UNIVERSITY of CALIFORNIA
Santa Barbara

Molecularly Informed Field Theories for Complex Formulation Design

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

by

My Nguyen

Committee in charge:

Professor Glenn Fredrickson, Co-Chair

Professor M. Scott Shell, Co-Chair

Professor Matthew E. Helgeson

Professor Omar Saleh

September 2023

The dissertation of My Nguyen is approved:

Professor Matthew E. Helgeson

Professor Omar Saleh

Professor Glenn Fredrickson

Professor M. Scott Shell

August 2023

Acknowledgements

I am deeply indebted to my family, friends, mentors, colleagues, and collaborators for their unwavering support on this journey.

I'm incredibly thankful for the guidance and mentorship of my advisors, Glenn Fredrickson and M. Scott Shell, at UCSB. Their influence has been instrumental in shaping me into a better researcher.

During my time at the University of Minnesota, I was incredibly fortunate to have Eray Aydil as my undergraduate advisor. His dedication to nurturing scientific curiosity and his support were pivotal to my academic path, inspiring me to pursue graduate school.

I'm grateful to have worked with Kevin Shen and Nick Sherck since day one as a graduate student. They are my closest collaborators and colleagues who have consistently shared their knowledge and engaged in numerous intellectual discussions with me. Additionally, I'm thankful for the individuals I've had the privilege to work with during my time at UCSB, including Kris Delaney, Charles Li, David Zhao, Chelsea Edwards, Kate Dolph, Doug Gerzetic, Sally Jiao, Evan Pretti, Matt Francis, Matt Helgeson, Omar Saleh, and all members of the Fredrickson and Shell groups.

Many thanks to Stephan Köhler, Brian Yoo, Josh Speros, and Rohini Gupta, as well as Keith Gutowski from BASF, with whom I have collaborated throughout the SCOUT project. The extensive discussions we've had have significantly contributed to shaping my thesis.

Most importantly, I'm grateful for my best friend, Hien Nguyen, who has stood by me and provided unwavering emotional support throughout this journey. Lastly, I dedicate this work to my parents, who have given me the opportunity to pursue my education in the US. Their sacrifices and support are something I'm forever thankful for.

Curriculum Vitæ

My Nguyen

Education

2023	Ph.D., Chemical Engineering, University of California, Santa Barbara Advised by Professors Glenn Fredrickson and M. Scott. Shell
2018	B.S., *Summa Cum Laude*, Chemical Engineering, University of Minnesota, Twin Cities

Professional Experience

2018-2023	Graduate Researcher, University of California, Santa Barbara
2016-2018	Undergraduate Researcher, University of Minnesota, Twin Cities
Summer 2017	Research and Development Intern, Bostik, Wauwatosa, WI

Publications

(* equal contribution)

1. **M. Nguyen***, K. Shen*, N. Sherck, S. Köhler, R. Gupta, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "A Molecularly Informed Field-Theoretic Study of the Complexation of Polycation PDADMA with Mixed Micelles of Dodecyl Sulfate and Ethoxylated Surfactants". *The European Physical Journal E*. Accepted (2023).
2. **M. Nguyen**, K. Dolph, N. Sherck, S. Köhler, R. Gupta, M. Francis, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Molecularly Informed Field Theory for Estimating Critical Micelle Concentrations of Intrinsically Disordered Protein Surfactants". In preparation.
3. **M. Nguyen***, N. Sherck*, S. Köhler, R. Gupta, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Developing Molecularly-Informed Field-theoretic Models for Carbohydrates". In preparation.
4. N. Sherck, B. Yoo, **M. Nguyen**, K. Shen, S. Köhler, P. Veld, J. Speros, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Rapid screening of polymer phase behavior using molecularly informed field theories and machine learning". In preparation.
5. A. J. DeStefano, **M. Nguyen**, G. H. Fredrickson, S. Han, R. A. Segalman. "Design of Soft Material Surfaces with Rationally Tuned Water Diffusivity". *ACS Central Science*. 1019-1024 (2023).

6. 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", *J. Colloid Interface Sci.* 638, 84-98 (2023).
7. **M. Nguyen**, N. Sherck, K. Shen, C. E. R. Edwards, B. Yoo, S. Köhler, J. Speros, M. E. Helgeson, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Predicting polyelectrolyte complex coacervation from a molecularly-informed field-theoretic model". *Macromolecules.* 55, 21, 9868–9879 (2022).
8. B. Das, J. T. Batley, K. L. Krycka, J. A. Borchers, P. Quarterman, C. Korostynski, **M. Nguyen**, I. Kamboj, E. S. Aydil, and C. Leighton. "Chemically Induced Magnetic Dead Shells in Superparamagnetic Ni Nanoparticles Deduced from Polarized Small-Angle Neutron Scattering." *ACS Applied Materials & Interfaces.* 14, 33491-33504 (2022).
9. N. Sherck, K. Shen, **M. Nguyen**, B.Yoo, S. Köhler, 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, 576 (2021).
10. K. Shen, N. Sherck, **M. Nguyen**, B. Yoo, S. Koehler, J. Speros, K.T. Delaney, M.S. Shell, G.H. Fredrickson. "Learning composition-transferable coarse-grained models: Designing external potential ensembles to maximize thermodynamic information." *The Journal of Chemical Physics.* 153, 154116 (2020).
11. J. T. Batley, **M. Nguyen**, I. Kamboj, C. Korostynski, E. S. Aydil, and C. Leighton. "Quantitative understanding of superparamagnetic blocking in thoroughly characterized Ni nanoparticle assemblies." *Chemistry of Materials.* 32, 6494-6506 (2020).

Oral Presentations

1. **M. Nguyen**, N. Sherck, S. Köhler, R. Gupta, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Predicting critical micelle concentrations for bio-based surfactants using molecularly informed field theories". California Research Alliance by BASF Spring Meeting 2023, Berkeley, CA.
2. **M. Nguyen**, K. Shen, N. Sherck, S. Köhler, R. Gupta, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Molecularly informed field-theoretic models of surfactant formulations". APS 2023. Las Vegas, NV.
3. **M. Nguyen**, M. S. Shell, G. H. Fredrickson. "Studying polyelectrolyte/surfactant complexation with molecularly informed field theories". Complex Fluids Design Consortium 2023. UCSB.
4. **M. Nguyen**, K. Shen, N. Sherck, S. Köhler, R. Gupta, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Multiscale simulations of surfactants from molecularly-informed field theories. California Research Alliance by BASF Fall Meeting 2022, Santa Barbara, CA.
5. **M. Nguyen**, M. S. Shell, G. H. Fredrickson. "Multiscale simulation of complex

coacervates with molecularly informed field theories from systematic, bottom-up coarse-graining". Annual Graduate Student Symposium 2022, UCSB.

6. **M. Nguyen**, N. Sherck, K. Shen, B. Yoo, S. Köhler, J. Speros, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Predicting Polyelectrolyte Complex Coacervation From A Molecularly-Informed Field-Theoretic Simulation Approach". APS 2022. Chicago, IL.
7. **M. Nguyen**, M. S. Shell, G. H. Fredrickson. "Phase behavior of polyelectrolyte-surfactant complex coacervates from molecularly-informed field theory". Complex Fluids Design Consortium 2022. UCSB.
8. **M. Nguyen**, N. Sherck, K. Shen, B. Yoo, S. Köhler, J. Speros, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Predicting Polyelectrolyte Complex Coacervation From A Molecularly-Informed Field-Theoretic Simulation Approach". AICHE 2021. Boston, MA.
9. **M. Nguyen**, M. S. Shell, G. H. Fredrickson. "Prediction of polyelectrolyte complex coacervation by bridging atomistic and field theoretic models". Complex Fluids Design Consortium 2021. UCSB.
10. **M. Nguyen**, K. Shen, N. Sherck, B. Yoo, S. Köhler, J. Speros, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Integrated particle and field-theoretic simulations for soft materials design". Polymer Physics Symposium 2020. Virtual.

Selected Poster Presentations

1. **M. Nguyen**, N. Sherck, S. Köhler, R. Gupta, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "Parametrizing field theory for applications in care chemicals". BASF California Research Alliance (CARA) Review Meeting 2022. Berkeley, CA.
2. **M. Nguyen**, N. Sherck, K. Shen, B. Yoo, S. Köhler, J. Speros, K. T. Delaney, M. S. Shell, G. H. Fredrickson. "SCOUT: Coacervation of poly(acrylic acid) and poly(allylamine hydrochloride). BASF Research Forum 2021. Virtual.

Teaching Experience

2020, 2021, 2022 ChE 210B: Advanced Topics in Equilibrium Statistical Mechanics

Leadership and Outreach Activities

2022	Publication Committee, Annual Graduate Student Symposium at UCSB
2020-2022	Organizing Committee, Graduate Simulation Seminar Series
2019	Volunteer, Materials Research Outreach Program, MRL at UCSB

Abstract

Molecularly Informed Field Theories for Complex Formulation Design

by

My Nguyen

Understanding the solution behavior of complex soft materials is crucial for designing and optimizing formulations that are relevant in everyday consumer products, including processed foods, detergents, hair care products, and various industrial applications such as lubricants, pesticides, and coatings. These formulations are highly multi-component and involve a wide range of charged molecules, such as polyelectrolytes, surfactants, and colloids, often in the presence of salt and other non-ionic (macro)molecules. While experimental investigations provide valuable insights, they are often limited in their ability to directly observe molecular-level interactions and explore the vast design space, encompassing numerous parameters such as composition, specific chemical species, macromolecule architecture, molecular weight, temperature, pH, and more. Computational simulations offer a powerful tool to complement experimental studies, providing a high-throughput screening approach to deepen our understanding of the underlying molecular interactions and the behavior of complex formulations.

In this thesis, we present a multi-scale simulation approach that parameterizes mesoscopic models of the field theory based on information obtained from small-scale atomistic simulations. We employ the relative entropy minimization framework to derive chemically-sensitive coarse-grained interaction parameters from all-atom simulations. Subsequently, we utilize the exact transformation to convert the coarse-grained particle-based model into field-theoretic form, facilitating the prediction of solution phase behavior. The

overall workflow preserves the chemical specificity in complex mixtures of interest, enabling *de novo* studies of solution phase behavior in the field theory without the need for any experimental input.

The simulation framework is highly adaptable and can be applied to investigate a wide range of soft-matter formulations. This thesis focuses on formulations that rely on the complexation of charged macromolecules. The presence of charged assemblies, such as micelles, in typical formulations introduces further complexity, including long length and time-scale phenomena that are intractable with other high-resolution simulation techniques. Through the exploration of various complex formulations, we demonstrate the predictive capability of this simulation workflow in exploring the thermodynamics and complex structures arising in such formulations. By integrating atomistic and mesoscopic simulation techniques, this work contributes to a fundamental understanding of the underlying mechanisms governing solution behavior in complex formulations. It offers valuable insights for the rational design and optimization of soft matter formulations, thereby contributing to advancements in various industries and sustainable chemical product development.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

This thesis aims to employ molecular simulations to gain insights into the solution behavior of representative formulations found in everyday consumer products. These formulations, besides being highly ubiquitous in processed foods,[3–5] detergent formulations,[6] hair care formulations,[7] also have high potential in advanced oil recovery,[8,9] wet adhesives,[10–13] and drug and gene delivery vehicles.[14–17] Their applications rely heavily on molecular-level interactions among the components and, more importantly, complexation that involves a wide range of natural or synthetic charged molecules, such as synthetic polyelectrolytes, biological polyelectrolytes (e.g., proteins, proteoglycans, polynucleotides), surfactants, and colloids, typically in the presence of salt and other nonionic macromolecules. For instance, rheological properties of shampoo are related to the molecular-level structure of the raw ingredients as well as the strength of complexation

between the charged macro-ions.[7,18,19] Similarly, the ability to enclose dirt particles in detergents heavily depends on the self-assembly of surfactants and their interactions with other molecules, such as polymers.[6] In these formulations, the presence of salt is particularly relevant as it significantly affects the electrostatic interactions and overall behavior of the charged molecules and, consequently, the macroscopic properties.[20–22] Therefore, understanding the solution behavior in these formulations is crucial for various industries as it directly influences their performance and functionality.

Experimental investigations provide valuable insights into the solution phase behavior,[20,21,23] but they are often limited by the difficulty of directly observing and quantifying molecular-level interactions. Moreover, realistic formulations are highly multicomponent, comprising both ionic and non-ionic macromolecules, small ions, and solvents and co-solvents, with numerous design parameters, including composition, specific chemical species, macromolecule architecture, molecular weight, temperature, pH, and more.[7,24] Exploring this vast design space through brute-force experimentation is challenging. Furthermore, with the increasing emphasis on sustainability and environmentally friendly practices in chemistry, it becomes crucial to intelligently probe and select ingredients from the green chemical space while maintaining product performance. Therefore, computational simulations offer a powerful tool to complement experimental studies and provide a detailed understanding of the underlying molecular interactions and the behavior of complex formulations.

Simulation techniques, such as molecular dynamics, Monte Carlo simulations or field-

theoretic simulations,[25–27] offers unique advantages in studying systems involving charged macromolecules. Simulations enable the investigation of molecular-level details and provide access to information that is otherwise challenging to obtain experimentally such as precise local density, chain conformation, hydrogen bonding, etc. They also allow for the systematic manipulation of various parameters, such as charge density, concentration, and salt concentration, facilitating a comprehensive exploration of the effects of these factors on the solution behavior of charged molecules. However, all-atom simulations,[28–30] while providing detailed chemical information and benefiting from extensive force field development efforts, are intractable for investigating complex formation due to the computational cost. To overcome this computational challenge, other studies have employed coarse-grained descriptions of the components. However, these approaches often rely on simplified toy models[31–33] or fitted parameters based on experimental observations[34,35] to represent the intricate interactions among different species. This limits the direct comparison between simulations and actual formulations, as well as restricts the studies to known compounds with available experimental data.

To address these challenges, we have developed and utilized a multi-scale simulation approach that involves parameterizing a mesoscopic model based on information obtained from atomistic simulations. Our work has showcased the predictive capability of this workflow for a range of representative soft matter formulations. Moreover, the simulation framework is highly adaptable and can be applied to investigate other soft-matter formulations beyond those considered in this thesis. This contribution sig-

nificantly enhances our fundamental understanding of the underlying mechanisms that govern solution behavior. It provides valuable insights into the factors influencing stability, self-assembly, and overall performance in complex formulations.

1.2 Complex coacervation in soft matter formulations

One of the most studied complexation mechanisms in mixtures comprised of charged macro-ions is coacervation.[36–39] This process is a liquid-liquid phase separation that is driven by the complexation free energy with entropic and enthalpic contributions.[40] Experiments suggest one major driving force of coacervation is the entropy gain from counterion release.[41] Although there is a loss in configurational and translational entropy of the macro-ions upon complexation, this contribution is small relative to the counterion entropy for high-molecular-weight macro-ions. The resulting dense phase comprised of the majority of the oppositely charged macro-ions is referred to as a complex coacervate. Complex coacervates typically exhibit high viscosity due to the dense packing of charged macromolecules within their liquid phase. This high viscosity is often advantageous for certain applications where enhanced stability and control over the encapsulated materials are desired as well as in personal care products such as shampoo where a gel-like consistency is preferred.[7,42]

Mixtures of oppositely charged polyelectrolytes are the simplest models for coacervate-

forming formulations.[1,21,43–46] Polyelectrolytes are polymers that contain ionizable groups along their backbone, which can dissociate into charged ions when dissolved in a polar solvent such as water. The degree of ionization refers to the extent to which the ionizable groups are dissociated into charged species. Based on the tendency of ionization, polyelectrolytes are classified into strong and weak polyelectrolytes. Strong polyelectrolytes exhibits a high degree of ionization in a solution for most reasonable pH values. In contrast, weak polyelectrolytes are often partially charged and their ionization states depend on the solution pH and the dissociation constant of the ionizable groups. Charge density, charge stochiometry, pH, temperature, and salt concentration are critical factors that play pivotal roles in influencing coacervation in these systems.[20,32,33,44] Although it has been observed that mixtures of polyelectrolyte pairs with high charge density can form solid precipitate at low salt concentrations,[20,44] this thesis mainly focuses on the regime of liquid complexes.

In realistic formulations, there are often additional charged assemblies present, such as micelles, as well as small molecules like active ingredients and perfumes alongside polyelectrolytes. Micelles, which form through the self-assembly of amphiphilic molecules, including surfactants and block copolymers, are particularly valued for their encapsulating properties, detergency, and foaming characteristics.[47–49] Recently, there has been growing interest in a class of colloidal systems called complex coacervate core micelles. These nanostructures feature a complex coacervate core and a neutral hydrophilic corona.[50,51] Complex coacervate core micelles are formed through the spontaneous assembly of ionic-

neutral copolymers and oppositely charged polyelectrolytes, making them an intriguing subject of investigation in the field. The presence of these charged assemblies introduces further complexity to the coacervation behavior and the overall solution properties.[24,52] Understanding the interplay between various interactions, including hydrophobic and electrostatic interactions, in such systems is crucial for the design and optimization of complex formulations.[53]

1.3 Theories and simulations for studying complexation

Voorn-Overbeek theory is a classical theory of coacervation that attempts to understand coacervation by approximating the complexation free energy as the combination of the Flory-Huggins mixing entropy and the Debye-Hückle electrostatic energy:[54,55]

$$\frac{a^3\beta F}{V} = \sum_i \frac{\phi_i}{N_i}\ln(\phi_i) + \sum_i \sum_{j<i} \chi_{ij}\phi_i\phi_j - \frac{\sqrt{4\pi}}{3}\left(\frac{l_B}{a}\right)^{3/2}\left(\sum_i \sigma_i\phi_i\right)^{3/2} \tag{1.1}$$

where $\beta = 1/k_BT$, a is the size of a monomer. The volume fraction, chain length, and charge density of species i are denoted ϕ_i, N_i, and σ_i, respectively. The Flory-Huggins χ_{ij} describes excluded volume interactions between species i and j. The last term is the Debye-Hückle where $l_B = e2/4\pi r k_B T$ is the Bjerrum length. Although being widely used to explain experimental observations,[35,50] it has certain limitations that should be

acknowledged. Primarily, this theory assumes a mean-field electrostatic environment around individual ions, which is only valid in dilute electrolytes (5-10 mM).[56] Consequently, the theory fails to provide accurate quantitative predictions at higher salt concentrations commonly encountered in coacervate systems, reaching up to approximately 3 M.[1] Additionally, the theory treats polymer charges as unconnected ions, which is an oversimplified representation of polyelectrolytes that often exhibit high linear charge density.[33]

The limitations of the Voorn-Overbeek theory emphasize the need for more advanced theoretical models and simulation techniques that can account for the intricate interactions and diverse molecular structures present in coacervating systems. One such theoretical model is the Edwards approach to polymer field theory properly connect the bound charges to the chain conformations.[57] For such models, the Gaussian approximation (often referred to in the literature as the random phase approximation) is a common analytical method providing the lowest-order perturbation correction to mean-field theory, where the latter is qualitatively inadequate for describing electrostatically driven phase separation.[58–64] Prior studies have shown that the Gaussian approximation captures a significant portion of the fluctuation effects.[27,63,65–69] At low charge densities, however, it overestimates the strength of charge correlations and the size of the two-phase region, i.e., the supernatant phase is predicted to be overly depleted of polyelectrolytes by many orders of magnitude.[65,68,70]

While traditional theories have been valuable for gaining fundamental understanding

of coacervation, they may not be suitable for investigating the intricacies of complex systems encountered in modern formulation design. These systems often consist of multiple species of different lengthscales and advanced architectures, necessitating a more comprehensive and sophisticated approach to accurately capture their behavior. In this context, simulations have emerged as powerful tools that offer insights into regimes beyond the reach of analytical techniques. Among particle-based simulation methods used in the field of soft matter formulation, dissipative particle dynamics (DPD) has gained popularity as a mesoscopic simulation technique.[71–73] In DPD, particles represent groups of atoms and interact through pairwise dissipative and random forces. This approach is particularly attractive due to its efficiency in accessing longer length and time scales compared to higher resolution techniques, such as all-atom molecular dynamics simulations. While the soft potentials used in DPD allow for longer time steps and the study of larger systems, probing macroscopic phenomena and understanding the complex behaviors of soft matter formulations can still be more efficiently achieved through continuum mesoscopic approaches, such as field-theoretic simulations (FTS).

FTS utilize the exact statistical field theory to account for field fluctuation effects, which are critical for the study of coacervation.[68,74–76] FTS replaces particle-particle interactions with interactions between individual particles and one or more fluctuating fields, allowing for efficient sampling of configurational integrals over particle configurations. As a result, this simulation method is uniquely suited to the study of high molecular weight and high density charged systems. Moreover, in the field representation the species'

chemical potentials are readily available from ensemble-averaged operators without the need for sophisticated, often expensive, free energy techniques required in molecular dynamics simulations.[77] A range of studies have used FTS to elucidate the effects of charge patterning, electrostatic environment and inclusion of explicit counterions on complex coacervation.[27,65–69,78]

Despite significant efforts in theory and simulation, incorporating the complex interaction environments arising from charged moieties remains challenging. For instance, theoretical models often assume that the polyanion and polycation, as well as the anion and cation, in complex coacervates have identical chemical structures except for opposite charges. However, this assumption leads to nonphysical symmetric phase diagrams concerning the oppositely charged species.[79,80] In many simulation studies, chemistry-specific effects are simplified through parameters such as persistence length or non-bonded cut-off distances.[26,31] Phenomenological parameters associated with coarse-grained potentials, such as χ parameters,[34] are also introduced to capture chemistry-specific interactions, but they require prior knowledge of the component chemistries and lack predictive capabilities.

In this regard, high-resolution approaches such as atomistic molecular dynamics (MD) offer a better alternative for incorporating the underlying chemical building blocks. A few studies have attempted to calculate the coacervation binodal using MD that employ the slab geometry to simulate the coexisting dense and dilute phases.[32] This method is, however, prone to finite-size errors due to the large interfaces between the coexising phases.

Moreover, due to the computational constraints, the system size is typically limited to a few nanometers and the simulation time is limited to nanoseconds or microseconds. These limitations restrict the direct investigation of long length and time-scale phenomena, such as the formation of coacervate, self-assembly of surfactants, diffusion, macro-ion exchange, and fission and fusion of charged assemblies, in MD simulations. Despite these challenges, MD simulations can provide valuable insights into the underlying molecular mechanisms and local structure within coacervating systems, which can inform and complement experimental observations.[28,31]

In this work, we employ a bottom-up simulation strategy that uses the atomistic simulations to parameterize field-theoretic models. Briefly, we use relative entropy coarse-graining to derive chemically-sensitive coarse-grained interaction parameters from all-atom simulations.[81] Subsequently, we employ the exact transformation to take the coarse-grained particle-based model into field-theoretic form,[74] facilitating the prediction of the solution phase behavior. The overall workflow preserves the chemical specificity in complex mixtures of interest, enabling *de novo* studies of solution phase behavior in the field theory bypassing the need for extensive experimental input. This approach enables us to explore the behavior of complex systems and the underlying mechanisms governing formation of complexation between macromolecules. Through the integration of the atomistic and mesoscopic simulation techniques, we aim to provide valuable insights into the equilibrium structural and thermodynamic properties of these systems, ultimately contributing to the rational design and optimization of soft matter formulations.

1.4 Outline

The thesis is organized as follow:

- Chapter 2: Constructing Molecularly Informed Field-Theoretic Models
- Chapter 3: Coacervation of Oppositely Charged Polyelectrolytes[82]
- Chapter 4: Complexation between Polyelectrolyte with Oppositely Charged Micelles[83]
- Chapter 5: Estimating Critical Micelle Concentrations of Intrinsically Disordered Protein Surfactants
- Chapter 6: Effect of the Acetylation Pattern on the Miscibility of Cellulose Acetate

Chapter 2 provides a comprehensive overview of the computational details involved in constructing field-theoreric models of complex formulations. We discuss the selection of the all-atom reference simulation, the mapping from all-atom to coarse-grained representation, and the particle-to-field transformation. These steps outline the main simulation workflow employed throughout this thesis. Subsequently in Chapter 3, we demonstrate that the proposed simulation framework accurately captures important effects in coacervation of a well-characterized polyelectrolyte mixture. Chapter 4 investigates the adsorption of polyelectrolytes on micelles and explores equilibrium complex structures, establishing a systematic approach to studying complex formulations. In the last two chapters, we extend the study to bio-based components which are promising biodegradable alternatives to the synthetic counterparts. Chapter 5 showcases the capability of

molecularly informed field theories in predicting the critical micelle concentration of intrinsically disordered protein surfactants. Finally in Chapter 6, we provide a discussion on how different acetylation pathways can lead to the diverse solution phase behaviors of cellulose acetate observed in experiments.

Chapter 2

Constructing Molecularly Informed Field-Theoretic Models

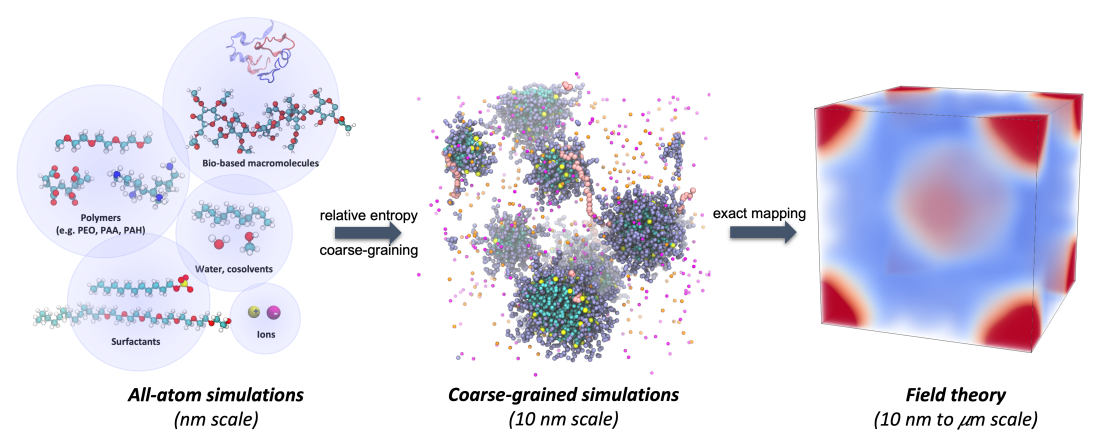

Figure 2.1: Schematic of the multi-scale simulation workflow to parameterize a molecularly informed field-theoretic model.

We first introduced molecularly informed field theories in our 2021 publication,[84] and since then, we have applied the workflow to study various systems, ranging from aqueous and non-aqueous mixtures to formulations involving surfactants, polyelectrolytes, and

block copolymers. In this chapter, our focus is to provide a comprehensive discussion on the construction of a molecularly informed field theory model, which can be summarized into three steps:

1. Performing reference all-atom (AA) simulations
2. Deriving the interaction parameters for the coarse-grained (CG) by minimizing the relative entropy
3. Transforming the CG particle-based model into the field representation

The overall workflow enables us to directly determine the free energy and chemical potential necessary for phase diagram prediction while preserving important information about the underlying chemical components.

