquatic organisms," *Ecotoxicol. Environ. Safety* **125**, 1–8 (2016).
- ¹⁷S. A. Sanders, M. Sammalkorpi, and A. Z. Panagiotopoulos, "Atomistic simulations of micellization of sodium hexyl, heptyl, octyl, and nonyl sulfates," *J. Phys. Chem. B* **116**, 2430–2437 (2012).
- ¹⁸F. H. Quina, P. M. Nassar, J. B. Bonilha, and B. L. Bales, "Growth of sodium dodecyl sulfate micelles with detergent concentration," *J. Phys. Chem.* **99**, 17028–17031 (1995).
- ¹⁹B. L. Bales, "A definition of the degree of ionization of a micelle based on its aggregation number," *J. Phys. Chem. B* **105**, 6798–6804 (2001).
- ²⁰A. Del Regno, P. B. Warren, D. J. Bray, and R. L. Anderson, "Critical micelle concentrations in surfactant mixtures and blends by simulation," *J. Phys. Chem. B* **125**, 5983–5990 (2021).
- ²¹B. Wen, B. Bai, and R. G. Larson, "Surfactant desorption and scission free energies for cylindrical and spherical micelles from umbrella-sampling molecular dynamics simulations," *J. Colloid Interface Sci.* **599**, 773–784 (2021).
- ²²R. Becker and W. Döring, "Kinetische behandlung der keimbildung in übersättigten dämpfen," *Ann. Phys.* **416**, 719–752 (1935).
- ²³J. A. Mysona, A. V. McCormick, and D. C. Morse, "Mechanism of micelle birth and death," *Phys. Rev. Lett.* **123**, 038003 (2019).
- ²⁴E. Aniansson and S. N. Wall, "Kinetics of step-wise micelle association," *J. Phys. Chem.* **78**, 1024–1030 (1974).
- ²⁵M. Nguyen, N. Sherck, K. Shen, C. E. Edwards, B. Yoo, S. Köhler, J. C. Speros, M. E. Helgeson, K. T. Delaney, M. S. Shell, and G. H. Fredrickson, "Predicting polyelectrolyte coacervation from a molecularly informed field-theoretic model," *Macromolecules* **55**, 9868 (2022).
- ²⁶V. Y. Borue and I. Y. Erukhimovich, "A statistical theory of globular polyelectrolyte complexes," *Macromolecules* **23**, 3625–3632 (1990).
- ²⁷K. T. Delaney and G. H. Fredrickson, "Theory of polyelectrolyte complexation—Complex coacervates are self-coacervates," *J. Chem. Phys.* **146**, 224902 (2017).
- ²⁸G. H. Fredrickson, V. Ganesan, and F. Drolet, "Field-theoretic computer simulation methods for polymers and complex fluids," *Macromolecules* **35**, 16–39 (2002).
- ²⁹G. H. Fredrickson and K. T. Delaney, "Direct free energy evaluation of classical and quantum many-body systems via field-theoretic simulation," *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2201804119 (2022).
- ³⁰S. H. Klass, M. J. Smith, T. A. Fiala, J. P. Lee, A. O. Omole, B.-G. Han, K. H. Downing, S. Kumar, and M. B. Francis, "Self-assembling micelles based on an intrinsically disordered protein domain," *J. Am. Chem. Soc.* **141**, 4291–4299 (2019).
- ³¹S. H. Klass, J. M. Gleason, A. O. Omole, B. Onoa, C. J. Bustamante, and M. B. Francis, "Preparation of bioderived and biodegradable surfactants based on an intrinsically disordered protein sequence," *Biomacromolecules* **23**, 1462–1470 (2022).
- ³²J. Xie and A.-C. Shi, "Formation of complex spherical packing phases in diblock copolymer/homopolymer blends," *Giant* **5**, 100043 (2021).
- ³³Y. Zhu, B. Zheng, L. Zhang, D. Andelman, and X. Man, "Formation of diblock copolymer nanoparticles: Theoretical aspects," *Giant* **10**, 100101 (2022).
- ³⁴M. Nguyen, K. Shen, N. Sherck, S. Köhler, R. Gupta, K. T. Delaney, M. S. Shell, and G. H. Fredrickson, "A molecularly informed field-theoretic study of the complexation of polycation pldadma with mixed micelles of sodium dodecyl sulfate and ethoxylated surfactants," *Eur. Phys. J. E* **46**, 75 (2023).
- ³⁵K. Shen, M. Nguyen, N. Sherck, B. Yoo, S. Köhler, J. Speros, K. T. Delaney, M. S. Shell, and G. H. Fredrickson, "Predicting surfactant phase behavior with a molecularly informed field theory," *J. Colloid Interface Sci.* **638**, 84–98 (2023).