2.1 All-atom simulations

In classical AA simulations, the interactions between the atoms in the system are described via a potential energy function of the atomic coordinates, also called a force field. The classical force field approximates the quantum ground-state potential energy surface due to electronic structure. Generally, such potential energy function for a system of n particles has the form

$$
\begin{aligned}
U_{AA}(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_n}) = & \sum_{bonds} k_b(r - r_0)^2 \\
& + \sum_{angles} k_\theta(\theta - \theta_0)^2 \\
& + \sum_{dihedrals} V_x(1 + cos(x\phi - \gamma)) \\
& + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{r_{ij}}{\sigma_{ij}} \right)^{-12} - \left(\frac{r_{ij}}{\sigma_{ij}} \right)^{-6} \right]
\end{aligned}
\tag{2.1}
$$

where $k_b, r_0, k_\theta, \theta_0, V_x, \gamma, q_i, q_j, \sigma_{ij}$ are parameters of the force field. These parameters are taken from a combination of electronic structure calculations on small molecules and experimental data, e.g., crystallographic and microwave data, vibrational frequencies, heat of vaporization. Thus, these classical force fields are semi-emperical. Classical force fields are developed by a wide community and there now exist accurate and systematic approaches to finding reasonably good models for a wide variety of organic systems. Some of the major families of classical force fields are AMBER, CHARMM, GROMOS, and OPLS.[85–88] Often time, we adopt force fields that have been developed for specific system of interest or generalize force fields that can reasonably reproduce experimental observations to perform AA simulations.

2.2 Coarse-graining

2.2.1 Atomistic-to-coarse-grained mapping

To obtain reference trajectories for coarse-graining, we translate AA trajectories by mapping center-of-mass coordinates of groups of atoms in the AA representation. Equiva-

lently, the coordinates of coarse-grained (CG) sites are defined as:

$$\mathbf{R}_I = \left(\sum_{i \in atoms\, in\, I} m_i \mathbf{r}_i \right) \left(\sum_{i \in atoms\, in\, I} m_i \right)^{-1} \tag{2.2}$$

where $\mathbf{R}_I$ are coordinates of CG site I, $\mathbf{r}_i$ and m_i are coordinates and mass of atom i in the AA model. While there are different ways to define this mapping function, each small molecule, including water and salt ions, is mapped to a single bead in our work. For macromolecules, a common approach is to map one residue to one CG bead. The charges of the CG beads, σ_α, are assigned based on the net charge of the corresponding group of atoms in the AA description. For example, the water bead carries a neutral charge, while the Na^+ and Cl^- ions carry a $+1$ and -1 charge, respectively. It is important to note that the choice of CG mapping in our studies is not unique, and different types of CG beads can lead to different force fields. However, the focus of this work is not on exploring different mapping schemes, and therefore, this aspect is beyond the scope of our research.

2.2.2 Relative entropy coarse-graining

After mapping the AA trajectories to the CG degree of freedoms, we use relative entropy coarse-graining[81,89–91] with these as references to derive CG interaction potentials that are amenable to efficient field-theoretic simulations. Briefly, this is a bottom-up parameterization strategy to derive interaction parameters for the CG model from the higher resolution AA trajectories. Since a detailed discussion of the method is outlined in a monograph by Shell in ref. 81, we only summarize the main points here. The relative

entropy, S_{rel}, describes how accurately a CG model can reproduce properties of an AA model and is given by

$$S_{rel} = \int \wp_{AA}(\mathbf{R}) \ln \frac{\wp_{AA}(\mathbf{R}))}{\wp_{CG}(\mathbf{R})} d\mathbf{R} \tag{2.3}$$

where $\wp_{CG}(\mathbf{R})$ is the equilibrium ensemble probability for the CG configurations, and $\wp_{AA}(\mathbf{R}) = \int \wp_{AA}(\mathbf{r})\delta(M(\mathbf{r}) - \mathbf{R})d\mathbf{r}$ is the equilibrium ensemble probability for atomic positions $\mathbf{r}$ that are projected onto CG coordinates $\mathbf{R}$.

The relative entropy can be interpreted as a measure of the information loss in moving from a higher resolution system to a system with fewer degrees of freedom. The optimal CG potential is determined by minimizing S_{rel} with respect to parameters of the potential. For a parameter λ, this gives the condition

$$\frac{\partial S_{rel}}{\partial \lambda} = \left\langle \frac{\partial U_{CG}}{\partial \lambda} \right\rangle_{AA} - \left\langle \frac{\partial U_{CG}}{\partial \lambda} \right\rangle_{CG} = 0 \tag{2.4}$$

where the angle brackets indicate averages over the AA and CG ensembles. If the AA and CG models' potential energy consists of pair potentials, the optimal CG pair interaction, $u_{CG}(R)$, satisfies

$$\begin{aligned} \frac{\delta S_{rel}}{\delta u_{CG}(R)} &= \left\langle \frac{\delta U_{CG}[u_{CG}(R)]}{\delta u_{CG}(R)} \right\rangle_{AA} - \left\langle \frac{\delta U_{CG}[u_{CG}(R)]}{\delta u_{CG}(R)} \right\rangle_{CG} \\ &= \left\langle \sum_{i<j} \delta[R - R_{ij}] \right\rangle_{AA} - \left\langle \sum_{i<j} \delta[R - R_{ij}] \right\rangle_{CG} \\ &= g_{AA}(R) - g_{CG}(R) = 0 \end{aligned} \tag{2.5}$$

where $g(R)$ is the radial distribution function. Hence, the optimal CG model is one that best reproduce the AA radial distribution functions.

2.2.3 Coarse-grained force field

In our CG interaction model, bonded interactions are described using a harmonic bond potential:

$$\beta U_{b,\alpha\gamma}(r) = \frac{3}{2b_{\alpha\gamma}^2} r^2 \tag{2.6}$$

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 non-bonded interactions, we opt for soft, regularized potentials that are amenable to the field theory. These nonbonded interactions consist of an excluded volume and smeared Coulomb interaction between all site pairs, including bonded pairs:

$$\beta U_{ev,\alpha\gamma} = v_{\alpha\gamma} e^{-r^2/2(a_\alpha^2 + a_\gamma^2)} \tag{2.7}$$

$$\beta U_{el,\alpha\gamma} = \frac{l_B \sigma_\alpha \sigma_\gamma}{r} \operatorname{erf}\left(\frac{r}{2\sqrt{a_\alpha^2/2 + a_\gamma^2/2}} \right) \tag{2.8}$$

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. The charge σ_α is fixed to the net charge of the corresponding CG bead. In addition, we also fix the 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 typically set to 0.31 nm in our work. The Bjerrum length l_B charaterizes the strength of the electrostatic

interactions and is chosen to be 0.74 nm, which is that of OPC water at 298 K and 1 atm[92]. The smeared Coulomb interaction behaves as an unscreened Coulomb $\sim 1/r$ at large separations r but is regularized to be finite at $r = 0$ to accommodate the soft-core repulsions of the Gaussian repulsive excluded volume potential. Our 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. The remaining parameters to be optimized, playing a role as λ in Eq. 2.4, are the root-mean-square bond lengths, $b_{\alpha\gamma}$, and the excluded volume strengths, $v_{\alpha\gamma}$.

For complex systems with many components, we derive these CG parameters in stages. This allows us to systematically reduce the parameter space in each coarse-graining step. Typically, the earlier stages involve studying pure component solutions or binary mixtures, while the later stages involve mixtures of multiple components. Once the parameters are determined for a specific stage, they remain fixed for subsequent steps in the coarse-graining process. This coarse-graining procedure subsequently determines a set of AA simulations one needs to perform.

2.3 Field theory transformation

The CG potential defined in Section 2.2.3 can be readily represented and simulated in a field theoretic representation via the Hubbard-Stratonovich-Edwards transformation. This field-theoretic transformation decouples nonbonded pair interactions, resulting in particles interacting only with an auxiliary field. As a result, particle coordinates can

be analytically integrated, yielding a partition function in terms of integrals over field configurations:

$$\mathcal{Z} = \int d\mathbf{r}^n \; e^{-\beta U(\mathbf{r}^n)} \rightarrow \int \mathcal{D}\mathbf{w} \; e^{-H[\mathbf{w}]} \tag{2.9}$$

where H is an effective Hamiltonian describing the statistical weight of the auxiliary field configuration $\mathbf{w}(\mathbf{r})$, and is systematically described in ref. 74. 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 Section 6.2.2. As a result, the field-theoretic transformation is *exact* for the CG model, and full field-theoretic sampling of the partition function is equivalent to performing coarse-grained molecular dynamics (CGMD). To accomplish this, complex Langevin simulations are employed to sample the complex weights in field theory models.[76] One critical advantage of the field representation is that the species' chemical potentials and free energy can be readily obtained from ensemble-averaged operators, eliminating the need for sophisticated and often computationally expensive free energy techniques required in molecular dynamics simulations.[77] This capability allows for the determination of phase diagrams that would otherwise be challenging to obtain using particle-based techniques.[93]

Another strength of the field-theoretic representation is that it is amenable to a host of analytical tools that facilitate its evaluation, including the mean field approximation, also termed self-consistent field theory (SCFT):

$$\mathcal{Z} \approx e^{-H[\mathbf{w}^*]} \equiv e^{-H^*} \tag{2.10}$$

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. In addition to the mean-field treatment, the Gaussian approximation is another frequently used analytical method that offers the lowest-order perturbation correction to mean-field theory, which is qualitatively insufficient for accurately describing phase separation driven by electrostatic interactions.[58–64] Despite not capturing the fluctuating field configurations, both of these analytical approximations prove to be surprisingly valuable in the investigation of self-assembly and macrophase separation phenomena in polymeric systems.[27,66,68,94]

2.4 Other considerations

There are two potential sources of errors can arise from this workflow: (1) the accuracy of the AA force field, and (2) the fidelity of the CG model to the AA reference. Error of type (1) can be mitigated by the development of new force fields with increased accuracy; this efforts have been a focus of many research groups and is outside the scope of this thesis. Improvement of error type (2), on the other hand, can be addressed by considering:

- CG potential energy functional
- AA-to-CG mapping scheme
- Coarse-graining protocol

CG models with more sophisticated potential energy functionals are anticipated to provide improved thermodynamic representations of the AA model. However, in the

present thesis, we litmit the treatment of excluded volume effects to pairwise Gaussian interactions (Eq. 2.7). This choice allows the ease of computing its functional inverse, which is necessary in the Hubbard-Stratonovich transformation. Nevertheless, recent investigations have proposed methods to obtain an effective inverse for realistic pairwise interactions, thereby enabling the incorporation of a broader range of pair interaction functions.[95,96]

The choice of mapping operator $\mathbf{M}$ is another factor that can affect the quality of the CG model. Previous studies showed that the 1-site and 3-site model of heptane, surprisingly, is better at capturing the radial distribution function than the 2-site model.[97] Furthermore, it has been observed that specific mapping functions provide more informative representations, leading to simpler potential of mean force (PMF) profiles that are easier to capture.[98] These findings underscore the significance of carefully selecting an appropriate mapping operator to enhance the fidelity of the CG model.

The coarse-graining protocol is critical to the resulting CG model. This involves decisions regarding the order of parameter optimization and the choice of simulation ensemble. In the bottom-up coarse-graining approach, the goal is to find an effective potential that accurately represents the underlying PMF. Typically, this optimization is performed at a single state point, which means that the resulting CG model may not be suitable for describing other state points that were not considered during the parameterization process. To address this limitation, one common approach is to perform the coarse-graining in the extended ensemble.[99] In this framework, a single CG model is

parameterized by combining information from an ensemble of AA reference simulations conducted at different state points, such as compositions, conformations, isomerization states, chemistries, and so on. A closely related approach, discussed in our recent work,[100] employs spatially-varying external potentials to inhomogeneous response in the composition. This can be viewed as an extended ensemble in the composition space, but with continuous changes in the composition. In fact, we have shown that with a simple potential energy functional such as that of Eq. 2.7, the external potential ensemble can produce a CG model that achieves near-quantitative capture of activity coefficients across the entire composition range. An application of this approach on aqueous solutions of NaCl is discussed in Appendix B.

Chapter 3

Coacervation of Oppositely Charged Polyelectrolytes

Reproduced in part with permission from:

Nguyen, M., Sherck, N., Shen, K., Edwards, C.E., Yoo, B., Köhler, S., Speros, J.C., Helgeson, M.E., Delaney, K.T., Shell, M.S. and Fredrickson, G.H., 2022. Predicting Polyelectrolyte Coacervation from a Molecularly Informed Field-Theoretic Model. Macromolecules, 55(21), pp.9868-9879.

3.1 Introduction

Many technological applications and everyday consumer formulations involve complexation of oppositely charged macro-ions, including processed foods,[3–5] detergent formulations,[6] advanced oil recovery,[8,9] wet adhesives,[10–13] and drug and gene delivery ve-

hicles.[14–17] The complex phase can form in aqueous mixtures consisting any number of natural or synthetic charged species, including synthetic polyelectrolytes, biological polyelectrolytes (e.g., proteins, proteoglycans, polynucleotides), surfactants and colloids, typically in the presence of salt. These mixtures can undergo a liquid-liquid phase separation with the majority of the oppositely charged macro-ions partitioning into the dense phase, which is referred to as a complex coacervate.[36–39]

The formation of complex coacervates is governed by the complexation free energy with entropic and enthalpic contributions.[40] Experiments suggest that one major driving force of coacervation is the entropy gain from counterion release.[41] Although there is a loss in configurational and translational entropy of the macro-ions upon complexation, this contribution is small relative to the counterion entropy in the limit of long polymers. The aforementioned entropic gain and the enthalpic driving force due to favorable electrostatic interactions weaken as the salt concentration is increased. Adding salt above a critical concentration leads to the dissolution of the complex coacervate due to the salt's screening of electrostatic interactions.[20,101] Many other variables have been shown to influence coacervation behavior, including the chemical nature of the charged species, stoichiometry, pH, polymer size, and temperature.[1,20,21,35,39,79]

Voorn-Overbeek theory is a classical theory of coacervation, which attempts to understand coacervation by approximating the complexation free energy as the combination of the Flory-Huggins mixing entropy and the Debye-Hückle electrostatic energy.[54,55] The theory, however, only captures the mean-field electrostatic environment around individ-

ual ions which is only valid in dilute electrolytes (5-10 mM),[56] resulting in quantitative inadequacies at typical coacervate salt concentrations (up to $\sim$ 3 M).[1] Furthermore, polymer charges are treated as unconnected ions; this is an unrealistic depiction of polyelectrolytes that can have high linear charge density. Other theoretical models such as the Edwards approach to polymer field theory properly connect the bound charges to the chain conformations.[57] For such models, the Gaussian approximation (often referred to in the literature as the random phase approximation) is a common analytical method providing the lowest-order perturbation correction to mean-field theory, where the latter is qualitatively inadequate for describing electrostatically driven phase separation.[58–64] Prior studies have shown that the Gaussian approximation captures a significant portion of the fluctuation effects.[27,63,65–69] At low charge densities, however, it overestimates the strength of charge correlations and the size of the two-phase region, that is, the supernatant phase is predicted to be overly depleted of polyelectrolytes by many orders of magnitude.[65,68,70]

Simulations, on the other hand, provide in-roads to probe regimes where analytical techniques cannot push further. Field-theoretic simulations (FTSs) utilize the exact statistical field theory to account for field fluctuation effects, which are critical for the study of coacervation.[68,74–76] FTS replaces particle-particle interactions with interactions between individual particles and one or more fluctuating fields, allowing for efficient sampling of configurational integrals over particle configurations. As a result, this simulation method is uniquely suited to the study of high-molecular weight and high-density charged

systems. Moreover, in the field representation the species' chemical potentials are readily available from ensemble-averaged operators without the need for sophisticated, often expensive, free energy techniques required in molecular dynamics (MD) simulations.[77] A range of studies have used FTS to elucidate the effects of the charge patterning, electrostatic environment and inclusion of explicit counterions on complex coacervation.[27,65–69,78]

Despite significant effort by theory and simulation with fine-tuned models beyond Voorn-Overbeek theory, the asymmetry in coacervation phase diagrams due to the different chemical structures of the involved charged moieties proves difficult to capture accurately. For example, it is often assumed in theoretical models that the polyanion and polycation (and the anion and cation) have the same chemical structure except for opposite charges, resulting in nonphysical symmetric phase diagrams with respect to the oppositely charged species.[79,80] In other studies, chemistry-specific effects are simply reduced to parameters such as the persistence length or non-bonded cut-off[26,31] or introduced by including phenomenological parameters associated with CG potentials (e.g., χ parameters).[34] The latter, although providing chemistry-specific interactions that are consistent with experiments, requires a priori knowledge of the component chemistries and thus is not predictive.

We present here a molecularly informed field-theoretic model of polyelectrolyte coacervation, which utilizes a bottom-up coarse-graining approach discussed in Chapter 2 and in our previous publication.[84] We build upon our prior work–a multiscale approach to simulating binary water-PEO phase behavior–by introducing more components and

electrostatic interactions. In short, the approach leverages the efficiency of the field theory in simulating systems that are large, dense, and composed of high-molecular weight species, while providing chemical specificity through small-scale, all-atom (AA) simulations. We derive parameters for field theory by performing a systematic coarse-graining of representative AA simulations, using a strategy based on the minimization of the relative entropy, S_{rel}.[81,89–91] The proposed simulation strategy enables efficient calculation of phase diagrams in the field representation while retaining chemical details via systematic coarse-graining of AA simulations. We demonstrate the predictive capability by comparing the phase behavior of a well-characterized model coacervate system consisting of poly(acrylic acid) (PAA) and poly(allylamine hydrochloride) (PAH) in a sodium chloride (NaCl) aqueous solution. PAA, which serves as the polyanion species, has a pK_a of 4.5 while PAH, which is the polycation species, has a pK_a of 8.5.[102,103] Coacervation in a mixture of PAA and PAH is sensitive to the pH, salt concentration, and molecular weight of the polyelectrolytes.[1,20,101] This sensitivity results from the variable ionization state of the weak polyelectrolytes and the screening of the long-range electrostatics upon the addition of salt.[104] For the sake of simplicity, we only investigate coacervation in mixtures of fully charged PAA and PAH, corresponding to a pH $\approx$ 6.5, with the polyelectrolytes having the same degrees of polymerization.

3.2 Computational details and methods

3.2.1 Reference all-atom systems and bottom-up coarse-graining procedure

We refer the reader to Chapter 2 for a detailed discussion on the workflow for developing a molecularly-informed field theory. Here, we define five CG bead types for the coacervate model: PAA and PAH monomers, Na^+, Cl^-, and water (Fig. 6.4). CG sites are generally defined as center-of-mass coordinates of groups of atoms in the AA representation. For the CG model presented in this paper, we use this mapping for each small molecule (water, Na^+ and Cl^-) and each polyelectrolyte monomer, to obtain mapped AA reference trajectories for coarse-graining. In addition, CG beads represent the PAA monomer (p_-), PAH monomer (p_+), Na^+, and Cl^- each carry an integer charge of ± 1.

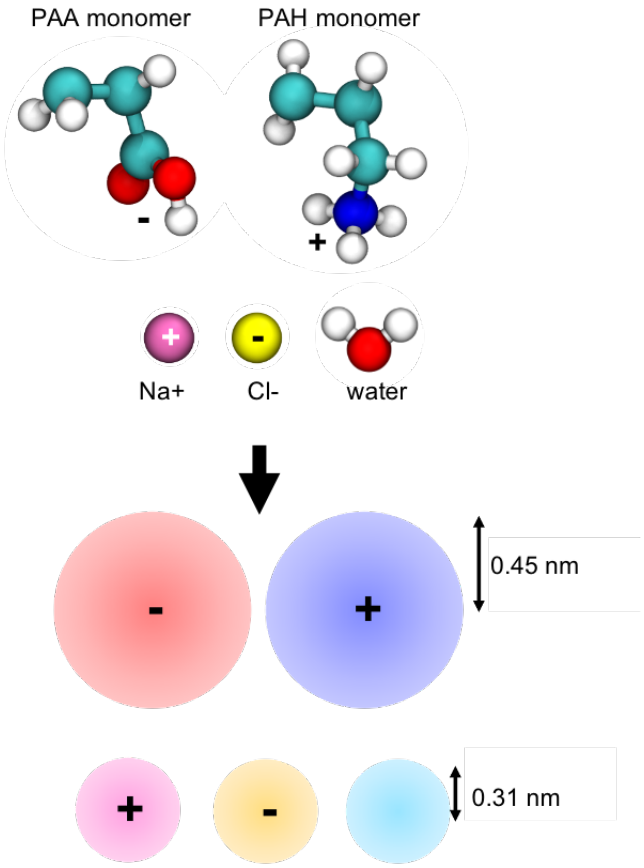

Figure 3.1: From left to right and top to bottom, the AA-to-CG mappings for PAA monomer (p_-), PAH monomer (p_+), Na^+, Cl^- and water (w). The CG bead radius is 0.45 nm for the polyelectrolyte monomers and 0.31 nm for the small molecules. See main text for details on the coarse-graining procedure.

We use three reference systems (Fig. 3.2) to parameterize the CG model: pure water, NaCl aqueous solution, and an aqueous polyelectrolyte-NaCl mixture. We use the optimal point charge (OPC) force field for water[92] and Joung-Cheatham's force field for Na^+ and Cl^- ions.[105] For PAA, we use the general Amber force field (GAFF2), while for PAH, we use the AMBER ff99 force field.[106,107] We employ the restrained electrostatic approach to assign atomic fixed-point partial charges for the polyelectrolytes at the HF6/31G* level in gas-phase calculations using the Gaussian16 software package.[108]

We find excellent agreement between aqueous solution densities from simulations and experiments for 45-mer PAA and 187-mer PAH in water at 1 atm and 298.15 K, when compared to experiments with synthetic polymers of matching average molecular weight (Fig. D.1). This agreement suggests that this set of force field parameters (OPC, Joung-Cheatham, GAFF2, and AMBER ff99) is a reasonable model for the polyelectrolyte mixture.

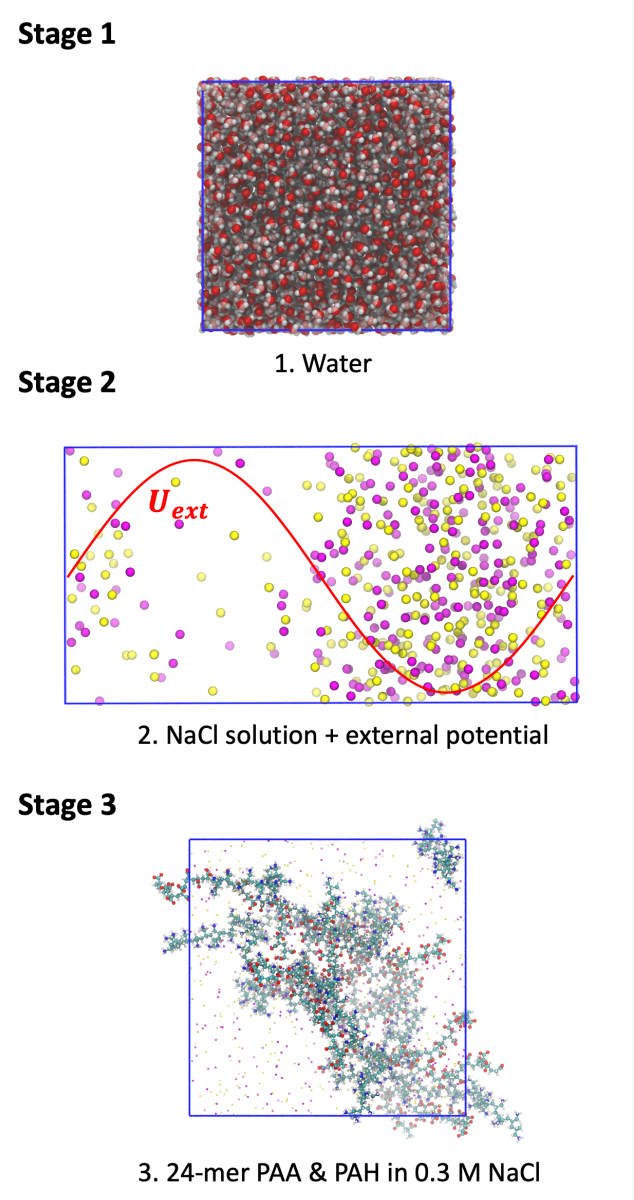

Figure 3.2: All-atom simulations used to derived coarse-grained interaction parameters for the polyelectrolyte coacervate model via the relative entropy coarse-graining framework.

The physical behavior of the AA system is coarse-grained into mesoscale models amenable to analytical conversion to a field theory, the CG force field is presented in Section 2.2.3. The field theory representation of the same coarse, particle-based model is then constructed using the Hubbard-Stratonovich-Edwards transformation as detailed in

Section 2.3. As done in our previous work,[84] we enforce specific interaction radii to retain the desired, long-length scale physics without resorting to an aggressive high-resolution AA to CG atom mapping. A convenient choice is the cube-root of the CG beads' specific volumes approximated from AA simulations of the pure component for water and single-component aqueous solutions for the other components: $a_w = a_{Na^+} = a_{Cl^-} = 0.31\,\text{nm}$, $a_{p-} = a_{p+} = 0.45\,\text{nm}$. We then use S_{rel} minimization to obtain the remaining 17 parameters in the following order: (1) $v_{w,w}$ from a pure water system in the NPT ensemble, (2) v_{w,Na^+}, v_{w,Cl^-}, v_{Na^+,Na^+}, v_{Cl^-,Cl^-}, and v_{Na^+,Cl^-} from NaCl aqueous solution in the external potential ensemble,[100] and lastly, (3) $v_{w,p-}$, $v_{w,p+}$, $v_{Na^+,p-}$, $v_{Na^+,p+}$, $v_{Cl^-,p-}$, $v_{Cl^-,p+}$, $v_{p-,p-}$, $v_{p+,p+}$, $v_{p-,p+}$, b_{p-}, and b_{p+} from the 10 weight % polyelectrolyte-NaCl mixture in the NPT ensemble. Once the parameters are determined, they are fixed in subsequent steps.

The necessary three reference AA simulations are conducted with the OpenMM simulation package.[109] We use a 1 nm cutoff for the direct space non-bonded interactions and use the particle mesh Ewald method to compute long range interactions for both Coulomb and Lennard-Jones (LJPME 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 the Monte Carlo barostat with an update frequency of $1/(25\ dt)$. Details on the system size, components, and simulation parameters are provided in Appendix D.

With the pure water reference simulation in the NPT ensemble, we parameterize the water-water interaction such that the CG model has a compressibility $\kappa_T \approx 0.062\ k_BT/a_w^3 \approx 4.51\ \times 10^{-10}\,\mathrm{Pa}^{-1}$, near that of OPC water. As discussed in our prior studies,[84] matching κ_T between AA and CG simulations uniquely determines a CG pressure of $P_{CG} \approx 8.5\ k_BT/a_w^3$.

We employ an external potential ensemble[100] to determine the self (Na^+ to Na^+ and Cl^- to Cl^-) and cross-interactions (Na^+ to Cl^-, Na^+ to water, and Cl^- to water) in a salt-water solution. For miscible mixtures, coarse-graining in the external potential ensemble shows improved thermodynamic fidelity and transferability of CG models over the widely used single, uniform-composition ensembles. As detailed in our previous work,[100] the coarse-graining is performed in a state of inhomogeneous response in the composition due to an applied spatially-varying external potential on different species. The resulting CG model is then optimized such that its interactions reproduce the same response as that by the AA reference system while capturing the dependence on the locally varying composition. Here, we observe improved NaCl mixing thermodynamics in water, evidenced by the mean ionic activity in Fig. B.3, when we derive the CG model in the presence of the sinusoidal external potential (visualized in Fig. B.1). We choose the optimal external potential ensemble to optimize the CG parameters by maximizing the trace of the Fisher information matrix (the Hessian of S_{rel} with respect to the interaction parameters); we find the maximum in the Fisher information matrix for external sinusoidal potentials to lie near an amplitude of 2 k_BT, Fig. B.4.

The last reference simulation is a solution of 10 weight % polyelectrolytes (50:50 mol/mol PAA to PAH; 24-mers, fully-ionized) and 0.3 M NaCl in the NPT ensemble. We model the polyelectrolytes as atactic polymers; they are built with the target dyad composition of around 0.44 meso and 0.56 racemic as per the Bernoullian distribution, Fig. D.2.[110,111] We note that the CG force field is composed of soft interaction potentials that do not account for the bending stiffness that partially dictates the chain conformation in the reference model. Chain conformation embeds information about the bending rigidity, intramolecule interactions, volume exclusion effects, and solvent screening; thus, it is essential for the CG model to reproduce this characteristic of the reference system. We reduce this complex interplay to a simple quantity - the average radius of gyration, R_g, and require the CG model to reproduce the average R_g of PAA and PAH in the reference AA simulation. This constraint is enforced during the S_{rel} minimization by modifying the objective function according to an augmented Lagrangian method:[112–114]

$$F_{obj}(\boldsymbol{\lambda}) = S_{rel}(\boldsymbol{\lambda}) - \zeta \left(\langle R_g \rangle_{CG}(\boldsymbol{\lambda}) - \langle R_g \rangle_{AA} \right) + \frac{c}{2} \left(\langle R_g \rangle_{CG}(\boldsymbol{\lambda}) - \langle R_g \rangle_{AA} \right)^2 \tag{3.1}$$

where $\boldsymbol{\lambda}$ are the CG force field parameters of interest, and the coefficient c, typically $\sim \mathcal{O}(10^{10})$, is chosen to bias the CG model's average radius of gyration, $\langle R_g \rangle_{CG}$, towards that of the mapped AA system,$\langle R_g \rangle_{AA}$. The Lagrange multiplier at iteration k is defined as $\zeta_k = \zeta_{k-1} - c \left(\langle R_g \rangle_{CG,k-1} - \langle R_g \rangle_{AA} \right)$. In contrast to neutral polymers, polyelectrolytes often have long persistence lengths ($\sim$ 10 nm as suggested by our simulations and other studies)[115,116] due to the repulsion of backbone charges. Hence, the mismatch in the chain conformation between the CG and AA systems will be significant without considering

chain stiffness in the coarse-graining, especially for stiff polymers. We tabulate all force field parameters for the CG model and provide them in Tables D.1 and D.2.

3.2.2 Multi-component phase boundary calculations

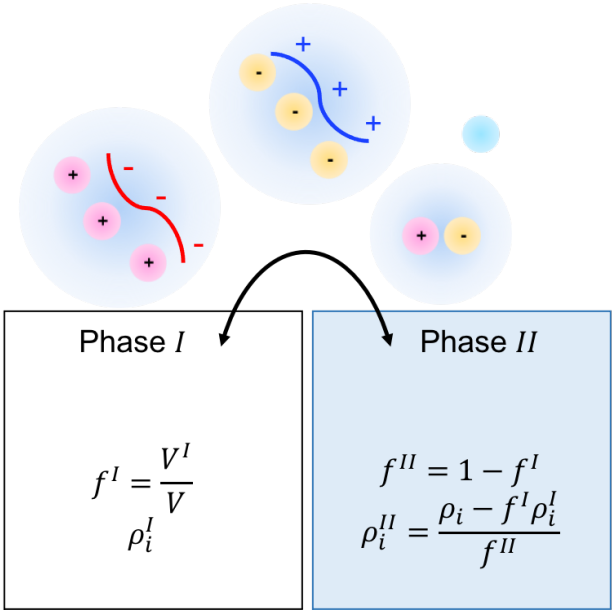

Figure 3.3: Schematic of the Gibbs ensemble. Coexisting dilute and coacervate phases are partitioned into separate simulation boxes. Neutral pairs of charged molecules ($p-/Na^+$, $p+/Cl^-$ and Na^+/Cl^-) and water are exchanged to achieve electrochemical equilibrium, and phase volumes are exchanged to achieve mechanical equilibrium.

We determine the multi-component phase equilibrium conditions within field theory in the Gibbs ensemble where the coexisting phases are partitioned into separate simulation boxes as illustrated in Fig. 3.3.[117] The overall Gibbs ensemble consists of two subsystems; each is treated in the canonical ensemble, with the distribution of mass and volume in each subsystem being constrained by the overall species densities. At phase coexistence, the overall free energy, F, is minimized, and the system is in mechanical and electrochemical

equilibrium,

$$\frac{\partial F}{\partial V^I} = -(P^I - P^{II}) = 0 \tag{3.2}$$

$$\frac{\partial F}{\partial n_m^I} = \mu_m^I - \mu_m^{II} + \sigma_m \Delta\Psi = 0 \tag{3.3}$$

where V is the simulation box volume, P is the pressure, n_m is the number of molecules m and μ_m is its chemical potential, and the superscript denotes phase I or II. The last term in Eq. 3.3 accounts for the electrostatic potential difference between the two phases and depends on the charge of molecule m, σ_m, and the Galvani potential, $\Delta\Psi$. Instead of exchanging individual molecular species, we exchange neutral pairs of charged molecules such that the equilibrium condition of Eq. 3.3 becomes

$$\frac{\partial F}{\partial n_i^I} = \mu_i^I - \mu_i^{II} = 0 \tag{3.4}$$

where i denotes the following neutral pairs: $p-/\mathrm{Na}^+$, $p+/\mathrm{Cl}^-$ and $\mathrm{Na}^+/\mathrm{Cl}^-$, in addition to water. We note that $n-1$ linearly independent neutral pairs are sufficient to span all possible values of $n-1$ independent densities in a system of n charged species. We define the effective chemical potentials of the neutral pairs from the chemical potentials of the charged molecules as

$$\mu_{p-/Na^+} = \mu_{p-} + N_{p-}\ \mu_{Na^+} \tag{3.5}$$

$$\mu_{p+/Cl^-} = \mu_{p+} + N_{p+}\ \mu_{Cl^-} \tag{3.6}$$

$$\mu_{Na^+/Cl^-} = \mu_{Na^+} + \mu_{Cl^-} \tag{3.7}$$

where N_{p-} and N_{p+} are the degrees of polymerization of PAA and PAH, respectively, and μ_i are the *per molecule* chemical potentials. The mass balance gives the expression for the overall monomer density of pair i

$$\rho_i = f^I \rho_i^I + f^{II} \rho_i^{II} \tag{3.8}$$

where the volume fractions of phase I is $f^I = V^I/V$ and phase II is $f^{II} = 1 - f^I$. Equilibrium is achieved by performing mass and volume swaps between the simulation boxes formulated as the following differential equations in a fictitious time t

$$\begin{aligned} \frac{df^I}{dt} &= P^I - P^{II} \\ \frac{d\rho_i^I}{dt} &= -(\mu_i^I - \mu_i^{II}) \end{aligned} \tag{3.9}$$

In practice, we conduct a series of iterative simulations. In each iteration, we perform two separate canonical simulations that represent phases I and II. We update the volume fractions and species' densities with the following scheme until the equilibrium conditions, Eq. (3.2),(3.4), are satisfied:

$$\rho_{tot}(k+1) = \rho_{tot}(k) - \Delta t_{\rho_{tot}}(P^I - P_{CG}) \tag{3.10}$$

$$f^I(k+1) = f^I(k) + \Delta t_f \left(P^I(k) - P^{II}(k)\right) \tag{3.11}$$

$$\rho_i^I(k+1) = \rho_i^I(k) - \Delta t_{\rho_i} \left(\mu_i^I(k) - \mu_i^{II}(k)\right) \tag{3.12}$$

$$f^{II}(k+1) = 1 - f^I(k+1) \tag{3.13}$$

$$\rho_i^{II}(k+1) = \frac{\rho_i - f^I(k+1)\rho_i^I(k+1)}{f^{II}(k+1)} \tag{3.14}$$

where k is the discrete time index, Δt_f, $\Delta t_{\rho_{tot}}$, and Δt_{ρ_i} are step sizes for volume fraction, overall density, and species density swaps, respectively. Eq. 3.10 updates the overall bead density, $\rho_{tot} = \sum_{m\in[p-,p+,Na^+,Cl^-,w]} \rho_m$, at fixed overall species fractions such that the coexisting phases will be at the pressure of the CG model, P_{CG}. This condition emulates the experimental conditions at constant temperature and pressure. We find that $\Delta t_{\rho_{tot}} = 0.002$, $\Delta t_f = 0.01$ and $\Delta t_{\rho_i} = 0.1 \min(\rho_i^I, \rho_i^{II})$ work well. From our definitions of neutral pairs, the bead density for each charged species follows as

$$\rho_{p-} = \rho_{p-/Na^+} \tag{3.15}$$

$$\rho_{p+} = \rho_{p+/Cl^-} \tag{3.16}$$

$$\rho_{Na^+} = \rho_{p-/Na^+} + \rho_{Na^+/Cl^-} \tag{3.17}$$

$$\rho_{Cl^-} = \rho_{p+/Cl^-} + \rho_{Na^+/Cl^-} \tag{3.18}$$

To rapidly screen phase behavior, we use a Gaussian approximation to evaluate the pressure and chemical potentials used in field-theoretic Gibbs ensemble simulations. Previous efforts have shown that the Gaussian approximation, while it only includes field fluctuations up to the second order, is semi-quantitative in reproducing the dense branch of the coacervate phase diagram.[27,65,67] Specifically in our work, we add Gaussian fluctuation modes of the electrostatic interactions on top of a mean-field treatment of excluded-volume interactions. The derivation of analytical expressions for the requisite thermodynamic quantities are provided Appendix A.

3.3 Results and discussion

3.3.1 Polyelectrolyte mixture with no added salt

Prior experiments have suggested that the propensity for coacervation decreases with the degree of non-stoichiometry in charged monomer compositions.[20,101] We probe the effect of charged monomer stoichiometry in a salt-free mixture consisting of PAA, PAH, the appropriate ion needed to neutralize the system, and water. The amount of PAA relative to PAH in the mixture is controlled by the PAA fraction parameter defined as $f_{p-} = \rho_{p-}/\rho_p$ where $\rho_p = \rho_{p-} + \rho_{p+}$ is the total bead number density of the polyelectrolytes. The bead density of PAH is thus $\rho_{p+} = (1 - f_{p-})\rho_{p-}/f_{p-}$. When $f_{p-} > 0.5 (< 0.5)$, PAA (PAH) is in excess and we include Na^+ (Cl^-) to neutralize the system.

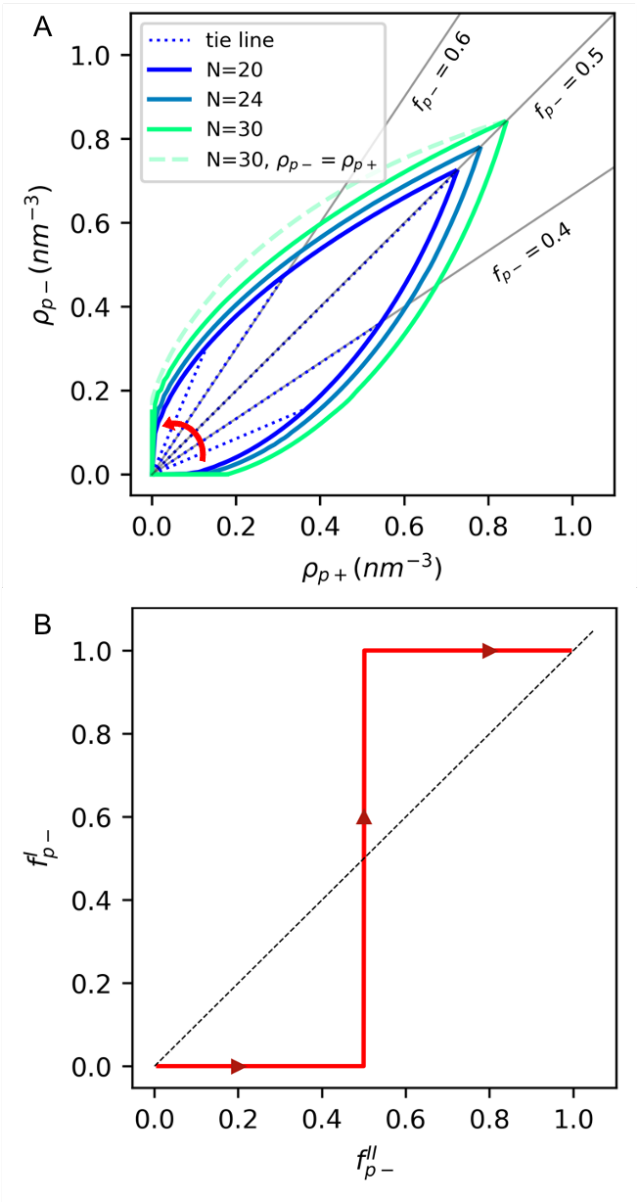

Figure 3.4: (A) Phase diagram for the salt-free condition of a PAA/PAH mixture at varying chain length, N. The solid lines denote the binodals, and dotted lines are example tie lines. Gray solid lines represent compositions corresponding to specific f_{p-} values, as annotated on the figure. The faded dashed line is the hypothetical binodal for excess PAA conditions at $N = 30$ if the phase diagram is symmetric ($\rho_{p-} = \rho_{p+}$). The red arrow shows a path traversing from a PAH-rich mixture ($f_{p-} \to 0$) to a PAA-rich mixture ($f_{p-} \to 1$) for $N = 20$. (B) Plot of the monomer fraction of PAA in the coexisting dilute, f^{I}_{p-}, and coacervate phases, f^{II}_{p-}, along this path at $N = 20$.

Fig. 3.4A shows the binodals for the polyelectrolyte chain lengths $N = 20$, 24 and 30.

The tie line, connecting the compositions in the dilute and coacervate phases, shortens with increasing deviation from charged monomer stoichiometry. The stoichiometric condition is denoted by the diagonal line, whereas other stoichiometries are gray lines above ($f_{p-} > 0.5$) or below ($f_{p-} < 0.5$) the diagonal. The increased propensity for coacervation with increasing N is reflected by both the longer tie lines (evident at $f_{p-} = 0.5$) and larger range of composition enclosed by the binodal.

For values of N shown here, the alignment of tie lines and the stoichiometry lines suggest that the coacervate retains the overall stoichiometry. This also indicates that most of the excess polyelectrolytes go into the coacervate phase (with the excess charges being neutralized by the counterions), while the dilute phase is almost devoid of both species of polyelectrolyte. This observation of the coacervate's stoichiometry relative to the overall mixture's is qualitatively different from previous theoretical work by Zhang *et al.*[79] In their work, the tie lines are almost parallel to the diagonal line, suggesting that the polyelectrolytes re-distribute in the coexisting phases such that the coacervate phase composition is less asymmetric in terms of the number of oppositely charged monomers as compared to that of the overall mixture. These differences could be the results of the enhanced charged connectivity in the Gaussian approximation of the field theory and chemical specific interactions that we employ here.

For the dilute phase, it is more illustrative to show the trajectory of the stoichiometry in the coexisting dilute phase (phase I) and coacervate phase (phase II), Fig. 3.4B, as we follow the path traversing across tie lines starting from $f_{p-} \to 0$ to $f_{p-} \to 1$ (red arrow

in Fig. 3.4A). Although the PAA fraction in the coacervate phase varies continuously from 0 to 1, this value is either near 0 for $f_{p-} < 0.5$ or near 1 for $f_{p-} > 0.5$ in the dilute phase, suggesting that the dilute phase, while being depleted of polyelectrolytes, is composed of mostly the polyelectrolyte species in excess. It is also inferred from Fig. 3.4B that the coacervate is stabilized for all range of stoichiometry, even at $f_{p-} \to 0$ and 1. We hypothesize that the stability of the coacervate over the full range of stoichiometry is due to the parameterization of the CG polyelectrolyte models at the phase separated state (at $f_{p-} = 0.5$ and 0.3 M NaCl). Interaction parameters derived from this reference state are biased toward the collapsed configurations of PAA and PAH, which promote phase separation even when only one polyelectrolyte species is present. This is related to the transferability of the polyelectrolyte model across stoichiometries which can be improved by employing the external potential ensemble in the parameterization step of the polyelectrolyte-NaCl mixture.

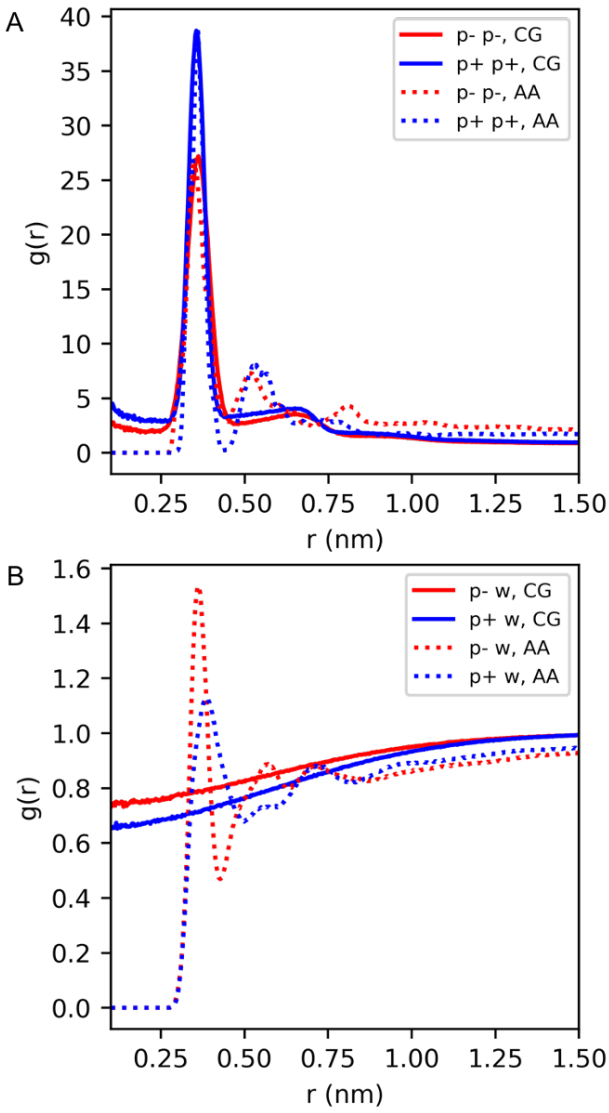

Figure 3.5: RDFs between the center-of-mass of the polyelectrolyte monomers and (A) that of the same polyelectrolyte species and (B) water in the CG simulation (solid lines) and the AA simulation (dotted lines) calculated from a mixture corresponding to simulation 3 in Fig. 3.2. Red and blue lines correspond to PAA and PAH monomers, respectively. The CG models do not retain liquid structuring of the AA model due to our choice of large interaction radii that only resolves long length-scale physics.

Lastly, we want to note that the phase boundaries are not symmetric across the diagonal; comparing actual phase boundaries to hypothetical boundaries by assuming

$\rho_{p+}^{I,II} = \rho_{p-}^{I,II}$ clearly shows the asymmetry (faded dashed lines in Fig. 3.4A and Fig. D.3). The coacervate dense phase incorporates slightly more PAH in non-stoichiometric mixtures with excess PAH (e.g., $f_{p-} = 0.3$), as compared to the incorporation of PAA when it is in excess by the same amount (e.g., $f_{p-} = 0.7$). The tendency of PAH to form a denser coacervate is explained by the radial distribution functions (RDFs) in a mixture of 10 wt % PAA and PAH at $f_{p-} = 0.5$ and 0.3 M NaCl (the composition of the third reference AA system discussed earlier). The analysis of the CG model shows that the PAH monomer-PAH monomer RDF has higher intensity than the PAA monomer-PAA monomer RDF (Fig. G.5A) while the PAH monomer-water RDF shows less pronounced structuring than the PAA monomer-water RDF (Fig. G.5B). This suggests that PAH is less soluble in water; the same trend is observed in the AA model (dotted lines). This asymmetry arises from the chemistry embedded in our model that leads to different interaction parameters among the charged moieties.

3.3.2 Polyelectrolyte mixture with added salt

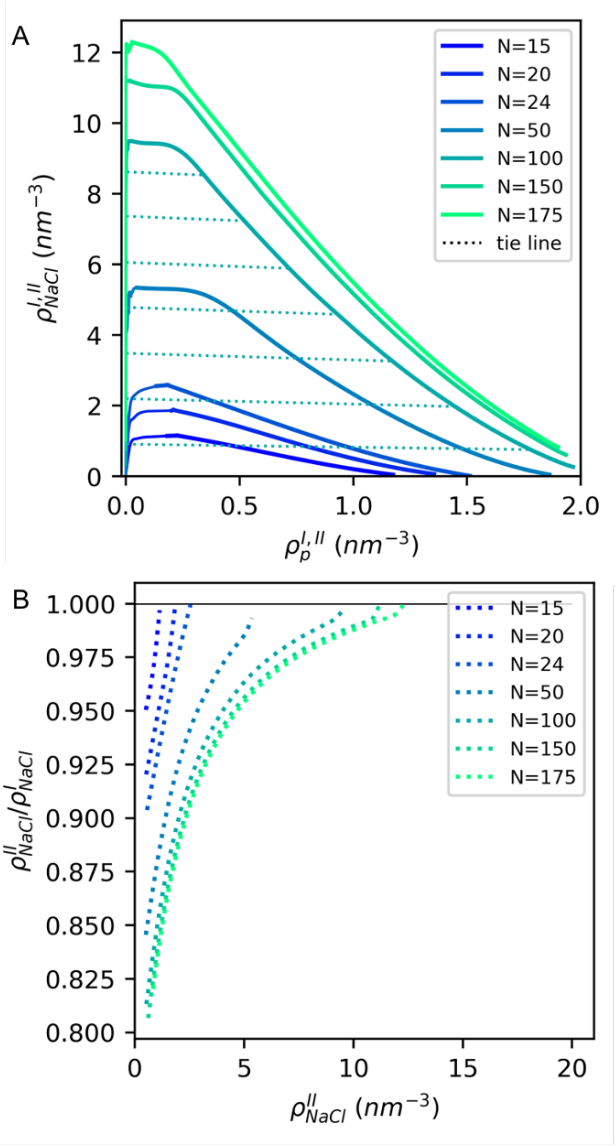

Figure 3.6: (A) Binodals at $f_{p-} = 0.5$ for varying N. Dotted lines are tie lines for $N = 100$. The negative slope in the tie lines suggests that there is slightly more salt in the dilute phase than in the coacervate phase. This is more evident in (B), a plot of the ratio of salt concentrations in the coacervate phase and the dilute phase as a function of the excess salt concentration in the coacervate. The ratio is always below unity for any values of the added salt concentration and N.

We now consider the addition of salt in a polyelectrolyte mixture where the ions stem from both the counterions (one counterion per polyelectrolyte charge) and added salt, i.e., $\rho_{Na^+} = \rho_{p-} + \rho_{salt}$ and $\rho_{Cl^-} = \rho_{p-} + \rho_{salt}$. Since the small ions have different partitioning behavior in the coexisting phases, we report their concentration in each phase as the smaller value of the Na^+ and Cl^- concentrations, i.e., $\rho^{I,II}_{NaCl} = \min\left(\rho^{I,II}_{Na^+}, \rho^{I,II}_{Cl^-}\right)$; this is equivalent to assessing the excess salt in each phase.

Fig. 3.6A shows how the dense branch concentration of the two-phase coexistence region for the stoichiometric mixture reduces with the salt concentration. The dilute branch concentration, on the other hand, becomes denser with added salt (Fig D.4) as the binodal region shrinks from both sides. Increasing the salt concentration reduces the effective electrostatic strength due to screening from the increased charge density. Both the entropic and enthalpic driving forces for coacervation diminish with increasing ionic strength before coacervation is thermodynamically unfavorable, at which point the complexation free energy becomes 0 (Fig. D.5). The salt concentration at this transition, ρ^*_{salt}, characterizes the salt resistance of the coacervates, i.e., the minimum amount of salt to dissolve the coacervate phase. The two-phase region also has a lower bound at $N \approx 15$ and increases with molecular weight until approximately saturating above $N \approx 150$. The higher solubility of shorter polyelectrolytes is not surprising and is a result of their increased translational entropy and weaker electrostatic correlation energy.[65]

The negatively sloped tie line suggests that the excess salt concentration in the dilute phase is slightly higher than that in the coexisting coacervate phase. This is readily

apparent from the ratio of the salt concentration in the coacervate phase with respect to that in the dilute phase in Fig. 3.6B, which is less than unity for all salt concentrations; observations that are consistent for the individual anion and cation species are shown in Fig. D.6. As salt is increased at constant N, or as N is decreased at constant ρ^{II}_{NaCl}, the polyelectrolyte concentration in the coacervate phase decreases. This reduces the excluded volume asymmetry between the two phases, and, as a result, the salt partitioning ratio increases from values less than 1 toward 1. While experimental efforts have shown that the salt partitioning behavior is not universal and depends on other factors such as the chemistry of the polyelectrolyte and charge density,[118,119] our prediction of higher salt partitioning in the dilute phase is consistent with the experimental observations for the stoichiometric PAA/PAH mixture.[1,2]

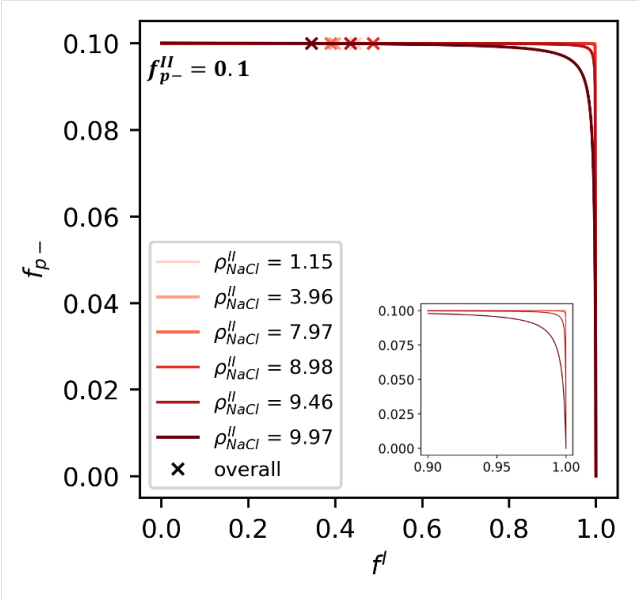

Figure 3.7: Overall polyelectrolyte stoichiometry f_{p-} along a tie line by varying the dilute phase volume fraction f^I for the case of the coacervate stoichiometry of 0.1. Different series represent tie lines corresponding to different excess salt concentrations in the coacervate (in nm^{-3}). Cross symbols denote the overall compositions used in Gibbs ensemble calculations to obtain the coexisting phases. Inset: expanded region near the dilute phase to highlight the continuous transition of the stoichiometry from 0.1 to 0 at moderate to high salt concentrations.