- ³⁶N. Sherck, K. Shen, M. Nguyen, B. Yoo, S. Köhler, J. C. Speros, K. T. Delaney, M. S. Shell, and G. H. Fredrickson, "Molecularly informed field theories from bottom-up coarse-graining," *ACS Macro Lett.* **10**, 576–583 (2021).
- ³⁷M. S. Shell, "Coarse-graining with the relative entropy," *Adv. Chem. Phys.* **161**, 395–441 (2016).
- ³⁸G. H. Fredrickson and K. T. Delaney, *Field-Theoretic Simulations in Soft Matter and Quantum Fluids* (Oxford University Press, 2023), Vol. 173.
- ³⁹G. Fredrickson *et al.*, *The Equilibrium Theory of Inhomogeneous Polymers* (Oxford University Press on Demand, 2006), Vol. 134.
- ⁴⁰J. Mullinax and W. Noid, "Extended ensemble approach for deriving transferable coarse-grained potentials," *J. Chem. Phys.* **131**, 104110 (2009).
- ⁴¹T. Sanyal, J. Mittal, and M. S. Shell, "A hybrid, bottom-up, structurally accurate, Gō-like coarse-grained protein model," *J. Chem. Phys.* **151**, 044111 (2019).
- ⁴²P. Robustelli, S. Piana, and D. E. Shaw, "Developing a molecular dynamics force field for both folded and disordered protein states," *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4758–E4766 (2018).
- ⁴³P. 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 Comput. Biol.* **13**, e1005659 (2017).
- ⁴⁴L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, "Packmol: A package for building initial configurations for molecular dynamics simulations," *J. Comput. Chem.* **30**, 2157–2164 (2009).
- ⁴⁵S. Izadi, R. Anandakrishnan, and A. V. Onufriev, "Building water models: A different approach," *J. Phys. Chem. Lett.* **5**, 3863–3871 (2014).
- ⁴⁶R. Zana, H. Levy, and K. Kwetkat, "Mixed micellization of dimeric (gemini) surfactants and conventional surfactants. I. Mixtures of an anionic dimeric surfactant and of the nonionic surfactants C₁₂E₅ and C₁₂E₈," *J. Colloid Interface Sci.* **197**, 370–376 (1998).
- ⁴⁷D. R. Perinelli, M. Cespi, N. Lorusso, G. F. Palmieri, G. Bonacucina, and P. Blasi, "Surfactant self-assembling and critical micelle concentration: One approach fits all?," *Langmuir* **36**, 5745–5753 (2020).
- ⁴⁸P. J. Flory, "Thermodynamics of high polymer solutions," *J. Chem. Phys.* **9**, 660 (1941).
- ⁴⁹M. L. Huggins, "Solutions of long chain compounds," *J. Chem. Phys.* **9**, 440 (1941).
- ⁵⁰P.-G. De Gennes and P.-G. Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, 1979).
- ⁵¹L. Leibler, H. Orland, and J. C. Wheeler, "Theory of critical micelle concentration for solutions of block copolymers," *J. Chem. Phys.* **79**, 3550–3557 (1983).
- ⁵²A. Ben-Naim and F. Stillinger, "Critical micelle concentration and the size distribution of surfactant aggregates," *J. Phys. Chem.* **84**, 2872–2876 (1980).
- ⁵³M. A. Floriano, E. Caponetti, and A. Z. Panagiotopoulos, "Micellization in model surfactant systems," *Langmuir* **15**, 3143–3151 (1999).
- ⁵⁴A. Bhattacharya and S. Mahanti, "Energy and size fluctuations of amphiphilic aggregates in a lattice model," *J. Phys.: Condens. Matter* **12**, 6141 (2000).
- ⁵⁵P. Carpena, J. Aguiar, P. Bernaola-Galván, and C. Carnero Ruiz, "Problems associated with the treatment of conductivity–concentration data in surfactant solutions: simulations and experiments," *Langmuir* **18**, 6054–6058 (2002).
- ⁵⁶D. L. Mobley and J. P. Guthrie, "FreeSolv: A database of experimental and calculated hydration free energies, with input files," *J. Comput.-Aided Mol. Des.* **28**, 711–720 (2014).
- ⁵⁷J. Lequieu, "Combining particle and field-theoretic polymer models with multi-representation simulations," *J. Chem. Phys.* **158**, 244902 (2023).
- ⁵⁸E. M. Lennon, G. O. Mohler, H. D. Cenicerros, C. J. García-Cervera, and G. H. Fredrickson, "Numerical solutions of the complex Langevin equations in polymer field theory," *Multiscale Model. Simul.* **6**, 1347–1370 (2008).