Using the same protocol, we construct phase diagrams for $N = 150$ under non-stoichiometric conditions. Although in general the stoichiometry of the coacervate phase can be different from that of the initial mixtures and depends on the overall composition, we find that the coacervate maintains the overall stoichiometry, and this observation is mostly independent of the overall composition. As a demonstration, for a pair of the coexisting phases we calculate the overall stoichiometry of different mixtures with overall compositions at different points along the tie line by varying the dilute phase volume fraction f^I, which controls the relative proportion of the two phases. The overall stoichiometry along the tie line for the coacervate stoichiometry $f^{II}_{p-} = 0.1$ is presented in Fig. 3.7 where different series represent tie lines at different salt concentrations.

The overall stoichiometry, f_{p-}, approaches the coacervate (dilute phase) stoichiometry as the dilute phase volume fraction approaches 0 (1). At low salt concentrations, the stoichiometry along the tie line is the same as that of the coacervate phase and has a discontinuity in the dilute phase ($f^I = 1$). At moderate to high salt concentrations, the stoichiometry varies continuously from 0.1 to 0 as we approach the dilute phase. This transitional region is indeed quite narrow and appears very close to the dilute phase ($f^I \gtrsim 0.8$) for all stoichiometries we investigated (Fig. D.7), suggesting most compositions enclosed by the binodals of Fig. 3.8A have the same stoichiometry as that of the corresponding coacervate phase. This is expected because the dilute phase has very low concentrations of polyelectrolyte and contributes very little to the overall polymer mass. Going forward, we refer to "non-stoichiometric" mixtures as those with the coacervate phase at non-stoichiometric conditions.

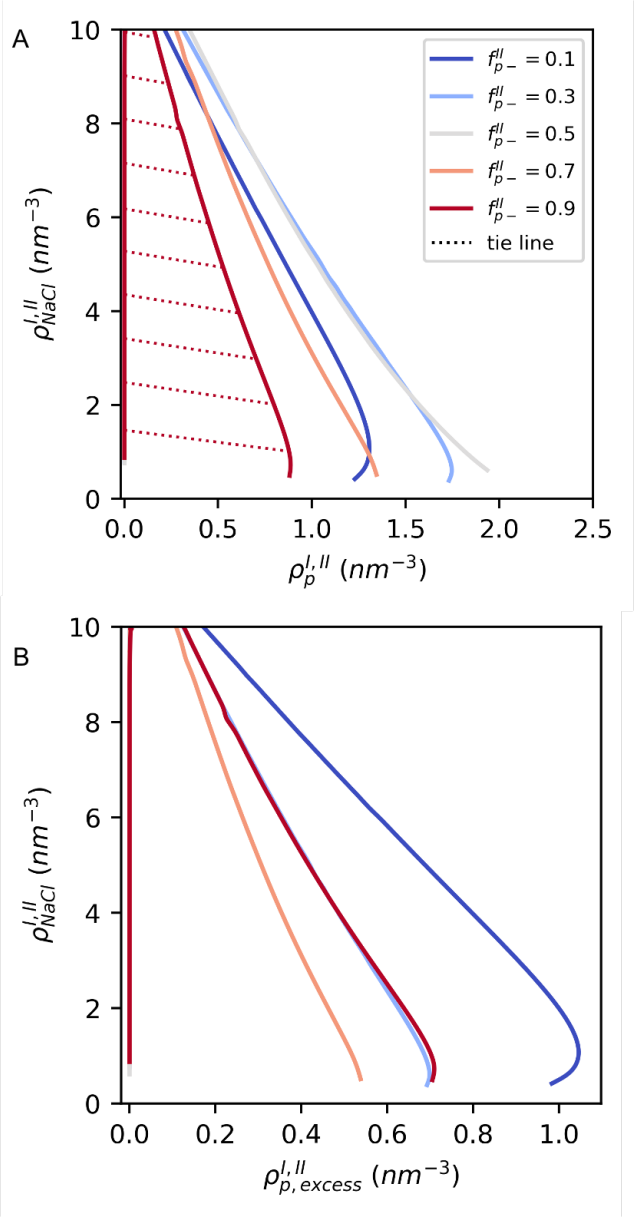

Figure 3.8: (A) Binodals for $N = 150$ at stoichiometries $f^{II}_{p-} = 0.1, 0.3, 0.5, 0.7, 0.9$ with dotted lines denoting tie lines. The reduced number of potential ion pairs in non-stoichiometric mixtures leads to the shrinkage of the coacervation region. (B) Same salt concentration vs the excess polyelectrolyte concentration $\rho_{p,excess} = |\rho_{p-} - \rho_{p^+}|$ in the coexisting phases. $\rho_{p,excess}$ is essentially 0 for the stoichiometric mixture.

Notably, the coacervation regions for non-stoichiometric mixtures are smaller than those for stoichiometric mixtures as evident in Fig. 3.8A. At the same overall polyelec-

trolyte concentration ρ_p, more asymmetric mixtures have fewer potential ion pairs. Thus, less salt is needed to screen the electrostatic interactions and dissolve the coacervate as compared to that under the stoichiometric condition. Unlike those in the stoichiometric mixture, cations and anions have different partitioning behavior in non-stoichiometric mixtures (Fig. 3.9) due to the electroneutrality constraint: excess-PAA (PAH) coacervate require excess Na^+ (Cl^-) in the same proportion to neutralize the charge. For $f_{p-}^{II} = 0.5$, the coacervate phase has equimolar amounts of oppositely charged monomers such that the small ions have freedom to partition in the dilute phase where it is less crowded. We show in Fig. 3.10 the partitioning ratio of "free ions" that do not neutralize the excess polyelectrolytes and have concentration $\rho_{free\ ion}^{I,II} = \rho_{Na^+}^{I,II} + \rho_{Cl^-}^{I,II} - \rho_{p,excess}^{I,II}$. This characterizes the degree of ion partitioning after accounting for the partitioning due to neutralization of the excess polyelectrolytes in both phases. As expected, when the polyelectrolyte complex is sufficiently neutralized, small ions preferentially partition in the dilute phase.

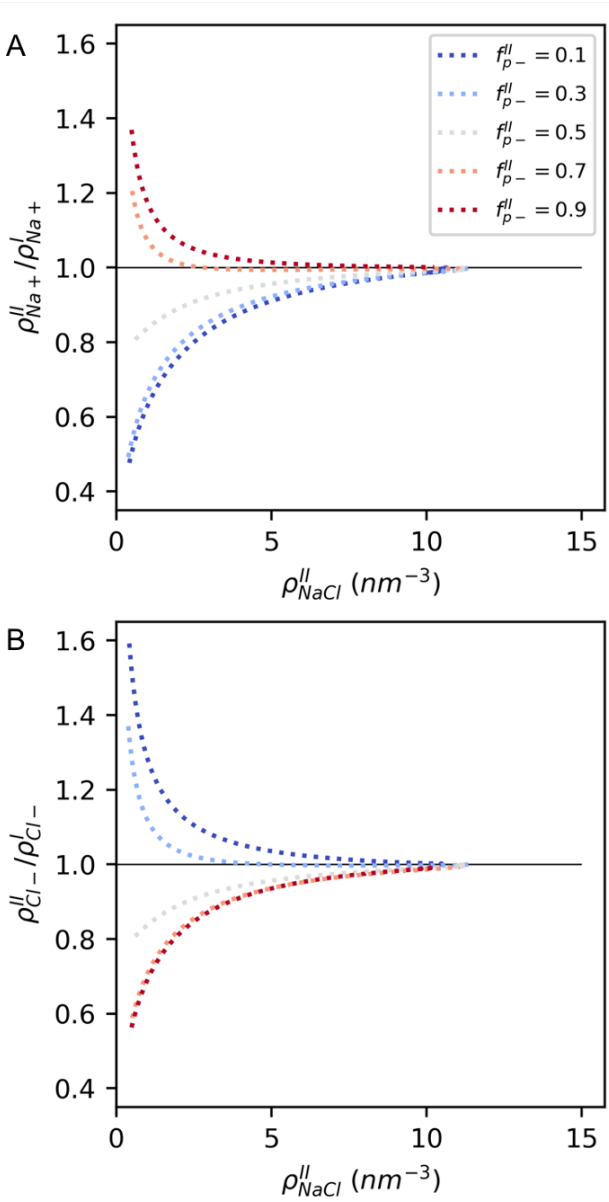

Figure 3.9: Partitioning ratios of (A) Na^+ and (B) Cl^- across excess salt concentrations in the coacervate for different f^{II}_{p-} at $N = 150$.

Interestingly, under non-stoichiometric conditions, the dense branch shows a looping-in shape that has been observed experimentally in other non-stoichiometric polyelectrolyte systems.[118] This looping-in feature is representative of a salting-out phenomenon,

where a homogeneous solution undergoes phase separation upon an initial addition of salt.[120] Friedowitz *et al.* attribute the looping-in feature in their phase diagrams to the competition of the counterion mixing entropy and the electroneutrality constraint in the coacervate when there is charge asymmetry. They argue that when excess polyelectrolytes preferentially partition into the coacervate, as is observed in our system, salt ions that do not participate in neutralization of the complex preferentially populate the dilute phase to maximize their translational entropy. Consequently, the polyelectrolyte concentration in the coacervate must increase in order to maintain the osmotic pressure balance between the two phases. With sufficient added salt, the salt partitions more evenly across the two phases, and eventually, the screening effect dominates, leading to a decrease in the coacervate concentration with increasing salt. Fig. 3.10 shows a drastic decrease in the free ion partitioning ratio for non-stoichiometric mixtures as compared to that under the stoichiometric condition, especially at low added salt concentrations, which is consistent with the earlier argument of the increased accumulation of free ions in the dilute phase. Furthermore, the salt concentration where the free ion partitioning ratio crosses ~ 0.8 (the partitioning of the $f_{p-} = 0.5$ mixture at low salt concentration) is qualitatively where the looping-in feature ends, suggesting that the looping-in shape in the dense branch is indeed related to the partitioning behavior of small ions.

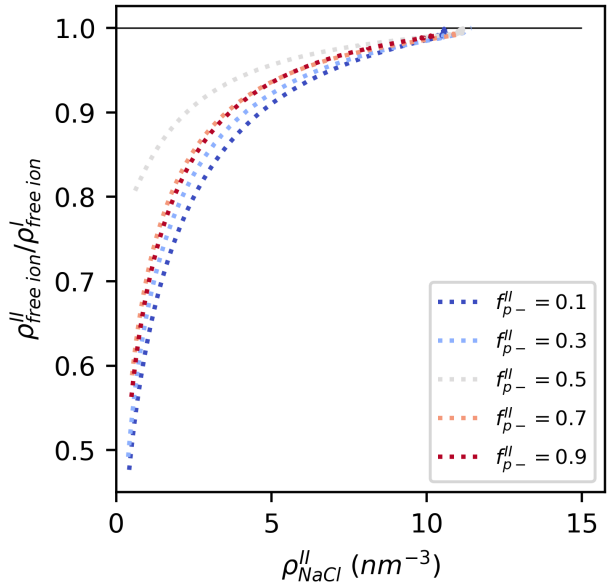

Figure 3.10: Partitioning ratios of the free ions, $\rho^{I,II}_{free\ ion} = \rho^{I,II}_{Na^+} + \rho^{I,II}_{Cl^-} - \rho^{I,II}_{p,excess}$, across excess salt concentrations for different f^{II}_{p-} at N=150. At low added salt concentrations, this ratio is significantly lower in non-stoichiometric mixtures as compared to that in stoichiometric mixtures.

Similar to that under the salt-free condition, we note that the change in the coacervation region is not symmetric between the excess PAA ($f^{II}_{p-} > 0.5$) and excess PAH ($f^{II}_{p-} < 0.5$) conditions. Specifically, the two-phase window for excess PAH is larger than that for the excess PAA case at the same degree of non-stoichiometry (e.g. $f^{II}_{p-} = 0.1$ vs 0.9, 0.3 vs 0.7, etc.). In addition, the polyelectrolyte complex carries more charge, characterized by the excess polyelectrolyte concentration $\rho^{II}_{p,excess} = |\rho^{II}_{p-} - \rho^{II}_{p^+}|$, in the case of excess PAH conditions relative to excess PAA conditions at the same ionic strength (Fig. 3.8B); thus, more counterions (Cl^- for $f^{II}_{p-} < 0.5$, Na^+ for $f^{II}_{p-} > 0.5$) are needed to neutralize coacervates with excess PAH (Fig. 3.9). Lastly, it is inferred that the propensity for coacervation at a fixed degree of non-stoichiometry is higher when the excess species

is PAH than when it is PAA. A similar observation is suggested by the experimental ternary phase diagrams at moderate salt concentrations by Chollakup *et al.* where the critical stoichiometry, beyond which coacervation does not occur, appears to be biased toward mixtures with a higher degree of excess PAH.[20] The fact that we can predict this asymmetry provides some validation of the workflow's ability to preserve the chemical specificity of the AA model in the CG model.

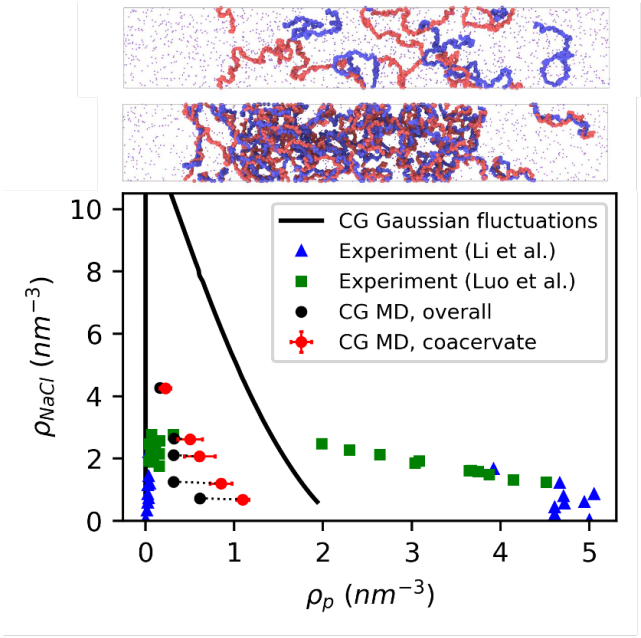

Figure 3.11: Phase diagrams for stoichiometric mixutures at $N = 150$ from the Gaussian approximation (black line) and CG MD (circles) compared to the experimental data by Li *et al.* (triangles) and Luo *et al.* (squares).[1,2] MD simulations are conducted in the NPT ensemble at 298.15 K and P_{CG} in a rectangular box of dimensions $\sim 11 \times 11 \times 80$ nm^3. Red circles reflect the polyelectrolyte and salt compositions in the coacervate, while black circles are the overall compositions. The polyelectrolyte composition in the dense branch from CG MD approaches the bulk with the increasing salt concentration much faster than the Gaussian approximation. The two simulation snapshots from CG MD, showing PAA (red chains) and PAH (blue chains), correspond to the highest (top) and lowest (bottom) salt concentrations shown here; water molecules are not shown for clarity.

The phase diagram from the Gaussian approximation for $N = 150$ at 1:1 stoichiometry

is directly compared to experimental results of the similar system reported by Li *et al.* and Luo *et al.* at pH = 6.5 (Fig. 3.11).[1,2] While qualitatively capturing the screening effect of added salt, the phase diagram shown here: 1) overestimates the critical salt concentration; and, 2) underestimates the polyelectrolyte compositions in the coacervate phase. Deficiency 1) is largely attributable to the Gaussian approximation which neglects higher-order fluctuations that become more important near the critical point (can be resolved by including fields other than the electrostatic potential). From a limited number of particle-based MD simulations of coacervates in equilibrium with a supernatant phase (circles in Fig. 3.11), we show that with full incorporation of fluctuations, the CG model indeed has a comparable salt resistance with experiments and becomes homogeneous above $\rho_{NaCl} \approx 4.5\ \text{nm}^{-3}$.

In contrast, the underestimation of the coacervate phase polymer density is attributed to the assumption of a constant dielectric screening effect of the solvent with added salt and polyelectrolyte. The current CG model assumes the Bjerrum length of pure water throughout the coexisting regions while in reality the Bjerrum length is likely larger in the dense coacervate (corresponding to a smaller dielectric constant). With an increased Bjerrum length, the attractive interactions between polyelectrolyte monomers will be stronger, resulting in more concentrated coacervate branch. A more sophisticated model that better describes the actual electrostatic environment would be one with a dielectrically active solvent. Specifically, one can model the solvent as a polarizable or fixed-dipole solvent such that the screening depends on the local environment;[121] we leave

investigating such a model to future work.

A potential further refinement to the workflow would be to coarse-grain the polyelectrolyte-polyelectrolyte, polyelectrolyte-ion and polyelectrolyte-water interactions in the external potential ensemble. Deriving the CG parameters in this ensemble improves thermodynamic faithfulness and transferability across state points (e.g., salt concentrations, polyelectrolyte concentrations and stoichiometries) by applying spatially varying external potentials to one or more species to force local composition variations and fluctuations.[100] From the profound improvement observed in the thermodynamics of the NaCl model across concentrations (Fig. B.3), we expect that this strategy will also improve the quality of the polyelectrolyte-NaCl model. However, probing fluctuations in both polyelectrolyte and salt compositions will require multiple external potentials to be applied on different components, the design of which is not obvious. Thus, further investigation is necessary to find an optimal reference ensemble for coarse-graining the polyelectrolyte-NaCl system.

3.4 Conclusions

This work presents phase diagrams for a PAA/PAH complex coacervate system from a molecularly-informed field theory that captures not only the effects of charged group stoichiometry, electrostatic screening with added salt, and salt partitioning behavior, but also chemistry-specific effects due to the molecular details of the polyelectrolytes (i.e., PAH-PAA asymmetries in non-stoichiometric mixtures). At the same overall polyelec-

trolyte concentration, the coacervation propensity decreases as the PAA ratio, $f_{p-} = \rho_{p-}/(\rho_{p-} + \rho_{p+})$, deviates from the stoichiometric condition ($f_{p-} = 0.5$). By reducing the number of potential ion pairs, the coacervate phase is more dilute, evidenced by shorter tie lines (Fig 3.4), and less salt is needed to dissolve the coacervate, resulting in a smaller two-phase region (Fig. 3.8A). The two-phase region shrinks at higher salinity (Fig. 3.6A and 3.8A) and the model predicts that small ions favor the dilute phase for stoichiometric mixtures (Fig. 3.6B). For non-stoichiometric mixtures, however, cations and anions have different partitioning patterns due to the electroneutrality constraint such that Na^+ (Cl^-) ions populate the coacervate phase when PAA (PAH) is in excess (Fig. 3.9). The looping-in behavior in non-stoichiometric mixtures at low salt concentrations, also observed previously in experiments,[118] is attributed to the competition between maximizing the translational entropy of the free ions (ions that do not neutralize the excess charge in the coacervate) and the electroneutrality constraint in the coacervate. While the simulations do not produce quantitative agreement with the experiments, they nevertheless capture qualitative trends with no fit parameters, providing an important screening tool. Nonetheless, with simple potential functional forms and assumption of a constant electrostatic screening environment throughout the phase separating regimes (constant Bjerrum length), the model qualitatively captures the coacervation response to the added salt concentration and mixing stoichiometries.

Atomic-scale details are extremely important in formulation and material design involving the fine-tuning of component chemistries. We believe that molecularly-informed

field theories are a promising way to overcome long-standing challenges in studying coacervation physics with either (1) traditional particle-explicit models that struggle to sample meaningful conformations or (2) phenomenological field theories that lack chemical specificity. For example, it is often assumed in theoretical models that the polyanion and polycation have the same chemical structure except for opposite charges, resulting in often nonphysical symmetric phase diagrams with respect to the two polyelectrolyte species. We have demonstrated a bottom-up coarse-graining methodology that molecularly informs the field-theoretic model with chemical details from atomistic simulations. A notable outcome is the prediction of a non-symmetric diagram where coacervation is more favorable when the polyelectrolyte in excess is PAH. We believe that the present coacervate model of PAA/PAH represents a significant step forward over prior simulation and theoretical studies in that it retains chemical specificity of the involved components while still making use of a computationally efficient field representation capable of rigorous phase diagram calculations.

Acknowledgements This work was supported by BASF Corporation through the California Research Alliance. K.T.D. and G.H.F. derived partial support from the CMMT Program of the National Science Foundation, Award DMR-2104255. 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.

C.E.R.E and M.E.H. were supported by the MRSEC Program of the National Science Foundation under Award No. DMR 1720256 (IRG-3).

Supporting Information see Appendix D

Chapter 4

Complexation between Polyelectrolyte with Oppositely Charged Micelles

Reproduced in part with permission from:

Nguyen, M., Shen, K., Sherck, N., Köhler, S., Gupta, R., Delaney, K.T., Shell, M.S. and Fredrickson, G.H., 2023. Complexation of Polycation PDADMA with Mixed Micelles of Sodium Dodecyl Sulfate and Ethoxylated Surfactants. The European Physical Journal E. Accepted.

4.1 Introduction

Solutions of polyelectrolytes and surfactants underpin many applications covering personal care products,[6,7,122–124] food products,[4,5,125] and drug encapsulation.[126–128] These systems feature an interesting range of self-assembly and phase behaviors, driven by the interplay of hydrophobic and electrostatic interactions.[53,129–132] Additionally, polyelectrolyte-surfactant systems are often highly multi-component, featuring at least one polyelectrolyte, a mixture of both neutral and ionic surfactants, cationic and anionic salt species, and water, with many parameters (composition, specific chemical species, molecular weight, temperature, etc.) that can be tuned to achieve desired properties.[23,133,134] The emergent self-assembly and phase behavior sets a range of rheological properties such as viscosity, lubrication, foamability, and wettability that are key performance targets in formulation design.[135,136]

There is a significant number of experimental works on the behavior of polyelectrolyte-surfactant mixtures. In polyelectrolyte-free solutions of surfactants, the critical micelle concentration (CMC) describes the onset of surfactant self-assembly into micelles.[53,137] In the presence of polymers or polyelectrolytes, there is usually a critical aggregation concentration (CAC) that precedes the CMC, where attractions between polymer and surfactant enables the formation of aggregates at concentrations lower than the surfactant's intrinsic CMC;[138] these aggregates are also referred to as complexes.[139] Hydrophobic polymers are expected to partition to the core of the resulting aggregates and micelles, while charged polymers typically decorate the surfaces of ionic surfactant aggregates

and micelles.[53,133,140,141] At higher concentrations, aggregates and micelles are known to undergo morphological changes into cylindrical micelles, spherical vesicles, lamellar structures, precipitates, and gels.[142–146] These structures in turn can be highly sensitive to the chemistry of the employed polyelectrolytes, surfactants, and salt.[23,130,144,147]

Mixtures of polyelectrolytes with oppositely charged micelles can also undergo phase separation through complex coacervation.[141,148–151] This charge-driven phenomenon results in a coacervate phase rich in macroions, and a supernatant phase lean in macroions, and naturally is highly sensitive to factors like the salt concentration, macroion surface charge, and polymer charge density. Complex coacervation is a very general phenomenon such that the macroions can be polyelectrolytes,[152] surfactant micelles,[141] charged colloids,[153,154] and proteins among others,[155–157] and as a result has garnered significant interest across many fields. Out of these various systems, mixtures of micelles with polyelectrolytes are particularly challenging to study and understand because of the propensity of micellar sizes and morphologies to change significantly with solution conditions. Additionally, experimental systems often employ *mixtures* of surfactants in order to tune micelle properties, underscoring the importance of understanding self-assembly.[23,141]

In this work we develop a multiscale modeling framework that directly addresses surfactant self-assembly in the presence of polyelectrolytes, and we focus on experimental studies by Dubin *et al.*[23,147] In a series of titration and dilution experiments, Dubin and coworkers revealed how polyelectrolyte-surfactant complexation is highly sensitive to polyelectrolyte linear charge density, micelle surface charge density, and the ionic

strength of the solution.[133,147,158] Additionally, the ratios of polyelectrolytes to surfactants were also shown to affect the resulting self-assembly behavior.[23] Taken together, these observations corroborate the intuition that electrostatic interactions are a primary driving force for the attraction of polyelectrolytes to charged surfactant micelles.

We model one of the systems they studied, featuring the cationic polyelectrolyte polydiallyldimethylammonium (PDADMA), anionic surfactant sodium dodecyl sulfate (SDS), nonionic ethoxylated surfactants (C_mEO_n), sodium chloride salt, and water (Fig. 6.4). In these studies, SDS surfactant solution was prepared above the CMC, and a mixture of PDADMA and C_mEO_n was iso-ionically titrated by the SDS micellar solution, thus steadily increasing the molar ratio of charged to neutral surfactants. Turbidity was monitored and demonstrated distinctive changes upon titration with the SDS solution. At modest mole fractions of charged surfactant SDS ($\sim$0.15-0.3) compared to the total surfactant concentration, the turbidity begins gradually increasing corresponding to the onset of the formation of soluble complexes, before exhibiting a sharp increase indicating a biphasic coacervation regime at higher mole fractions.[147] In addition to quantifying the effects of charge, Dubin and coworkers also demonstrated how changing the nonionic surfactant species C_mEO_n (i.e., by changing the length n of the hydrophilic EO group) quantitatively changes the observed transitions.

While there are numerous theoretical and simulation models of oppositely charged polyelectrolytes,[34,70,118,152,159–168] there is comparatively less theoretical work addressing the complexation of polyelectrolytes with surfactants.[139,140,146,169–172] Directly simulating

these systems by molecular dynamics (MD), even coarse-grained MD, is usually confined to relatively small systems. Additionally, the equilibration of micellar structures is itself limited by the slow processes of diffusion, surfactant exchange, and micelle fission and fusion.[173–176] The characteristic time for these processes is easily on the order of μs for common surfactants.[177,178] Adding slowly equilibrating polymer chains only makes it more difficult to confidently equilibrate the resulting systems. As a result, many molecular dynamics simulations of polyelectrolyte-surfactant systems model the micelles as pre-assembled colloidal particles.[154,179] Similarly, many theoretical treatments of polyelectrolyte-surfactant mixtures also avoid the difficult equilibration of surfactants into micelles altogether,[139] and make approximations such as treating the micelles as spheres with fixed size and charge.[139,180,181] In these models, the micelles act as external potentials on the surrounding polyelectrolyte solution, which is in turn modeled using field-theoretic models. Such approaches neglect the self-assembly of surfactants into micelles in order to facilitate evaluation of the mixture phase behavior.

In contrast to modeling micelles as colloids with fixed properties, in this study we explicitly study the self-assembly of surfactants and polyelectrolytes in soluble complexes. We achieve this by developing a *molecularly informed* field-theoretic model based on the strategy outlined in Chapter 2 and our publications (Fig. 6.4).[82,84,182] Briefly, the approach utilizes the field theory to efficiently simulate large, dense systems containing high-molecular-weight species, while providing chemical specificity through small-scale, all-atom simulations. Specifically, we use relative entropy coarse-graining to derive

chemically-sensitive coarse-grained interaction parameters from all-atom simulations.[81] The chemical specificity preserved by this approach allows us to make direct comparisons to the mixed micelle experimental studies by Dubin and coworkers.[147] To facilitate the calculation of *equilibrium* micelle properties we transform the coarse-grained particle-based model into field-theoretic form.[74] This alternative representation of the same particle-based model provides facile access to the free energy, and has been used to evaluate micelle properties and self-assembled morphologies of surfactants.[140,182]

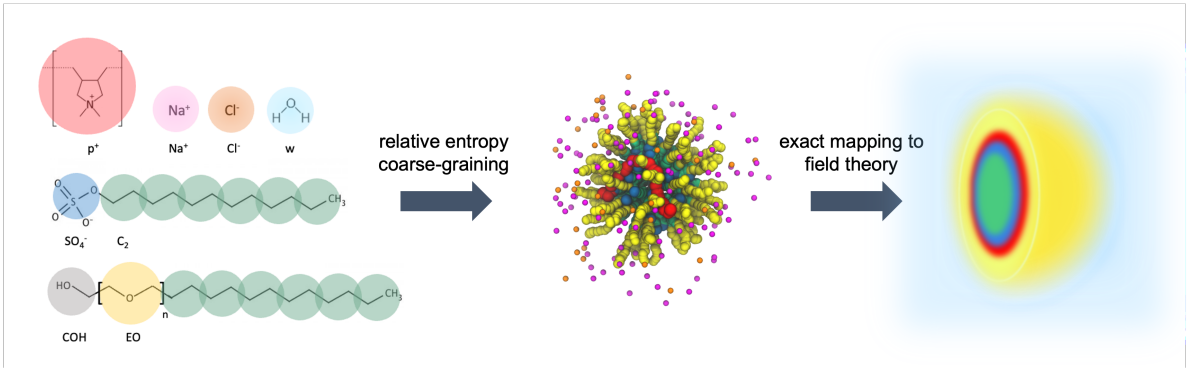

Figure 4.1: Schematic of the the multi-scale simulation workflow for deriving a field-theoretic model of polyelectrolyte and micelle complex. The left panel lists the components we consider in this work. From left to right and top to bottom, chemical structures of the polycation PDADMA monomer, Na^+, Cl^-, water, anionic surfactant SDS and non-ionic surfactant $C_{13}EO_n$ overlaid by the corresponding coarse-grained bead types. The middle panel shows a coarse-grained model parameterized by relative entropy minimization, while the right panel shows the exact mapping from a coarse-grained particle-based description of a PDADMA-SDS/$C_{13}EO_n$ mixed micelle complex to a field-theoretic model (cross-section of the micelle).

Using this approach, we self-consistently determine the mixed micelle size and morphology in tandem with polyelectrolyte adsorption. Subsequently, we show that the adsorption of polyelectrolytes to micelles correlates well with experimental measures of turbidity, and our model correctly predicts the effect of salt concentration and nonionic surfactant identity. The proposed simulation strategy facilitates the rigorous determina-

tion of equilibrium micelle structures, which is challenging to achieve in particle-based simulations, while considering the intricate balance of interactions among all components present in the solution. This establishes a basis for examining complexation propensity in other polyelectrolyte-surfactant mixtures, enabling efficient exploration of formulation space.

4.2 Computational details and methods

4.2.1 All-atom simulations

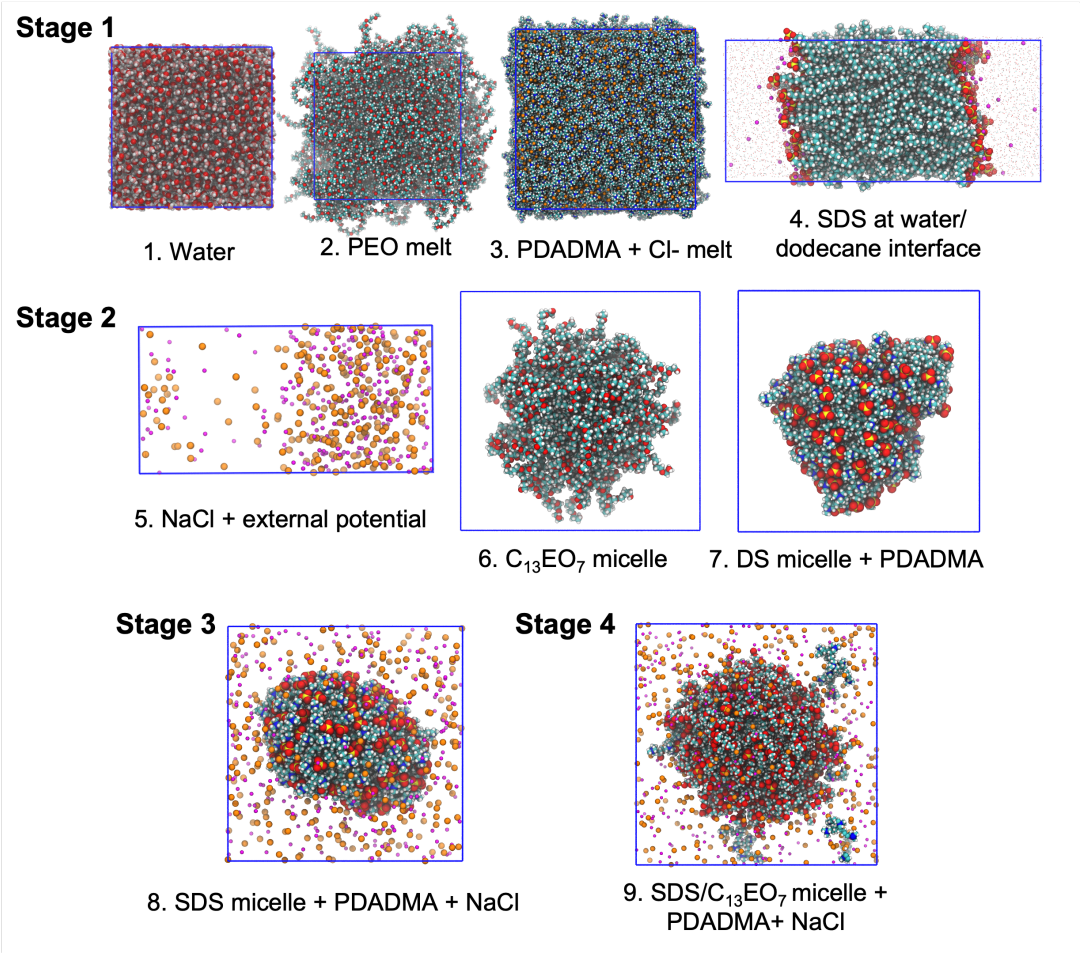

Figure 4.2: All-atom simulations used to derived coarse-grained interaction parameters via the relative entropy coarse-graining framework. Simulations details are provided in Table G.1

We parameterize the pair-wise interactions and bonded interactions for the eight coarse-grained (CG) bead types (Fig. 6.4, first panel) via four stages based on nine reference all-atom (AA) simulations as presented in Fig. 6.3. We use the Optimal Point Charge (OPC) 4-point water model[92] and the Joung-Cheatham ion model.[105] For the PDADMA and the

ethylene oxide block on $C_{13}EO_7$, we use the General Amber Force field (GAFF2).[106] We model the alkyl block (appearing in SDS, $C_{13}EO_7$, and dodecane) with the Lipid 14 force field.[183] We adopt the same force field used in our previous work for the headgroup sulfate SO_4^-,[182] which combines the parameters from Yan *et al.* (2010)[184] and subsequently adjusted nonbonded Lennard Jones interaction between sodium ions and oxygens bound to the sulfate headgroup to reproduce the surface tension of SDS deposited at a water-vacuum interface.

We conduct reference AA simulations with the OpenMM simulation package.[109] We use a 1 nm cutoff for the direct space non-bonded interactions and use the Particle Mesh Ewald method to compute long range Coulomb and Lennard-Jones interactions (LJPME 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 the Monte Carlo barostat that is updated every 25 time steps. We generate the initial configurations for the simulations with the Packmol package.[185] Details of the system sizes are provided in Table G.1 in Appendix E.

4.2.2 Bottom-up coarse-graining procedure

As detailed in Chapter 2, after performing AA simulations as described in the previous section, we use relative entropy coarse-graining[81] with these as references to derive CG interaction potentials (Section 2.2.3) that are amenable to efficient field-theoretic simu-

lations. We translate AA reference trajectories for coarse-graining by mapping 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 Na^+ and Cl^- ion to a single bead with $+1$ and -1 charge, respectively. Each monomer of PDADMA is mapped to one bead that bears a $+1$ charge. The 12-carbon alkyl tails on both SDS and $C_{13}EO_n$ are mapped to six neutral C_2 beads of two carbons each. The sulfate head group is represented as a single SO_4^- bead of charge -1. Lastly, the ethylene oxide repeating unit (CH_2OCH_2) on $C_{13}EO_n$ is mapped to one neutral EO bead and the terminal CH_2OH group is modeled as a neutral COH bead. We fix the Gaussian regularization range, a_α, of each bead species to approximately the cube root of its molecular volume. By this convention, the water interaction range, a_w, is set to 0.31 nm in this work. Values for other bead types are listed in Table G.3. The CG bead types and corresponding CG molecules are presented in the first panel of Fig. 6.4.

We derive CG parameters in successive stages; once the parameters are determined, they are fixed in subsequent steps. In the first stage, we determine CG parameters for the interactions between intramolecular bead species. The water-water repulsion v_{ww} was derived in Chapter D from a pure water AA simulation to reproduce the compressibility of OPC water, $\kappa_T \sim 0.062\ k_BT/a_w^3 \sim 4.51 \times 10^{-10}$ Pa^{-1}. This determines the CG pressure of $P_{CG} = 8.5\ k_BT/a_w^3$ that we use in the subsequent coarse-graining stages. We follow the previously published coarse-graining procedure[182] to parameterize pairwise and bonded interactions for the CG beads in SDS ($v_{C_2C_2}$, $v_{SO_4^-SO_4^-}$, $v_{C_2SO_4^-}$, $b_{C_2C_2}$, and $b_{C_2SO_4^-}$) from

SDS deposited at a water/dodecane interface at 298.15 K. The EO-EO repulsion $v_{EO\,EO}$ and root-mean-square bond length $b_{EO\,EO}$ are derived using a neat PEO simulation in the NPT ensemble. Similarly, pairwise excluded volume and bonded parameters for the polycation monomer, $v_{p^+p^+}$ and $b_{p^+p^+}$, are also determined from neat PDADMA chains in the NPT ensemble.

In the second stage, we parameterize the ions and cross-interactions involving water as well as those between PDADMA and SDS. We adopt the ion parameters ($v_{Na^+Na^+}$, $v_{Cl^-Cl^-}$, $v_{Na^+Cl^-}$, v_{Na^+w}, and v_{Cl^-w}) from our previous polyelectrolyte work (Chapter 3)[82] in which we coarse-grained from an aqueous NaCl solution in the external potential ensemble.[100] From the AA simulation of the non-ionic surfactant $C_{13}EO_7$ micelle, we derive the following parameters: $v_{COH\,COH}$, $v_{COH\,C_2}$, $v_{COH\,EO}$, v_{C_2EO}, $v_{COH\,w}$, $v_{C_2\,w}$, $v_{EO\,w}$, b_{C_2EO}, and $b_{COH\,EO}$. Lastly, we determine the cross-interactions among PDADMA, SDS and water ($v_{p^+C_2}$, $v_{p^+SO_4^-}$, v_{p^+w}, $v_{SO_4^-w}$) from a mixture of PDADMA with a SDS micelle.

In stage three, we derive the ions-PDADMA and ions-SDS pair-wise interaction parameters ($v_{p^+Cl^-}$, $v_{p^+Na^+}$, $v_{C_2Cl^-}$, $v_{C_2Na^+}$, $v_{SO_4^-Cl^-}$, and $v_{SO_4^-Na^+}$) from a similar PDADMA and SDS micelle system in the presence of NaCl. In the last stage, the remaining parameters (v_{p^+COH}, v_{p^+EO}, $v_{COH\,Cl^-}$, $v_{COH\,Na^+}$, $v_{COH\,SO_4^-}$, $v_{EO\,Cl^-}$, $v_{EO\,Na^+}$, and $v_{EO\,SO_4^-}$) are derived from a mixture of a SDS/$C_{13}EO_7$ micelle, PDADMA, and NaCl, around the composition range that we target in this study. We tabulate the parameters in Tables G.2 and G.3.

4.2.3 Coarse-grained molecular dynamics (CGMD)

CGMD simulations are conducted using a Langevin Dynamics integrator. The Langevin relaxation time τ_{CG} is taken as the unit of time. Due to the soft nature of the CG interactions and bonds, large time steps of 0.1 τ_{CG} are feasible. The inner-loop of relative entropy optimization requires CGMD simulations, which we run for $2 \times 10^5 - 1.5 \times 10^6$ time steps ($2 \times 10^4 - 1.5 \times 10^5$ τ_{CG} of simulation time), sufficient to equilibrate slow relaxation modes in the systems, such as the surface area in the interfacial system used to parameterize SDS interactions with water and salt.

4.2.4 Micelle simulations with the field theory

As discussed in Section 2.3, the nonbonded CG potentials defined in Section 2.2.3 can be readily represented and simulated using a field-theoretic representation by means of a Hubbard-Stratonovich-Edwards transformation. This field-theoretic transformation decouples nonbonded pair interactions, resulting in particles interacting only with an auxiliary field. As a result, particle coordinates can be analytically integrated, yielding a partition function in terms of integrals over field configurations. In this chapter, we invoke the mean-field approximation (Eq. 2.10) to study surfactant self-assembly. SCFT provides a readily-accessible approximation of the free energy, which is a powerful tool for evaluating the relative stability of competing structures particularly in comparison to conventional MD-based methods requiring significant, often intractable efforts to evaluate phase free energies. For example, minimizing the free energy at constant concentration

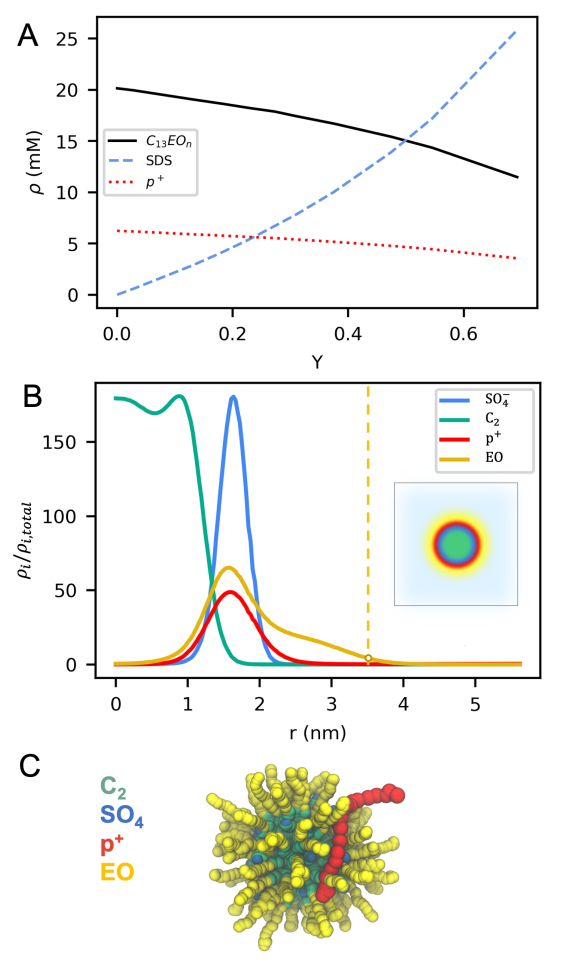

Figure 4.3: (A) Concentrations of SDS (chain basis), $C_{13}EO_n$ (chain basis), and DADMA (monomer basis) as the titration proceeds. (B) An example density profile and a cross section of the micelle from SCFT. Dashed line indicates the 0.1 EO locus introduced in the main text. A representative snapshot from CGMD is shown in (C).

with respect to the simulation cell size allows one to determine equilibrium sizes and aggregation number of molecular self-assemblies. Additionally, one can compare the free energies of different morphologies (e.g., spherical, cylindrical, lamellar) under stress-free conditions to determine the putative equilibrium structure.

In this work, we take advantage of this particular strength of the field-theoretic transformation to characterize the equilibrium self-assemblies of the multi-component polyelectrolyte-surfactant mixture. We perform simulations of micelles at varying compositions as in the titration experiment outlined by Dubin *et al.* (2017).[147] Briefly, 60 mM SDS in NaCl is added to solutions of 1 g/L PDADMA (with Cl^- counterions), 20 mM $C_{13}EO_n$, and the same NaCl concentration as in the SDS stock solution. As the titration proceeds, the anionic surfactant fraction increases and the composition changes according to Fig. 4.3A. This fraction is defined by the surfactant number densities as:

$$Y = \frac{\rho_{SDS}}{\rho_{SDS} + \rho_{C_{13}EO_n}}. \tag{4.1}$$

Fig. 4.3B presents an example density profile of a micelle from SCFT. A CGMD snapshot of an equivalent particle-based model is shown in Fig. 4.3C. The mixed micelle comprises of the non-ionic surfactant $C_{13}EO_n$ and SDS with an alkyl-rich core and a corona of SO_4^-, EO, and COH groups. At conditions where the micelle surface charge is above a critical value, the density profile shows an enrichment of PDADMA around the peak concentration of the SO_4^- head group, indicating the adsorption of polyelectrolytes on the micelle surface.

We define several variables to characterize surfactant assembly. First, the number

density of PDADMA monomers adsorbed onto the micelle is calculated by summing over their densities $\rho_{p^+,m}$ at mesh points m that are enclosed within a cut-off distance from the micelle center:

$$\rho_{p^+,adsorbed} = \sum_{m \in r=[0,0.1\ EO\ locus]} \rho_{p^+,m} \tag{4.2}$$

Here, we choose the cut-off distance where the EO density is 10% of its peak value to the right of the peak (dashed line in Fig. 4.3B); we refer to this as the "0.1 EO locus". We note that results are not very sensitive to the exact choice of this locus, as long as the adsorbed polyelectrolyte layer lies fully within its radius. Consequently, the fraction of adsorbed PDADMA monomers is

$$f_{p^+,adsorbed} = \frac{\rho_{p^+,adsorbed}}{\sum_m \rho_{p^+,m}} \tag{4.3}$$

The micelle surface charge density is calculated based on contributions from the surfactant only, i.e., SDS and $C_{13}EO_n$. Since SO_4^- is the only charged species, the surface charge density is

$$\sigma = \frac{\sigma_{SO_4^-}}{4\pi r_s^2} \sum_{m \in r=[0,0.1\ EO\ locus]} \rho_{SO_4^-,m} \tag{4.4}$$

where $\sigma_{SO_4^-} = -1$ e is the charge of SO_4^- bead. The cut-off distance r_s is taken to be at the 0.1 EO locus and is used to estimate the surface area in the normalization factor.

The aggregation numbers for SDS and $C_{13}EO_n$ are

$$N_{agg,SDS} = \Delta V \sum_{m \in r=[0,0.1\ EO\ locus]} \rho_{SO_4^-,m} \tag{4.5}$$

$$N_{agg,C_{13}EO_n} = \frac{\Delta V}{n} \sum_{m \in r=[0,0.1\ EO\ locus]} \rho_{EO,m} \tag{4.6}$$

where n is the number of EO beads in each non-ionic surfactant molecule. In addition,

we measure the core and micelle radii at the 0.1 C_2 and 0.1 EO loci, respectively.

4.3 Results and discussion

4.3.1 SDS titration

We assess the the binding between the polyelectrolyte and the surfactant micelle *via* the amount of 100-mer PDADMA chains adsorbed onto the micelle, as defined in Eq. 4.3, as the SDS mole fraction Y increases during the titration. Fig. 4.4A shows that the fraction of adsorbed PDADMA relative to the total PDADMA in the simulation box increases with Y for both SDS/$C_{13}EO_{11}$ and SDS/$C_{13}EO_7$ mixed micelle systems in 0.4 M NaCl, indicating more PDADMA binds to the mixed micelles with the addition of SDS to the mixture. This is expected as micelle surface charge density becomes more negative with increasing anionic surfactant mole fraction (Fig. 4.4B) which strengthens electrostatic interactions. Experimental measurements that infer micelle charge *via* zeta potential measurements also find that the micelle surface charge density increases with increasing SDS content.[147]

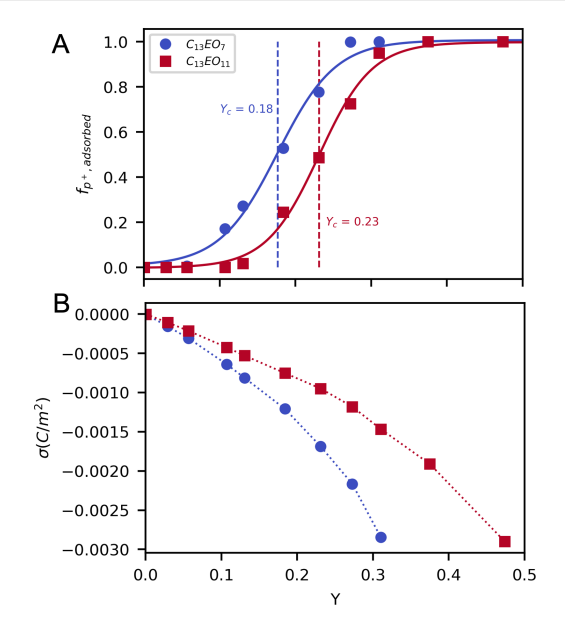

Figure 4.4: (A) SCFT-predicted binding isotherms showing the fraction of bound PDADMA (100-mer) as the anionic surfactant mole fraction, Y, increases during the course of the titration at a constant NaCl concentration of 0.4 M for SDS/$C_{13}EO_{11}$ and SDS/$C_{13}EO_7$ micelles. Solid lines are fits to sigmoid functions and Y_c is the Y value at the inflection point. (B) Corresponding micelle surface charge density.

CGMD provides further insight about the interactions between the micelle and polycations. In the first set of CGMD runs, Fig. 4.5, we initialize a single micelle using the aggregation numbers obtained from SCFT simulations and observe the interaction between a 24-mer PDADMA and the micelle for 2×10^4 τ_{CG}. As expected from SCFT results, PDADMA does not interact with the micelle at $Y = 0$ whereas it adsorbs to the micelle surface when the SDS content increases to $Y = 0.31$. We further investigate the inter-micelle binding at $Y = 0.31$ by doubling the system size to realize two identical micelles and one 48-mer PDADMA. The polycation initially binds to one of the micelles at $t = 0$ and quickly binds to both as the simulation proceeds (Fig. 4.6A). This inter-micelle binding results in an aggregation of the two micelles, evidenced by the decrease in the

center-of-mass distance between them as shown in Fig. 4.6B. This observation aligns with experimental hypothesis that the increased number of multi-micelle aggregates increases solution turbidity near the onset of complexation.[147]

Previous experimental studies[23,147] have quantified the onset of polyelectrolyte-micelle complexation with the critical composition Y_c, indicated by the initial increase in the turbidity. For comparison, we choose the inflection point of the PDADMA binding isotherms shown in Fig. 4.4A as Y_c. While this choice is arbitrary, it has relevant implications: complexation is considered to begin when a substantial number of polycation monomers (50% in this case) adsorb onto micelles, and the complexation is detectable through increased turbidity. Notably, the calculated Y_c values from the simulations are 0.18 and 0.23 for SDS/$C_{13}EO_7$ and SDS/$C_{13}EO_{11}$ mixed micelles, respectively, which are in good agreement with the Y_c values of 0.17 and 0.28 from experimental works by Dubin *et al.* (2017, 2018).[23,147] Good quantitative (Y_c of SDS/$C_{13}EO_7$) and qualitative (increasing Y_c from SDS/$C_{13}EO_7$ system to SDS/$C_{13}EO_{11}$ system) agreement with experimental

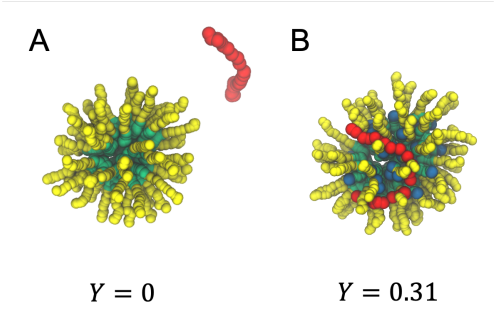

Figure 4.5: CGMD snapshots of a single SDS/$C_{13}EO_{11}$ micelle with a 24-mer PDADMA in 0.4 M NaCl at (A) $Y = 0$ (no SDS) and (B) $Y = 0.31$, respectively, with the aggregation numbers taken from SCFT solutions at same conditions. CGMD confirms that PDADMA binds to the micelle at high Y, as suggested by the SCFT binding isotherms. Water and NaCl molecules are not shown for clarity.

results suggests that the Y value at the inflection point of the binding isotherm can reliably describe the onset of complexation observed in previous experimental works. We note that the qualitative trend of SDS/$C_{13}EO_{11}$ micelles exhibiting higher Y_c values than SDS/$C_{13}EO_7$ micelles remains consistent, even when considering other fractions of bound PDADMA as the criteria for the onset of complexation. This agreement helps to validate that our coarse-graining procedure produces a reasonable parameterization of this complex multicomponent system.

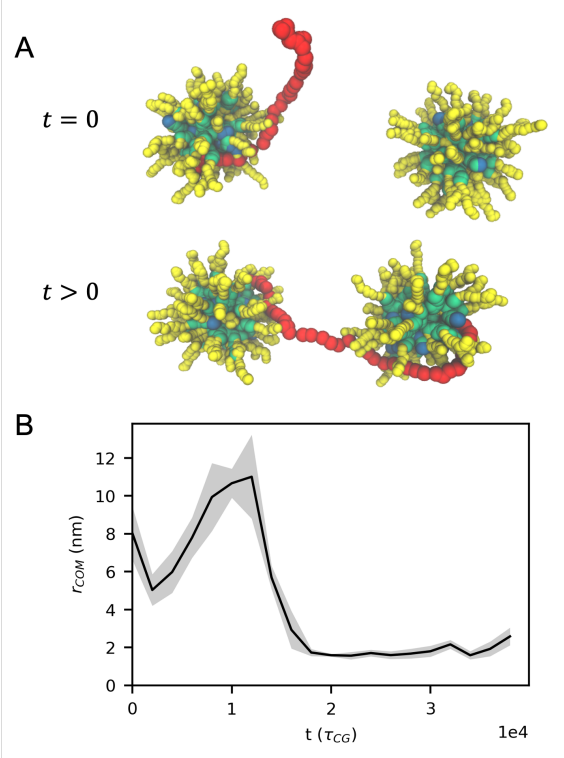

Figure 4.6: (A) CGMD snapshots of the same system of Fig. 4.5B at $Y = 0.31$ but at double the system size with a 48-mer PDADMA. PDADMA initially binds to one micelle $t = 0$ then bridges the two micelles as the simulation proceeds. (B) Center-of-mass distance between the two micelles during the course of the simulation. For reference, the average simulation box size length is 24.3 nm.

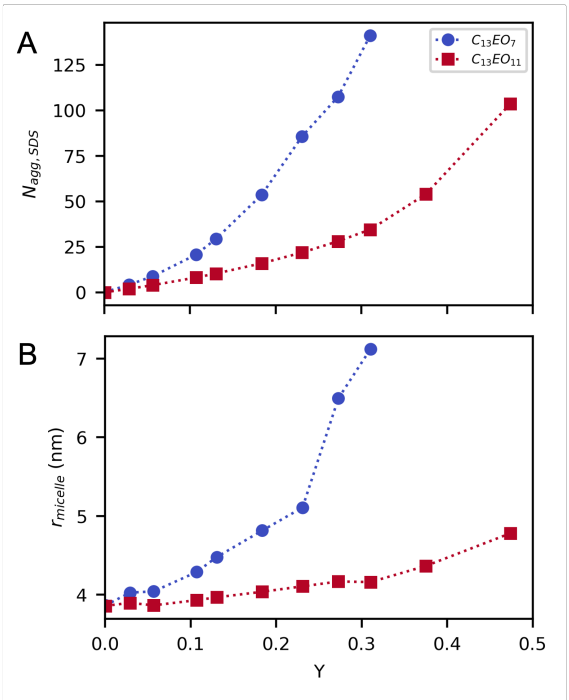

Figure 4.7: (A) SDS aggregation number and (B) micelle radius measured at 0.1 EO locus as a function of Y.

Dubin *et al.*[147] defined polyelectrolyte-micelle binding affinity as the resistance of polyelectrolyte-micelle complexes at fixed salt concentration to dissociation by the addition of non-ionic surfactant (decreasing Y), where smaller resistances indicate higher affinity. Thus, from this point of view, lower Y_c in the SDS/$C_{13}EO_7$ system from both simulations and experiments implies that SDS/$C_{13}EO_7$ micelles have higher polyelectrolyte binding affinity than SDS/$C_{13}EO_{11}$ micelles. This is supported by the higher micelle surface charge density when the non-ionic surfactant is $C_{13}EO_7$ (Fig. 4.4B). At a given value of Y, SDS/$C_{13}EO_7$ micelles have a higher SDS aggregation number $N_{agg,SDS}$ than SDS/$C_{13}EO_{11}$ (Fig. 4.7A). The radius of the SDS/$C_{13}EO_7$ micelle is also larger than that of the SDS/$C_{13}EO_{11}$ micelle as expected from significantly higher aggregation num-

bers of SDS and the nonionic ethoxylated surfactant (Fig. E.4). Despite having larger surface area, SDS/$C_{13}EO_7$ micelles still have higher surface charge density (Fig. 4.7B) which leads to stronger electrostatic interactions between PDADMA and SDS/$C_{13}EO_7$ micelles. Interestingly, the critical micelle surface charge density σ_c (σ at the Y_c transition) is more negative for SDS/$C_{13}EO_7$ than SDS/$C_{13}EO_{11}$ (Fig. E.3), suggesting higher charged micelles are required for complexation with the former system. We also note that experimentally-reported hydrodynamic radii of the complexes are in the range of 4-8 nm, which is in relatively good agreement with our simulation predictions.

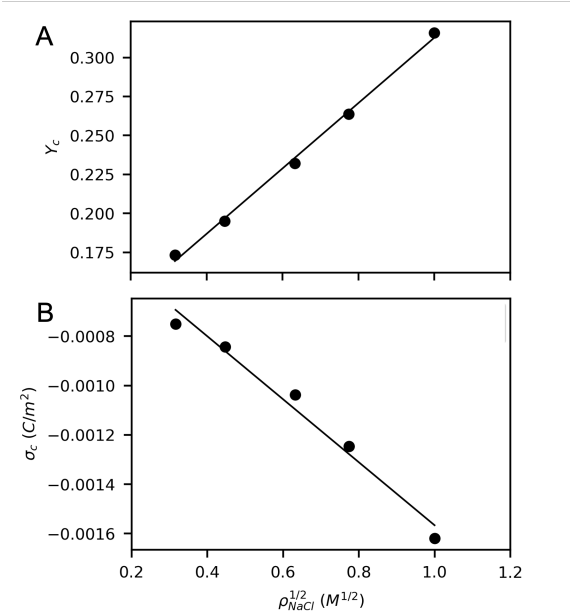

Figure 4.8: Dependence of critical (A) SDS content Y_c and (B) micelle surface charge density on added salt concentration for SDS/$C_{13}EO_{11}$ micelles. Solids lines are linear regressions with respect to the square root of the salt concentration.

Next, we investigate the effects of salt on the complexation between PDADMA and SDS/$C_{13}EO_{11}$ micelles by repeating the titration procedure at various NaCl concentrations. Fig. 4.8 presents critical conditions for polyelectrolyte-micelle complexation at

added salt concentrations between 0.1 and 1.0 M. While the micelle surface charge density at fixed Y is essentially constant with salt concentration (Fig. E.6), the critical SDS content Y_c shifts upward. The increase in Y_c with increasing salt concentration coincides with an increase in the magnitude of the critical micelle surface charge σ_c, which accounts for the salt's screening of electrostatic interactions. Notably, McQuigg *et al.* has also observed the linear dependence of the critical SDS content, Y_c, and micelle surface charge density σ_c with the square root of the added salt concentration in their experimental study of a similar system, PDADMA and SDS/$C_{12}EO_6$ micelles.[186] According to their simplified model for the binding of a polyelectrolyte to an oppositely charged colloid, such scaling arises when the electrostatic potential in the vicinity of the colloidal particle is less than 0.5 k_BT and the colloidal radius is large relative to the Debye length. As shown in Fig. E.8, the electrostatic potential at the 0.1 EO locus in our work is also less than 0.5 k_BT, and the micelle radius, either the core or total micelle radius (Fig. E.7C, D), is consistently larger by a factor of ∼2-10 than the Debye length, which varies from 0.96 to 0.30 nm in the 0.1-1.0 M salt concentration range. This implies that McQuigg's simplified model offers a reasonable representation of the micelles in our study, and the consistency in the scaling is justifiable.

4.3.2 Mesophase transitions

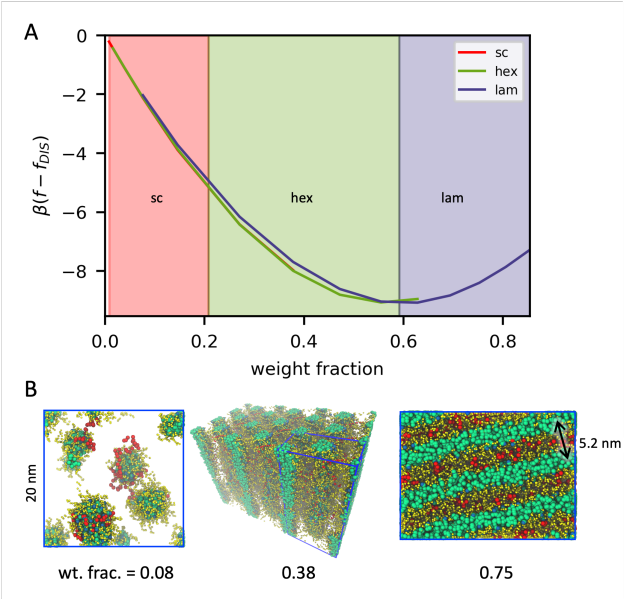

Figure 4.9: (A) SCFT free energy densities for mixtures of PDADMA and SDS/$C_{13}EO_{11}$ micelles in 0.4 M NaCl at $Y = 0.3$ and stoichiometric charge. The considered phases are simple cubic sphere (sc), hexagonal cylinder (hex), and lamellae (lam). The disordered free energy is used as reference values. (B) CGMD snapshots at solid (PDADMA, SDS, and $C_{13}EO_{11}$) weight fractions 0.08, 0.38, and 0.75.

So far, we have demonstrated that the CG model presented here shows good agreement with experimental observations for the binding of PDADMA and SDS/$C_{13}EO_n$ micelles in the low concentration regime. Next, we utilize field theory to determine self-assembled structures across a larger range of compositions for mixtures of PDADMA and SDS/$C_{13}EO_{11}$ micelles in 0.4 M NaCl. We set the SDS mole fraction to a fixed value of $Y = 0.3$ and maintain a charge stoichiometry of 1 between PDADMA and SDS. Then, we search for stable phases at increasing weight fractions of PDADMA, SDS, and $C_{13}EO_{11}$. As detailed elsewhere, the free energy density for a given mesophase, which can be directly accessed in SCFT, is minimized with respect to the simulation cell size.[187] A

structure is stable when its free energy density is the lowest relative to other candidate structures as well as the homogeneous disordered phase. Here, we consider three phases: simple cubic spheres, hexagonal cylinder, and lamellae (Fig. 4.9).

Fig. 4.9A presents free energy densities of these three phases relative to that of the homogeneous disordered phase. At low weight fractions, less than 0.21 weight fraction, SCFT predicts the stable structure is simple cubic with discrete micelles. When fluctuations are included, i.e., in CGMD, unbinding of the lattice occurs and and we observe a solution of spherical micelles (Fig. 4.9B). Previous assessments of polyelectrolyte-micelle binding reported in Section 4.3.1 lie in this regime (the solid weight fraction at $Y = 0.5$ on the titration path is ~ 0.015). As the concentration increases above 0.21 weight fraction, we cross over to a region where infinitely long cylindrical structures are stable. Since fluctuations are not included in SCFT, the infinitely long hexagonal cylindrical structure serves as an idealized estimate to an elongated micelle, and can correspond to both wormlike micelles as well as well-ordered hexagonal cylinder phases. Nevertheless, the infinitely long cylinder is a reasonable approximation, as evidenced by the CGMD snapshot of elongated micelles at 0.38 weight fraction. Lastly, lamellar structures, which are also captured in CGMD, are predicted to form above 0.59 weight fraction. This overall sequence of microstructures follows fairly standard, commonly-reported sequences (solution of spherical/rod micelles $\rightarrow$ ordered cubic and/or cylindrical phases $\rightarrow$ lamellar structures) in other surfactant systems.[182,188] Additionally, we show in Fig. E.2 that equilibrated micelle structures from SCFT closely match those of CGMD (corresponding to

micelles and simulation conditions shown in Fig. 4.5A, B). This suggests the mean-field treatment invoked in this work is sufficiently good to describe surfactant self-assembly and polyelectrolyte-micelle complexation, in line with the observation in our previous publication.[182]

Lastly, we note that in CGMD it is difficult to properly equilibrate micelle (both spherical and cylindrical) and periodic cell sizes to compare with the equilibrium cell dimensions obtained in SCFT. The exception is for lamellar structures, for which the lamellae can reorient in the simulation cell and thereby find an equilibrium interlayer spacing. For the lamellar simulation condition (weight fraction 0.75) shown in Fig. 4.9 we find that the CGMD interlayer spacing is ≈ 5.2 nm, which is in good agreement with the SCFT equilbirium layer spacing of 5.1 nm (Fig. E.9), confirming that SCFT is a good approximation for simulating these surfactant systems.

4.4 Conclusions

In this work, we have developed a molecularly informed field theory that faithfully describes the complexation behavior between polyelectrolytes and micelles. The model uses relative-entropy minimization to systematically coarse-grain from all-atom simulations. The resulting coarse-grained models are then transformed into a field-theoretic description, which enables rapid self-consistent field-theoretic simulations. Most importantly, field-theoretic simulations allow for direct the evaluation of free energies, and hence the rigorous determination of equilibrium micelle size and structure. In turn, we were able to

show how self-assembled micelles change over varying compositions of the nonionic and ionic surfactants, the polyelectrolyte, and salts. Such calculations are possible but much more difficult to perform accurately with equivalent particle-based representations owing to long time scales associated with micelle equilibration processes. Using our workflow, we were able to build a fully-parameterized model of the same molecular system considered by Dubin *et al.*[147] and studied how *mixed micelle* size and properties can change dramatically as surfactant concentration and mixing ratios change along experimental titration paths. Additionally, at high surfactant concentrations, we were also able to locate morphological transitions to cylindrical and lamellar structures to confirm that, for the system under consideration, no cylindrical micelles are expected for the titration paths we considered.

We found that the degree of polyelectrolyte adsorption correlated well with experimentally observed turbidity transitions that announce the onset of polyelectrolyte-micelle complexation. The turbidity transition was also estimated to within 5% of the experimentally reported anionic surfactant mole fraction. Additionally, our model correctly predicts that nonionic ethoxylated surfactants with shorter hydrophilic groups undergo complexation transitions at lower mixing ratios of anionic to nonionic surfactants. While ethoxylated surfactants with shorter head groups tend to form larger micelles than their counterparts with longer head groups, the micelles formed from ethoxylated surfactants with shorter head groups also exhibit higher surface charge density, thus explaining their stronger interaction with polyelectrolytes. Lastly, the critical mole ratio of anionic surfac-

tants and surface charge density of micelles were shown to vary linearly with the square root of salt concentration, in agreement with experimental observations.[186]

In conclusion, we have demonstrated that molecularly informed field theories are a powerful tool for exploring the self-assembly of multi-component systems like polyelectrolyte-surfactant mixtures. These field theories have the potential to be chemically specific and allow direct comparison with experiment. Using this tool, we studied how equilibrium self-assemblies of mixed micelles change with solution mixing conditions, and found that polyelectrolyte adsorption curves correlate well with experimental turbidity curves and the onset of soluble polyelectrolyte-micelle complexes. Although previous work[182] and the semi-quantitative agreement with experiments validate that the mean-field approximation used in this work is qualitatively good to describe surfactant self-assembly, including fluctuation effects could further improve predictions of the field-theoretic model. Within the confines of field theory, composition fluctuations that are ignored in the mean-field approximation could be incorporated using techniques such as complex Langevin sampling.[74,76] By fully sampling the partition function, complex Langevin recovers the same thermodynamic properties as CGMD while retaining rapid equilibration. However, while collective variables like the density are readily accessed within a field theoretic simulation, single-molecule properties require more care.[94] In this study, we focused on investigating the soluble polyelectrolyte-surfactant complexes which only covers a small region of a much richer phase diagram in these systems.[133] Future work can extend this study and rigorously determine phase boundaries in biphasic regions with complex coacervation

between polyelectrolytes and surfactants.[82]

Acknowledgments

This work was supported by BASF Corporation through the California Research Alliance. We thank Keith Gutowski for generously sharing papers and engaging in insightful discussions with us. G.H.F. and K.T.D. also derived partial support from the National Science Foundation CMMT Program under grant number DMR-2104255. M.S.S. acknowledges funding support from the National Science Foundation through Award No. CHEM-1800344. K.S. 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.

Supporting Information see Appendix E

Chapter 5

Estimating Critical Micelle Concentrations of Intrinsically Disordered Protein Surfactants

Reproduced in part with permission from:

Nguyen, M., Dolph, K., Delaney, K.T., Shen, K., Sherck, N., Köhler, S., Gupta, R., Francis, M.B., Shell, M.S. and Fredrickson, G.H. Estimating Critical Micelle Concentrations of Intrinsically Disordered Protein Surfactants. In preparation.

5.1 Introduction

The self-assembly of amphiphilic molecules, such as block co-polymers, surfactants, and biomolecules, plays a critical role in many natural and industrial processes. Examples

include the formation of biological cell membranes through lipid molecules, the micellization of surfactants in detergents, and drug encapsulation.[6,17,128,189] 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.[47–49] 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.).[190–197] Moreover, with the rising emphasis on sustainable and environmentally friendly practices in chemistry, there is a growing demand for the development of high-throughput screening methods as our 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.[198]

Current computational tools such as coarse-grained molecular dynamics[190] and dissipative particle dynamics[191,192] are commonly used to calculate the CMC in particle-based simulations. Most studies are performed in the NPT or NVT ensemble and use the concentration of free surfactants as an estimate for the CMC. To mitigate the need for large simulation boxes required near the CMC (typically on the order of mM for com-

monly studied surfactants),[191] most particle-based simulation work at concentrations much higher than the CMC rely on the assumption that the free surfactant concentration is constant above the CMC. This assumption has been proven to be inaccurate, especially for ionic surfactant systems.[193,199–201] Studies have shown that employing empirical corrections to account for crowding effects due to aggregate formations can provide more accurate predictions of the CMC.[193,194,202] These corrections, however, are system dependent and 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 the extrapolation and high computational cost of the large simulation boxes required to study near the CMC. In addition to the length-scale challenge, the self-assembly involves inherently long-time-scale processes related to diffusion, and micelle fission and fusion, which occur on the order of microseconds.[173–176] While atomistic simulations are intractable for capturing such the time-scale of phenomena, coarse-grained models also face challenges[193] in sufficiently sampling the free surfactant concentration and equilibrium distribution of aggregate sizes required for accurate estimation of the CMC in particle-based approaches.

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 micelle state by comparing its grand free energy with the homogeneous (non-aggregated) state at the same chemical potentials. One advantage of using the grand canonical en-

semble is the reduced simulation box size required for studying micelle formation. Unlike some previous 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, which is precisely the definition of the CMC.

While the grand canonical ensemble method is in principle exact and direct, it requires 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,[195] as chemical potentials and free energies can be directly evaluated through analytical approximations, including mean-field and Gaussian approximations[58,65,82], or numerically computed without approximation through field-theoretic simulations via complex Langevin sampling.[75,77] 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 chapter, we apply the multiscale simulation framework to enable 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 Francis and coworkers.[15,203] This class of bio-inspired, protein-based surfactants possesses a remarkable degree of tunability, which arises 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 offers more precise control over chain length and the individual building block sequence than the synthetic counterparts. Importantly, prior studies demonstrated that these surfactants possess encapsulating properties similar to commonly used synthetic counterparts with the CMC $\sim 10\,\mu$M.[15,203] 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 lim-

ited. 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 of often circumvent this by reducing the complex interactions to hydrophilic and hydrophobic interactions in simplified heteropolymer systems .[204,205] Here, we employ *molecularly informed* field theory, as described Chapter 2 .The method uses relative entropy coarse-graining[81] 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 the field-theoretic representation.[74] Because free energies and chemical potentials are readily calculated by operators in the field-theoretic representation, this approach enables us to directly determine the grand canonical free energy and chemical potential for CMC calculation while preserving important information about the underlying chemical components.

5.2 Computational details and methods

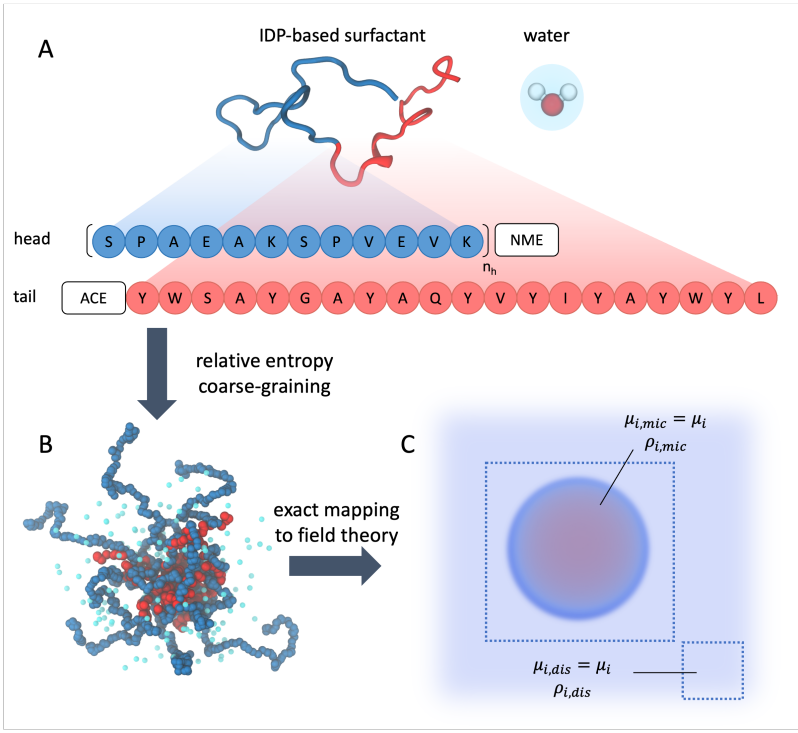

Figure 5.1: Schematic of the multi-scale simulation workflow to construct a molecularly informed field-theoretic model of IDP surfactants. (A) Species involved in the all-atom system which include the IDP surfactant and water. Instead of simulating the full surfactant sequence, we split the surfactant 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, $\mu_{i,mic}$, and disordered, $\mu_{i,dis}$, states of compositions $\rho_{i,mic}$ and $\rho_{i,dis}$, respectively.

5.2.1 All-atom simulations

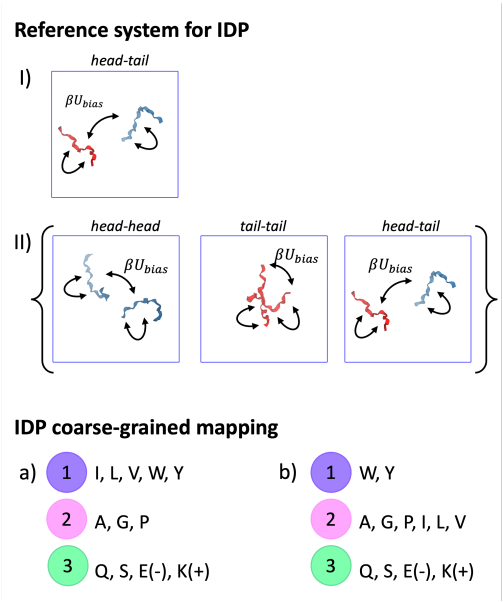

Figure 5.2: Reference AA systems and CG mapping schemes considered in constructing the IDP surfactant model.

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. 5.1A. The hydrophilic head sequence is inspired by neurofilament heavy arm side-chain protein found in human neurons 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. 5.1A. 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. 5.2: a simulation of the head and tail fragments, and an extended ensemble[99,206] 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 25500 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 to get 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:

$$\beta U_{bias}(r) = v_{bias} e^{-r^2/4a^2} \tag{5.1}$$

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 Section 6.2.2. 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. F.1 for the tail-tail simulation.

We employ the a99SB-disp force field, which was developed by Robustelli *et al.*. (2018) to accurately describe both folded and disordered proteins in tandem with the modified TIP4P-D water model.[207] We conduct reference AA simulations with the OpenMM simulation package.[109] 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 (LJPME 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 $1/(25\ dt)$. We generate the initial configurations for the simulations with the Packmol package.[185]

5.2.2 Bottom-up coarse-graining procedure

We follow the coarse-graining procedure detailed in Chapter 2. For simplicity, we use one universal bond length for all amino acid pairs, $b_{\alpha\gamma} = b$ in Eq. 2.6. We translate AA reference trajectories for coarse-graining by mapping 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 and lysine, 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. 5.2. 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. 2.8. 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 type. 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 in our previous publication.[84] In this step, we derive v_{ww} in the NPT ensemble at the CG pressure P_{CG} to 3.218 k_BT/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. 5.2. 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 assessment of the CMC that will be discussed in Section 5.4. We tabulate the parameters in Tables F.1-F.5.

5.2.3 Calculating the critical micelle concentration with field theory

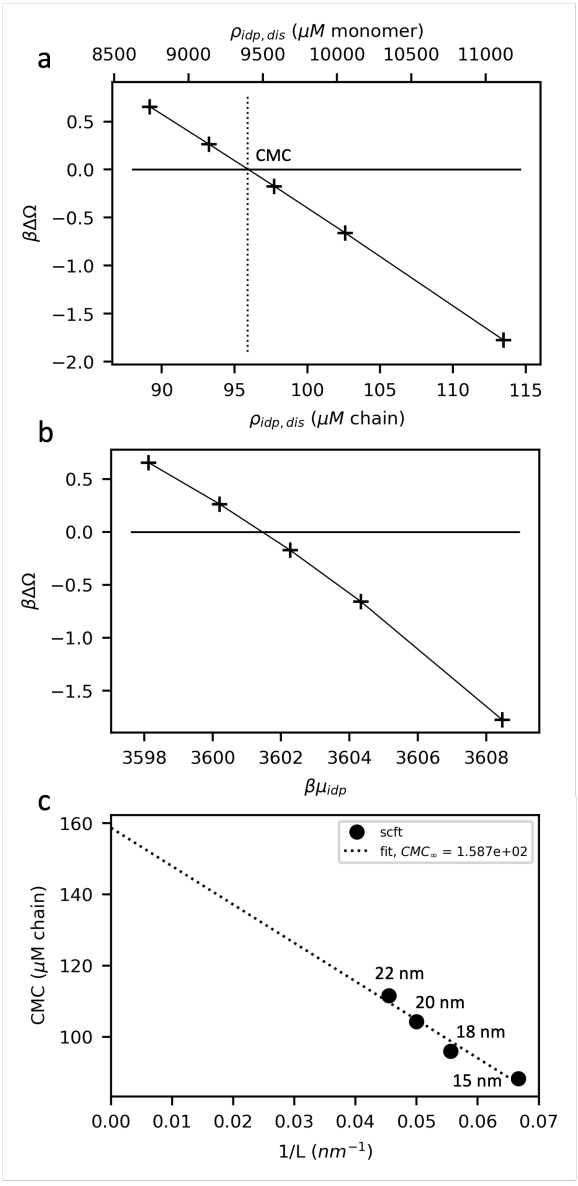

Figure 5.3: Example grand free energy difference $\beta\Delta\Omega$ between a spherical micelle and the homogeneous phase as a function of (A) surfactant concentration in the homogeneous phase $\rho_{idp,dis}$ (chain and monomer basis) and (B) surfactant chemical potential $\beta\mu_{idp}$. (C) To account for finite-size errors, we extrapolate the CMC linearly with respect to the inverse of the box size length.

To determine the CMC, we compute the grand free energy difference between a spherical micelle and the homogeneous phase in chemical equilibrium with that micelle, following the procedure outlined by Zhou *et al.* (2011).[195] Specifically, we perform SCFT simulations of a spherical micelle in the *canonical ensemble* for various values of IDP mole fraction, $\phi_{idp,mic}$. This results in a series of micelles at different IDP and water chemical potentials, $\mu_{w,mic}$ and $\mu_{idp,mic}$, respectively. The composition of the homogeneous disordered state that is in chemical equilibrium with each of these micelle states is determined from a *grand canonical* simulation at $\mu_{i,dis} = \mu_{i,mic} \equiv \mu_i$. This process is illustrated in Fig. 5.1C, 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. Fig. 5.3A-B shows example 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 $\rho_{idp,dis}$, 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 approximately 10 to 22 nm in side length. We then extrapolate the CMC values against the inverse box side length, $1/L$, (Fig. 5.3C) and extract the CMC at $1/L \to 0$ or as the micelle simulation box size approaches infinity ($L \to \infty$). We

repeat these steps for each surfactant model and report the mean and median values of the CMC.

5.3 Determination of the critical micelle concentration via pyrene-based fluorescence assay

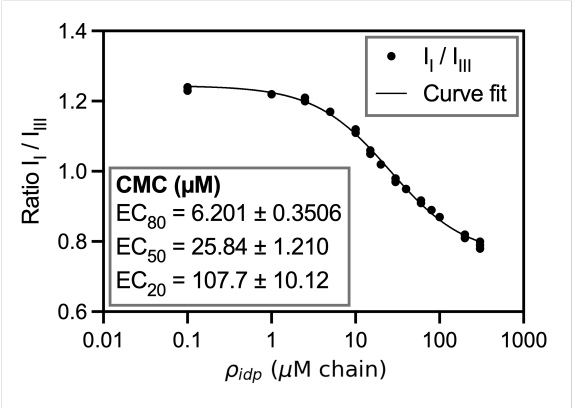

Figure 5.4: Pyrene I_I/I_{III} fluorescence emission ratio across concentrations of the IDP surfactant with $n_h = 6.5$. Solutions containing 0.1 μM to 300 μM surfactant in 2 μM pyrene and 10 mM phosphate buffer, pH 6.5 was excited at 330 nm and the emission was recorded at 373 nm (I_I) and 384 nm (I_{III}).

We experimentally determine the CMC of the surfactant with the hydrophilic block length of $n_h = 6.5$ experimentally by a solvatochromic pyrene-based fluorescence assay, as described previously.[15,203] 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 μM pyrene solution in 10 mM phosphate buffer, the I_I/I_{III} ratio is approximately 1.3 and lowers to approximately 0.8 when pyrene is 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:

$$EC_{50} = EC_F \left(\frac{100 - F}{F} \right)^{1/HillSlope} \tag{5.2}$$

$$y = Bottom + \frac{Top - Bottom}{1 + (EC_{50}/x)^{HillSlope}} \tag{5.3}$$

In Fig. 5.4, we report the CMC and its standard error as the EC_F of this nonlinear fit, where F is 80, 50, or 20. Previous studies suggest that the inflection point, EC_{50}, is a better approximation for surfactants with CMC values less than 1 mM.[208,209]

5.4 Results and discussion

5.4.1 Coarse-grained model evaluation

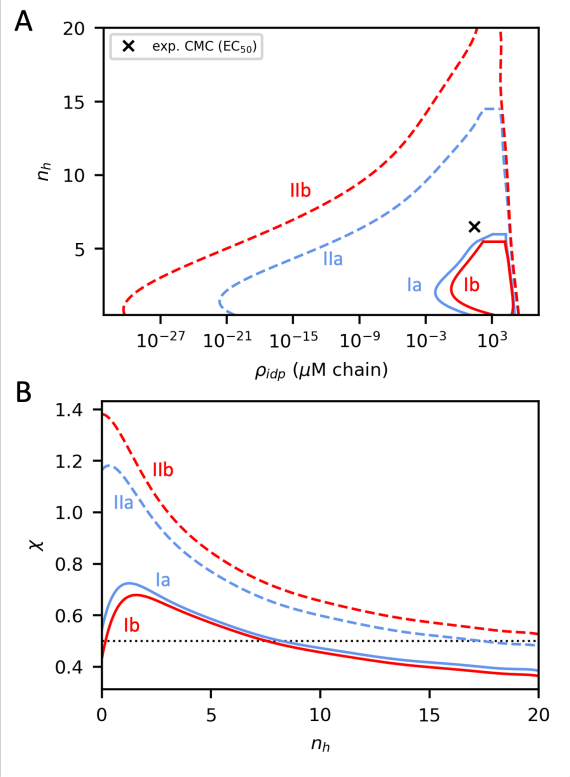

Figure 5.5: (A) Binodals calculated from representative parameter sets for four IDP models at varying number of hydrophilic repeating units, n_h. Symbol denotes the experimentally determined CMC (EC_{50} value) at $n_h = 6.5$. (B) Corresponding χ parameters against n_h. Dotted line denotes $\chi = 0.5$.

To calculate the CMC, it is necessary to have a well-defined homogeneous phase in co-existence 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 Section 3.2.2. Fig. 5.5A shows the binodals using representative parameter sets for the four different IDP models. As the number of the 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 approximately $\sim 10^{-10}$ μM. 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 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. Based on these observations, we only proceed to focus on models Ia and b for the CMC calculation.

To better understand the differences between the four CG models, we simplify the parameter space by reducing the 10 pair-wise interactions to a single effective Flory-Huggins parameter, χ.[210,211] 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 towards phase separation. In this work, χ is defined as

$$\chi = v_{ref}\left(\rho_w^*\rho_{idp}^* U_{idp\,w} - \frac{1}{2}(\rho_{idp}^{*2}U_{idp\,idp} + \rho_w^{*2}U_{w\,w})\right), \tag{5.4}$$

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 from using the mean-field approximation (detailed in Appendix A) as follows:

$$\rho_i^* = \frac{-1 + \sqrt{1 + 2U_{ii}P_{CG}}}{U_{ii}}. \tag{5.5}$$

Eq. 5.4 and 5.5 involve the excluded volume parameter U_{ij} between *molecules* i and j, which is defined by summations over bead and molecule species:

$$U_{ij} = \sum_{\alpha,\gamma\,\in[w,1,2,3]}\ \sum_{i,j\,\in[w,idp]} u_{\alpha\,\gamma} f_{i,\alpha} f_{j,\gamma} N_i N_j \tag{5.6}$$

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 $N_{idp} = 12\ n_h + 20$, 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. 5.5B. As a first approximation, phase separation typically occurs at $\chi \gtrsim 0.5$[212], 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

Table 5.1: 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	-	6.20 ± 0.35 (EC_{80}) 25.84 ± 1.21 (EC_{50}) 107.7 ± 10.1 (EC_{20})		19.6 ± 4.9 [a]	-

[a] reported in ref. 203

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 from the small center-of-mass distance between the tail fragments ($\sim$ 1 nm, Fig. F.2), the aggregated state is the dominant conformation, 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 F.6). 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.

5.4.2 Critical micelle concentration

While the CMC calculation procedure with SCFT is deterministic, the derivation of CG parameters via relative entropy minimization involves stochastic samplings from short CGMD simulations, leading to CG parameter variation. Therefore, we repeat the parameterization process 20 times and perform the necessary calculations for each

of these replicates to obtain statistically meaningful values. In Table 5.1, 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 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 EC_{50} 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, volume occupied by micelles, exhibit rapid *continuous* changes through the CMC instead of a sharp transition.[213] Therefore, different methods of inferring the CMC can yield slightly different values.[194,214–217] Specifically in this case, it is reasonable to expect micelles to form at any concentration between the EC_{80} and EC_{20} 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.

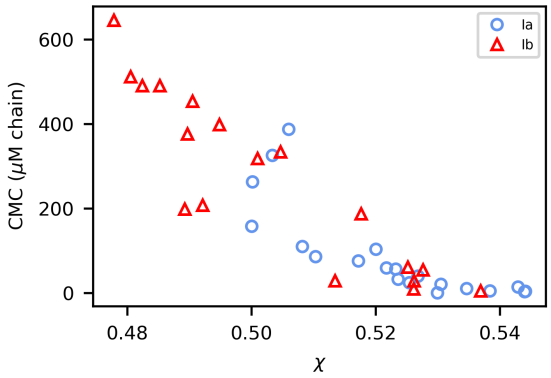

Figure 5.6: CMC at $n_h = 6.5$ calculated from 20 replicates for models Ia and Ib against χ. Higher average χ of model Ia suggests that this model is slightly more hydrophobic than model Ib, resulting in a lower average CMC value.

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. 5.6, where higher χ values promote micellization, resulting in smaller CMCs. This is reflected in model Ia, which has a higher average χ value, and thus 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. Fig. 5.7 illustrates that even small perturbations in the interaction parameters can significantly affect the CMC. Specifically, a mere increase of approximately 0.02 k_BT in the excluded volume strength between species 1 and 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_BT across 20 replicates for both models Ia and Ib (Fig. F.4), 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 kJ/mol or 0.8 k_BT.[218]

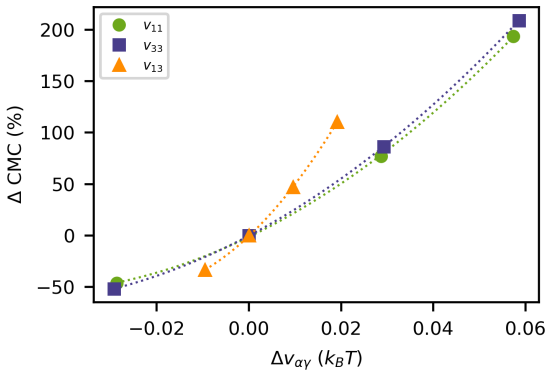

Figure 5.7: Sensitivity analysis of the CMC with respect to the change in excluded volume parameters v_{11}, v_{33}, and v_{13}. A plot that shows the percentage of change in the CMC with respect to the change in the excluded volume parameter from the base value.

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 the relative entropy minimization.[81] In this study, we have chosen 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 potentially further reduce parameter variations, and this will be carefully considered in future work.

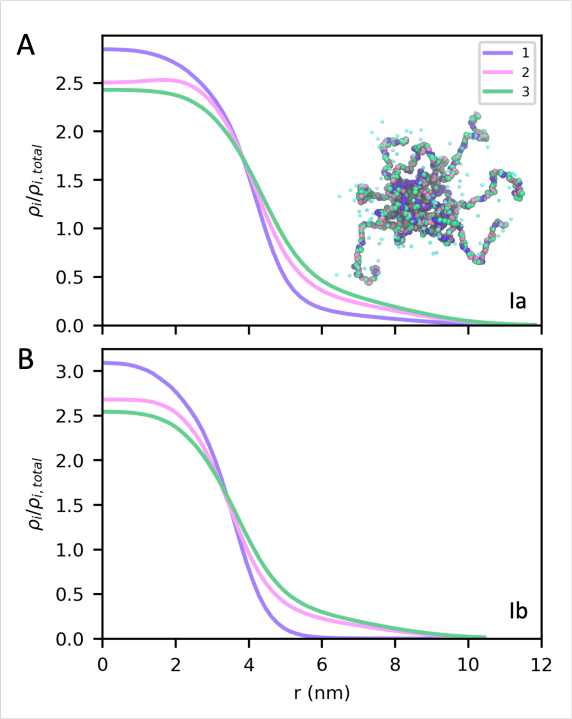

Figure 5.8: Density profile of micelles at $n_h = 6.5$ from representative parameter sets for models (A) Ia and (B) Ib at the corresponding CMCs. Inset is a snapshot of the micelle from CGMD reconstructed based on the equilibrium aggregation number calculated in the field theory.

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. 5.8). 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 that determined by dynamic light scattering experiments[203] as reported in Table 5.1. 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. 5.8. This can be done through pre-assembling micelles based on the SCFT-predicted aggregation numbers or implementing a backmapping strategy proposed by Lequieu.[219] The flexibility to transform between the 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.

5.5 Conclusions

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 chemical-specificity 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 have 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.[191,192] 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. 5.2). 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 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 from 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, thus 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 in Chapter 4,[83] incorpo-

rating w-field fluctuations using techniques such as complex Langevin sampling[74,76] 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.[77]

In conclusion, our study showcases the capability of molecularly informed field theories in systematically predicting 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.

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 (NIH T32GM066698). We thank Paul Huang and Kueyoung Kim for 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 number DMR-2104255. M.S.S. acknowledges funding support from the National Science Foundation through Award No. CHEM-1800344. K.S. 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 (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; NSF DMR 2308708) at UC Santa Barbara.

Supporting Information see Appendix F

Chapter 6

Effect of the Acetylation Pattern on the Miscibility of Cellulose Acetate

Reproduced in part with permission from:

Nguyen, M., Sherck, N., Köhler, S., Gupta, R., Shell, M.S. and Fredrickson, G.H. Elucidating the Effect of the Acetylation Pattern on the Miscibility of Cellulose Acetate from a Molecularly Informed Field-Theoretic Approach. In preparation.

6.1 Introduction

Carbohydrates are appealing from a sustainability perspective as they are bio-sourced and often biodegradable, such as cellulose.[220] Cellulose is seen to hold great potential for chemical modification due to the three hydroxyl functionalities per glucose monomer, however since cellulose is crystalline they remain shielded. Among the derivatives of

cellulose, cellulose acetate (CA) is one of the most widely used chemistries, utilized in films, membranes, fibers, and drug delivery, among others.[221] Cellulose itself is insoluble in many solvents including water due to strong intra- and inter-molecular hydrogen bonding between chains with a high degree of crystallinity which greatly hampers its utility in industrial formulations. To improve cellulose solubility, the alcohols are functionalized (e.g., acetylated) disrupting the intra- and inter-cellulose hydrogen bonding network. Throughout the literature there are numerous studies demonstrating that tuning the degree-of-substitution (DS) of the alcohols (yellow highlights in Fig. 6.1) with acetate, modulates cellulose solubility in a variety of solvents.

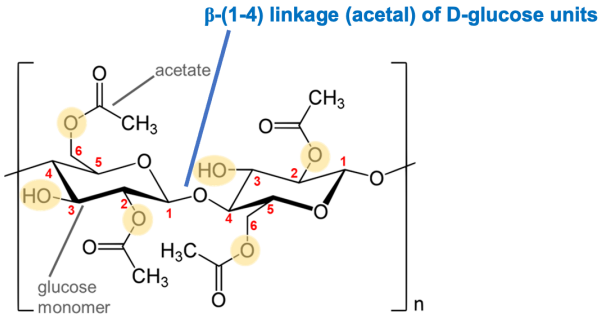

Figure 6.1: Representative cellulose structure consisting of two glucose monomers, each with two alcohols at the C6 and C2 positions substituted with acetate groups (DS=2). The yellow highlights denote the oxygens in the alcohols available to acetylation, while the red numbers denote the carbon numbering.

Broadly speaking, in the literature there are two routes to achieve partially acetylated CA. The first and oldest technique is a 2-step process that involves first completely acetylating the cellulose (DS $\sim$ 3) followed by partial deacetylation under acidic or basic conditions.[222,223] This is a harsh process and leads to degradation of the acetal linkages along the cellulose backbone. More recent developments have aimed at 1-step strate-

gies that are less harsh and less involved, such as acetylation in an ionic liquid.[221] A critical difference between the one- and two-step synthesis techniques is that they typically yield different ratios of substitution between the three alcohol sites. Albeit, in the two-step method during the deacetylation step, reacetylation is believed to occur which obscures direct measurement of alcohol reactivities through quantification of the relative DS substitutions.[224] In the one-step method the C6 position is preferably acetylated with relative ratios of acetylation for C6:C3:C2 of 14.1:3.7:1.0, respectively (total DS was 0.63).[221] Kamide *et al.* have observed that one-step protocols indeed lead to a greater fraction of C6 substitution relative to the two-step method; however they didn't observe a region of water solubility for CA produced by a one-step protocol.[223] Since Kamide *et al.* observed a more even DS among the alcohol sites in the two-step method, they concluded that acetylation at both the C2 and C3 positions is more important for water solubility than for C6 alone.

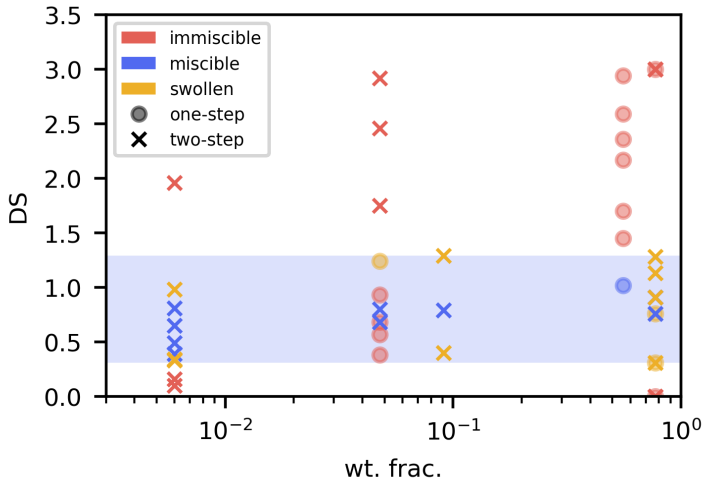

Figure 6.2: Cellulose acetate solubility data compiled from the literature. Circle and cross symbols denote one-step and two-step routes while , respectively. Colors denote insoluble (red), swollen/partially soluble cellulose acetate (yellow), and soluble (blue). Shaded blue region indicates soluble or swollen cellulose acetate from the *two-step* synthesis route.

Data on cellulose acetate (CA) solubility in water, compiled from various literature sources,[221,223–230] indicate a loosely defined region of solubility for CA in water, spanning DS values between 0.3 to 1.3 (Fig. 6.2). Unfortunately, many of the literature sources do not clearly define their protocols for assessing "solubility". There are several factors that can change the solubility in addition to the DS and synthesis method (one- or two-step): (1) CA molecular weight and polydispersity, (2) temperature, (3) method to mix and assess if soluble, and (4) the relative ratios of acetylation at three alcohol sites which is somewhat correlated with the particular protocol, either one- or two-step. When considering all available data (Fig. 6.2, both circle and cross symbols), there is uncertainty around where CA and water are miscible, with both insoluble and soluble datasets overlayed; however only considering two-step data reveals a clear DS region

appears for which CA is either soluble or swollen, shown in the shaded blue region of Fig. 6.2. Overall, there is a lack of detailed investigation into the factors that contribute to the miscibility of CA in water. Further research and detailed studies are needed to gain a better understanding of the factors influencing CA solubility and to establish a more comprehensive and accurate phase diagram.

In this work, our main objective is to develop a molecularly informed field-theoretic model to systematically study the impact of DS on the aqueous phase behavior of CA, validating the potential of this multiscale approach in studying polysaccharides. The construction of our CA model is based on the discussions presented in Chapter 2. Leveraging the computational efficiency of the field theory, we determine the phase boundary, which not only sheds light on how acetylation pattern and DS influence the miscibility of CA but also precisely identifies the CA composition that leads to miscible samples. More importantly, this work delineates the role of distinct acetylation sites and underscores their significance in CA's miscibility. Such insights are vital for guiding the design and synthesis of water-soluble CA with targeted properties for a wide range of applications and provide a computational platform for in silico screening of CA formulations.

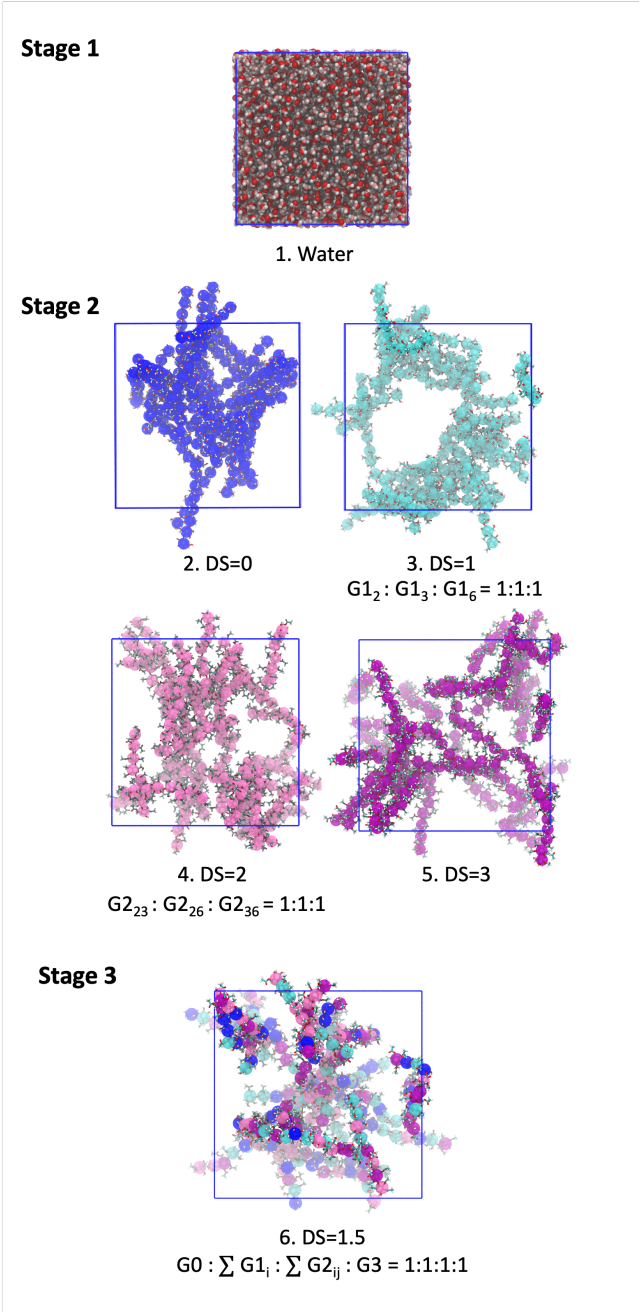

Figure 6.3: All-atom simulations used to derived coarse-grained interaction parameters via the relative entropy coarse-graining framework. In simulations 2-6, we overlay the coarse-grained bead types corresponding to the repeating units defined in Fig. 6.4; water is not shown for clarity. We denote underneath simulations 3, 4, and 6 the relative ratio of different glucose monomers in the simulation. Gx denotes a glucose repeating unit with x alcohol sites being substituted for acetate. For G1 and G2 monomers, the subscripts denote the acetylated alcohol sites (2, 3, or 6). Simulations details are provided in Table G.1

6.2 Computational details and methods

6.2.1 All-atom simulations

We parameterize the pair-wise and bonded interactions for the coarse-grained (CG) model via three stages based on six reference all-atom (AA) simulations as presented in Fig. 6.3. We use the CHARMM carbohydrates atomistic force field[231] with modifications for the dihedrals of the acetate group on the primary alcohol pulled from the CHARMM General FF (CGenFF), along with the Optimal Point Charge (OPC) 4-point water model.[92]

The first simulation is a pure water box with the average side length of 4.7 nm. For simulations 2-6 (refer to Fig. 6.3), we perform molecular dynamics simulations on cellulose oligomers, each consisting of 8 repeat units, with varying DS. During these simulations, we randomly select alcohols along the backbone of each cellulose chain to acetylate, while ensuring that the acetate groups are distributed equally among the three alcohol sites whenever applicable (simulations 3, 4, and 6). In simulations 2-5, we additionally constrain the DS of all repeating units to be the same. In contrast, in simulation 6, we consider mixed DS values (DS=0, 1, 2, or 3) for the repeating units and ensure all there is an equal number of monomers for each DS value.

We conduct reference AA simulations with the OpenMM simulation package.[109] 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 (LJPME 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 $1/(25\ dt)$. We setup the initial configurations for the CA using the CHARMM builder feature inside of VMD 1.9.3[232] (http://www.ks.uiuc.edu/Research/vmd/) and solvate the chains with water using Packmol.[185] Details of the system sizes are provided in Table G.1.

6.2.2 Bottom-up coarse-graining

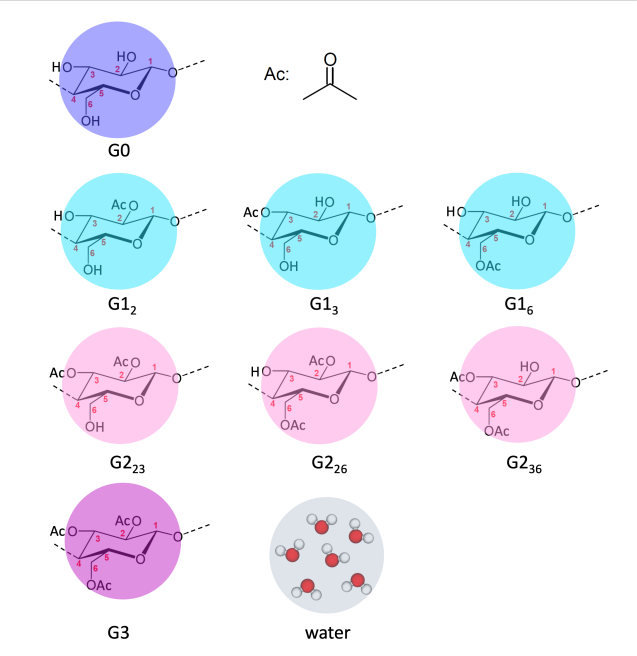

Figure 6.4: Schematic of 9 CG bead types for unsubstituted (G0), partially substituted ($G1_i$ and $G2_{ij}$), fully substituted (G3) glucose monomers, and CG water composed of, on average, 6 atomistic water molecules. Subscripts i, j denote the alcohol site that acetylation occurs; $i = 2, 3, 6$ corresponds to sites C2, C3, and C6, respectively.

We translate AA reference trajectories for coarse-graining by mapping center-of-mass coordinates of groups of atoms in the AA representation to CG sites as shown in Fig.

6.4. Specifically, we map each repeating glucose unit of cellulose into a single neutral bead. We use eight bead types to represent the different chemistries of the glucose monomers based on the DS and location of acetate group. There are three monomer types with DS = 1, $G1_2$, $G1_3$, and $G1_6$, corresponding to acetylation at C2, C3, and C6 sites, respectively. Similarly, $G2_{23}$, $G2_{26}$, and $G2_{36}$ represent monomers of DS = 2 with sites C2 and C3, C2 and C6, and C3 and C6 acetylated, respectively.

In this work, we represent several atomistic water molecules by a single neutral bead. This is inspired by the challenge of studying long length and time-scale phenomena in explicit-solvent simulations, even with the reduced resolution in coarse-grained simulations. It also aims to ensure uniform bead sizes across all CG bead types. Here, we use the k-means clustering algorithm[233–235] to identify clusters of water molecules in each reference trajectory frame by minimizing the within-cluster sum of variances of coordinates. This procedure is discussed in Appendix C. In short, we perform the clustering such that the average cluster size is 6, i.e., each CG bead of water represents, on average, 6 atomistic water molecules. We fix the Gaussian regularization range, a_α, of each bead species to approximately the cube root of its molecular volume, estimated from AA simulations. Since the molecular volume of all CG bead types are comparable, we set $a_\alpha = 0.6$ nm for all species.

We derive CG parameters in successive stages; once parameters are determined, they are fixed in subsequent steps. In the first stage, we determine the water-water repulsion v_{ww} from pure water AA simulation to reproduce the compressibility of OPC water, $\kappa_T \sim$

4.51×10^{-10} Pa^{-1}. This determines the CG pressure of $P_{CG} = 285.99\ k_BT/nm^3$ that we use in the subsequent coarse-graining stages.

In the second stage, we parameterize the excluded volume interactions between glucose monomers of same DS values (DS = 0, 1, 2, and 3) and their cross interactions with water via 4 reference simulations, each composed of uniformly substituted cellulose molecules in water (simulations 2-5 in Fig. 6.3). The parameters derived in this step are $v_{G0\,G0}$, $v_{G1_i\,G1_j}$, $v_{G2_{ij}\,G2_{kl}}$, $v_{G3\,G3}$, $v_{G0\,w}$, $v_{G1_i\,w}$, $v_{G2_{ij}\,w}$, and $v_{G3\,w}$, where the subscripts i, j, k, and l denote the alcohol sites that are acetylated ($i, j, k, l \in [2, 3, 6]$).

In stage three, we derive the remaining cross-interactions among the glucose monomers ($v_{G0\,G1_i}$, $v_{G0\,G2_{ij}}$, $v_{G0\,G3}$, $v_{G1_i\,G2_{jk}}$, $v_{G1_i\,G3}$, and $v_{G2_{ij}\,G3}$) along with one universal bond length, b, for all monomer types from a simulation at DS = 0.5 (simulation 6 in Fig. 6.3) where monomers are randomly substituted while maintaining the ratio G0 : $\sum_i$G1$_i$: $\sum_{ij}$G2$_{ij}$: G3 as 1:1:1:1 and acetate groups are distributed equally to three alcohol sites. We tabulate the parameters in Tables G.2 and G.3.

6.2.3 Coarse-grained molecular dynamics (CGMD)

CGMD simulations are conducted using a Langevin Dynamics integrator. The Langevin relaxation time τ_{CG} is taken as the unit of time. Due to the soft nature of the CG interactions and bonds, large time steps of 0.02 τ_{CG} are feasible. During the relative entropy minimization step, CG configurations are sampled from short trial CGMD which we run for 5×10^4 time steps, or 2.5×10^6 τ_{CG}. We employ a cutoff of 3 nm ($5 \times a_\alpha$) for

the non-bonded pair-wise interactions, sufficient for the excluded volume interactions to become negligible after the cutoff.

6.2.4 Phase diagram calculation with the field theory

To identify the two-phase boundary for CA aqueous mixtures, 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 Section 3.2.2.

6.3 Results and discussion

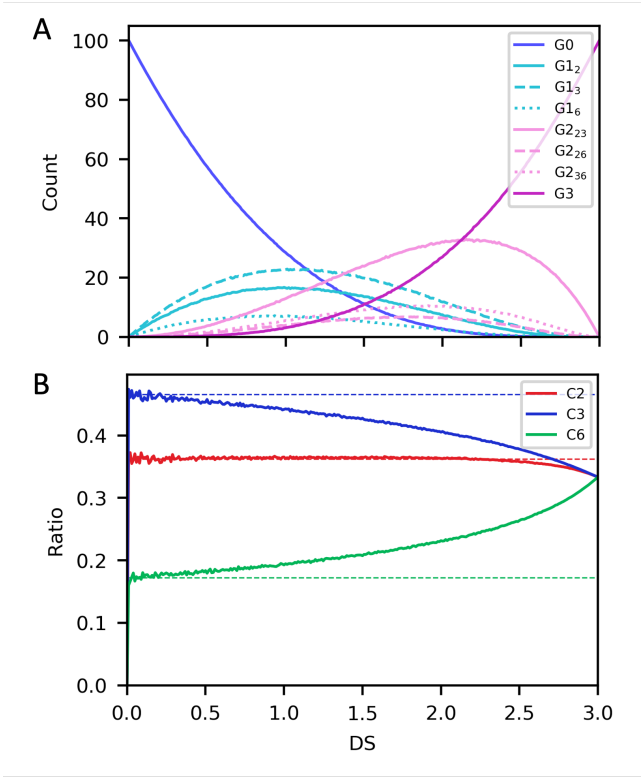

Figure 6.5: (A) Count of 8 CG bead types comprising CA at varying DS for $N = 100$ and $r_{set} = 0.36 : 0.47 : 0.17$. r_{set} represents the relative ease of substitution at the C2, C3, and C6 sites (see main text for its definition). The count is averaged over 1000 sequence generations at each DS value to obtain the most representative sequence. (B) Relative substituent ratios at C2, C3, and C6 corresponding to the representative sequences in (A). Dashed lines represent r_{set}.

At zero and full substitution (DS = 0 and 3), cellulose is composed solely of the unsubstituted and fully substituted monomers, G0 and G3, respectively. For intermediate DS values, we determine the monomer composition of cellulose by assigning acetate independently to three alcohol sites (C2, C3, and C6) based on the ratios C2:C3:C6 $\equiv r_{set}$. At a given DS value, we restrict the number of substitutions to DS $\times N$, where N is the degree of polymerization. This process is repeated 1000 fold to obtain an average monomer

composition at each DS value. An example of the resulting monomer counts averaged over the generated sequences at DS values ranging from 0 to 3 is shown in Figure 6.5A. In this specific example, we set the acetylation ratio $r_{set} = 0.36 : 0.47 : 0.17$, which aligns with the average relative substituent composition reported for water-soluble CA samples by Buchanan *et al.* (1991).[236]

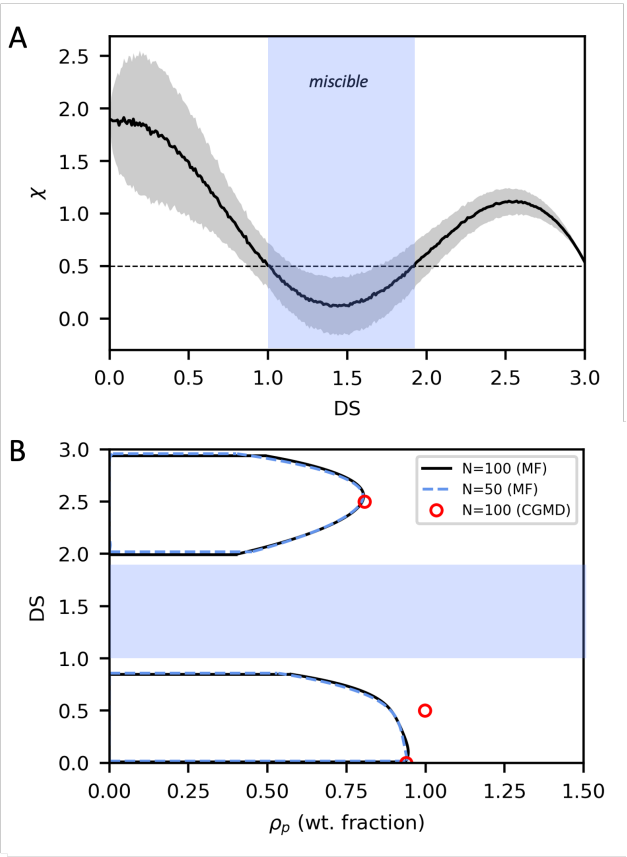

Figure 6.6: (A) Effective χ of CA as a function of DS for the representative sequences in Fig. 6.5. Shaded gray area denotes the standard deviation of χ over 1000 sequence generations. A miscibility window, denoted by shaded blue region, is predicted for intermediate DS values in which $\chi < 0.5$. This is consistent with results in (B), phase diagrams calculated using the Gibbs ensemble method for $N = 50$ and 100. Open red symbols indicate compositions of the dense phase calculated from CGMD for $N = 100$. Again, the shaded blue region is the miscible region from (A).

We note that for DS > 1.0, the design space defined by the ratio C2:C3:C6 becomes

more constrained as DS increases (illustrated by the shaded region in Fig. G.1). For instance, at DS=3, it is not feasible to have a monomer composition with all substitutions at the C2 alcohol and none at C3 or C6 alcohols (C2:C3:C6 = 1:0:0). In this case, all three sites need to be substituted, such that only the ratio C2:C3:C6 = 1:1:1 is possible. This illustrates that when we set a specific C2:C3:C6 (r_{set}) ratio during acetate assignment, we are not directly determining the resulting ratio C2:C3:C6 in the generated cellulose sequence. Consequently, by defining the ratio C2:C3:C6 during acetate assignment, we influence the probability of substitutions at the respective sites, but the actual relative substituent ratio, r_{actual}, is also determined by the design space imposed by the chosen DS value. Figure 6.5B presents the actual substituent ratio at the three alcohol sites, showing that r_{actual} is not equal to r_{set}. Instead, r_{actual} varies with DS and approaches 1:1:1 at DS=3.

We first characterize the propensity for macrophase separation as a function of DS by the effective χ parameter as defined in Eq. 5.4. The binary interaction χ approximates the overall affinity between CA in solution where a higher value of χ indicates a greater tendency towards phase separation. In our definition of χ, only the relative composition of monomer types is important and χ is independent of patterning, which means that the specific arrangement or pattern of the monomers within the cellulose sequence does not influence its value. After obtaining χ for all the generated sequences at a particular DS, we calculate the average of these values to arrive at a representative or typical χ value for the given DS. We plot the average χ as a function of DS for $r_{set} = 0.36 : 0.47 : 0.17$

in Fig. 6.6A. As a first approximation, phase separation occurs at $\chi \gtrsim 0.5$,[212] suggesting that the our model predicts CA is insoluble for DS $\lesssim 1$ and $\gtrsim 2$.

Fig. 6.6B shows the phase diagram for two different chain lengths, $N = 100$ and $N = 50$, calculated directly in the mean-field limit using the Gibbs ensemble approach based on the average monomer compositions (Fig. 6.5A) at $r_{set} = 0.36 : 0.47 : 0.17$. The miscibility window is observed at intermediate DS values from 1.0 to 2.0, in line with the predictions made by the effective χ values. In contrast to the χ analysis, the binodal calculation rigorously accounts for composition dependence which provides the specific composition and DS at which CA becomes miscible. The agreement between the binodal phase diagram and the effective χ values suggests the viability of using the approximate effective χ as a proxy to quickly determine phase separation in CA solutions. Additionally, we find that the 2-phase boundary is insensitive to the chain length N (Fig. 6.6B), indicating that the results presented here has approached the long-chain limit. This suggests that results obtained at $N = 100$ can be safely extrapolated to larger chain lengths that are used in experiments. Additionally, the close agreement between the CA composition in the dense phase as predicted by the mean-field approximation and CGMD simulations at DS values of 0, 0.5, and 2.5 further suggests that the mean-field approximation provides a reasonable description of phase separation in this system.

More importantly, we are able to qualitatively reproduce the miscibility window at intermediate DS values that have been previously observed for DS $\sim 0.3 - 1.3$ by various experimental studies[224,228,236] for CA as summarized in Fig. 6.2B. The emergence of the

miscible region with increasing DS in CA has been attributed to the disruption of the hydrogen bonding network within the cellulose matrix.[221,224] Intra-molecular hydrogen bonding contributes to the rigidity of CA, which helps to maintain cellulose's crystallinity, while inter-molecular hydrogen bonding keeps the close packing of cellulose chains. As the AA simulations suggest (discussed shortly), when DS increases, more alcohol groups are replaced by acetate groups, leading to a disruption of the hydrogen bonding network between cellulose molecules. Consequently, the unsubstituted alcohol groups become more exposed and readily form hydrogen bonds with water. When a significant portion of the alcohol groups is substituted by acetate at high DS, the chain becomes hydrophobic because acetate groups cannot form as many hydrogen bonds with water as the alcohol groups. Consequently, this leads to the reappearance of the immiscibility window at high DS values.

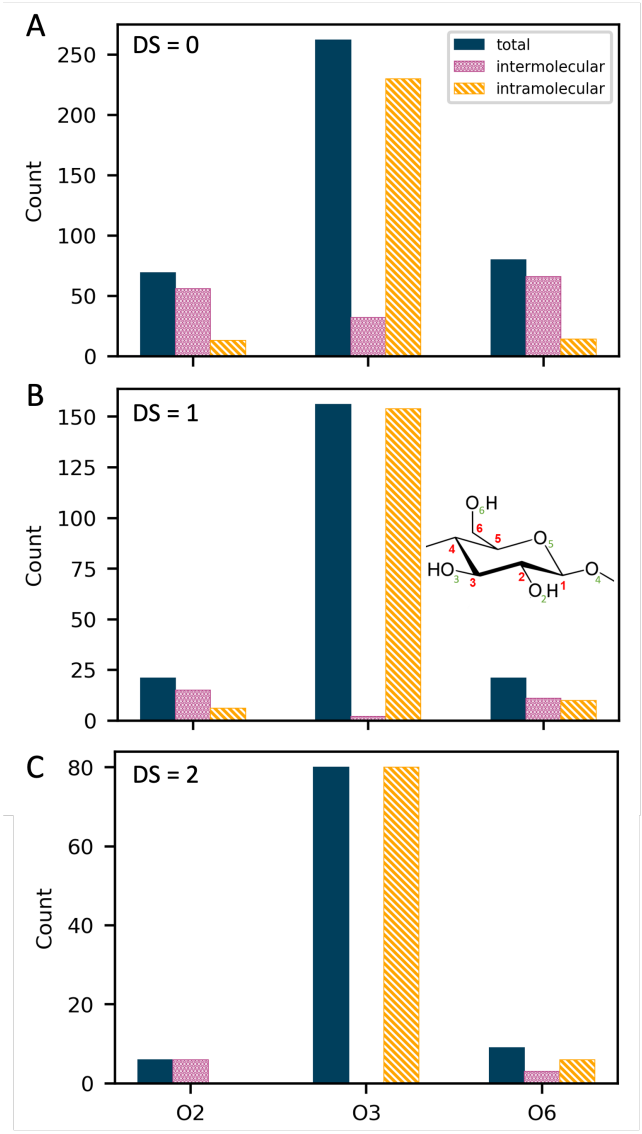

Figure 6.7: Number of total, inter-, and intra-molecular hydrogen bonds between monomers of cellulose that involve the alcohol groups at the C2, C3, and C6 sites calculated from AA simulations of 8-mers at (A) DS=0, (B) DS=1, and (C) DS=2, corresponding to simulations 2, 3, and 4 in Fig. 6.3. In each simulation, all monomers have the same DS value and acetate groups are equally distributed to all sites. Inset shows the schematic of the monomer with oxygen (green) and carbon (red) numbering.

To understand the origin of the miscibility window predicted by our model at low DS values, we compute the effective χ values for homopolymers made of each of the 8 monomer types. This provides a simple metric to assess the individual monomer's pref-

erence for hydration. Notably, Fig. G.2 illustrates that the $G1_3$ monomer has the lowest effective χ value, below 0.5, in comparison to the other monomers. This finding suggests that the emergence of miscibility, as predicted by our model at intermediate DS values, stems from the increase in hydrophilicity associated with the increase in the number of $G1_3$ monomer relative to the more hydrophobic monomers like G0 and $G1_2$. Notably, this observation aligns with experimental observations that selectively acetylating C3 results in water-soluble CA because it disrupts hydrogen bonding involving the alcohol group at the C3 position, which plays a crucial role in maintaining the crystalline structure of cellulose.[221,224]

We verify this hypothesis by calculating the number of hydrogen bonds between all monomers from reference AA simulations of 8-mers, where all monomers have the same DS value and substituents are equally distributed to all alcohol sites at DS values 0, 1, and 2 (simulations 2, 3, and 4 in Fig. 6.3, respectively). As depicted in Fig. 6.7, there are more total hydrogen bonds involving the oxygen at the C3 alcohol group (O3) compared to those involving the oxygen at the C2 and C6 alcohol groups (O2 and O6, respectively). Consequently, disrupting the hydrogen bonding at C3 through the substitution of C3 alcohol groups leads to a more significant increase in water solubility of CA. The ability of the CG model to qualitatively capture these subtle effects of substitution sites on solubility underscores the model's ability to predict nontrivial phase behavior without any experimental input.

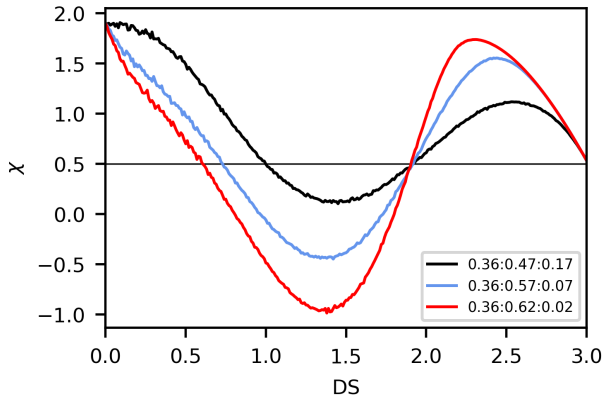

Figure 6.8: Effective χ of CA as a function of DS for sequences generated from three r_{set} values. Black series is from $r_{set} = 0.36 : 0.47 : 0.17$, same as Fig. 6.6A. Blue and red series are for sequences with a higher substitution rate at C3. Their actual relative substituent compositions are shown in Fig. G.3.

While Fig. 6.6 illustrates that the miscibility window occurs approximately between DS values of 1 and 2, which is higher than observed in experiments, Fig. 6.8 demonstrates that this miscibility window is influenced by the relative composition of substituents at three alcohol sites. By increasing the substitution rate of C3 by 15% and correspondingly reducing it at C6 (from modifying r_{set} accordingly), the lower boundary of the miscibility window shifts to 0.6.

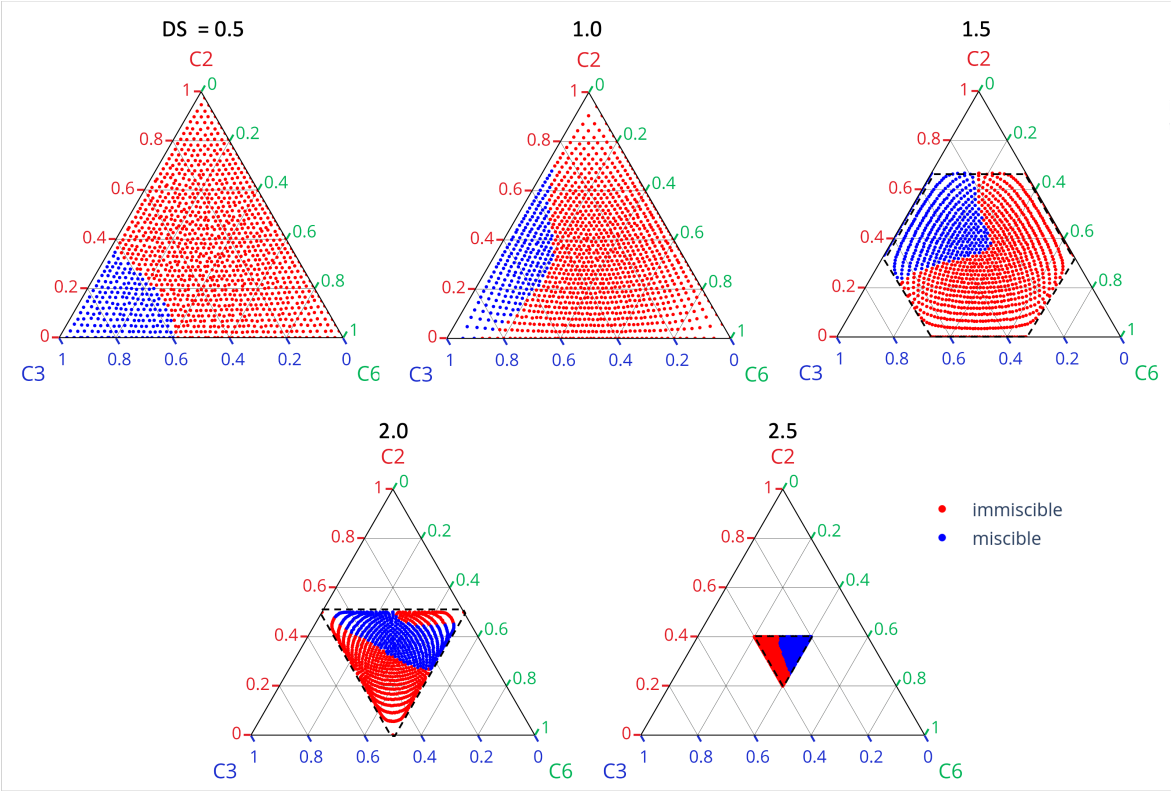

Figure 6.9: Miscibility assessed by evaluating the effective χ parameter in the acetylation composition space for DS = 0.5, 1.0, 1.5, 2.0, and 2.5. Blue cirles denote water-soluble CA with $\chi < 0.5$ while red cirles correspond to water-insoluble CA with $\chi \geq 0.5$. The black dashed line indicates the outer boundary of possible compositions for DS values larger than 1.

Using the developed model we further probe substituent compositions to develop design principles for engineering water-soluble CA. Specifically, at a fixed DS value, we calculate the effective χ for sequences spanning all possible values of C2:C3:C6 by varying r_{set} as described earlier. Similar to the data presented in Fig. 6.6A, we obtain the average χ and actual relative acetylation ratios, r_{actual}, over 1000 sequence generations at each r_{set} value. As we have previously established a strong correlation between effective χ values and miscibility, we use χ as a metric to assess miscibility: $\chi < 0.5$ corresponds to miscible samples while $\chi \geq 0.5$ corresponds to immiscible ones. Based on the results

depicted in Fig. 6.9, which shows the miscibility of CA in the acetylation composition plane (r_{actual}), water-soluble CA is achieved when the majority of substitutions occur at C3 sites for DS values below approximately 0.5. At intermediate DS values, around 1.0-1.5, samples with substituents evenly distributed at C2 and C3 sites exhibit water solubility. This is consistent with the experimental study by Miyamoto that lowering the relative substitution at C6 via altering the synthesis route results in water-soluble CA.[224] Finally, at high DS values, water-soluble CA can be obtained when the substituents are distributed relatively evenly among all three sites, with a slightly higher propensity for C2 and C6 sites.

In a more detailed analysis of the hydrogen bonding interactions presented in Fig. 6.7, we find that most of the hydrogen bonds involving O3 are intra-molecular with the heterocyclic oxygen (O5) (Fig. G.4). We hypothesize that substituting alcohol groups at the C3 site can have a detrimental effect on backbone rigidity, leading to increased flexibility. Consequently, the remaining alcohol groups become more exposed to water, supporting our earlier observation that enhanced substitution at C3 sites results in water-soluble CA at low DS values. On the other hand, at high DS values, the backbone of CA is already sufficiently flexible, and removing additional intra-molecular hydrogen bonds between O3 and O5 has minimal impact on the miscibility. Instead, inter-molecular hydrogen bonding involving O2 and O6 becomes more crucial in maintaining the cohesion between cellulose chains. Consequently, substitutions at C2 and C6 sites become necessary to disrupt the cellulose network at higher DS values. This analysis

provides a proposed mechanism that elucidates the role of specific hydrogen bonding interactions and how they contribute to the solubility behavior of CA at varying DS values. By understanding the impact of these interactions, one can strategically choose a synthesis route that selectively acetylates either the alcohols at C2, C3, or C6 sites to achieve the desired water solubility at a given DS.

Lastly, one inconsistency we observe in the current model is that CA becomes miscible at DS=3, indicated by the disappearance of the two-phase region. In fact, DS=3 is right on the cusp of immisicibility with a $\chi \sim 0.5$ (see Fig. 6.6)A. Importantly, we also observe this weakening of immiscibility with increasing acetylation in the AA simulations, as evidenced by the decrease in the peak intensity of the radial distribution function (RDF) between inter-molecular C1 carbons as DS approaching 3 (Fig. G.5). Thus, CA also becomes more hydrated at DS=3 in the AA simulations. The fact that the CG model correctly captures the AA model's solution phase behavior at DS=3 indicates that the model performs well in this regard. Nevertheless, the discrepancy with experimental findings might be attributed to the crystallization of CA, a phenomenon not captured in AAMD due to the long timescales and the relatively small molecular weight of the CA used in our study. Consequently, the model inadequately considers the free energy gain from crystallization, which can drive phase separation at high DS values. While acknowledging this limitation's impact on DS $\rightarrow$ 3 and planning to address it in future work, our focus has been on addressing CA miscibility within the intermediate DS range, which is more relevant in formulation design.

6.4 Conclusions

In this study, we developed a molecularly informed field-theoretic model for cellulose acetate (CA), an important ester derivative of cellulose, and investigated its phase behavior in aqueous solutions. Our coarse-grained (CG) model represents CA with 8 monomer types, categorized by the degree-of-substitution (DS) and the specific acetylation sites. Particularly, our model predicts a miscibility window at intermediate DS values (approximately 0.6-2), which qualitatively aligns with experimental findings. This is especially remarkable since we are able to capture the nontrivial phase behavior, where CA solutions become insoluble $\rightarrow$ soluble $\rightarrow$ insoluble with increasing DS, all achieved without relying on experimental input.

While experimental studies have a wide range of DS values over which CA is water-soluble, we are able to use a molecularly informed field theory to directly probe CA solution behavior as a function of the relative DS at the three alcohol sites, C2, C3, and C6. Selective acetylation of C3 alcohols leads to water-soluble CA at low DS values (DS $\lesssim 0.5$) because C3 alcohols participate in intra-molecular hydrogen bonding with heterocyclic oxygens, crucial for maintaining cellulose's rigidity and well-packed crystalline structure. By acetylating C3 sites, chain flexibility increases at low DS, exposing the remaining alcohol groups and allowing them to form hydrogen bonds with water.

At intermediate DS values (approximately 1-1.5), water-soluble CA is achieved when C2 and C3 alcohols are selectively acetylated. However, at high DS values (DS $\gtrsim 2$), we predict water-soluble CA with evenly distributed acetate, having a slightly higher

propensity for C2 and C6 sites. We hypothesize that the shift in the relative DS for water solubility from higher substitution at C3 to C2 and C6 with increasing DS is due to the fact that the CA backbone is already sufficiently flexible at higher DS, minimizing the impact of removing additional intra-molecular hydrogen bonds involving C3 alcohols. Instead, inter-molecular hydrogen bonding involving C2 and C6 alcohols becomes crucial in maintaining the cohesion between cellulose chains, making substitutions at C2 and C6 sites necessary to disrupt the cellulose network at higher DS values.

Overall, our model provides insights and proposes a mechanism for CA's solubility at different DS values. This understanding can guide the design and synthesis of water-soluble CA with tailored properties for a wide range of applications. While in this work we have focused solely on cellulose and its acetate derivative, this methodology is generally applicable to other carbohydrates and their derivatives in a wide variety of solvents. We believe there is great promise in coupling this level of modeling with wet-lab experimentation to aid in the engineering of next-generation polymeric materials with tailored properties, such as solubility.

Acknowledgments This work was supported by BASF Corporation through the California Research Alliance. G.H.F. derived partial support from the National Science Foundation CMMT Program under grant number DMR-2104255. M.S.S. acknowledges funding support from the National Science Foundation through Award No. CHEM-1800344. Use was made of computational facilities purchased with funds from the National Science Foundation (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; NSF DMR 2308708) at UC Santa Barbara.

Supporting Information see Appendix G

Chapter 7

Conclusions and Outlook

In this thesis, we present a novel framework for constructing a predictive multi-scale simulation workflow to investigate the solution phase behavior of soft matter formulations. By parameterizing large-scales field-theoretic models based on small-scale atomistic simulations, we overcome the long-standing challenges in studying complexation driven by electrostatics with either (1) traditional particle-explicit models that struggle to sample meaningful conformations or (2) phenomenological field theories that lack chemical specificity. We address the first limitation by the leveraging the computational efficiency of the field theory in simulating dense systems composed of high-molecular-weight molecules. In addition, the direct access to free energies and chemical potentials in the field theory allow us to rigorously determine the thermodynamics that govern the macroscopic properties. Chemical specificity is achieved through a bottom-up coarse-graining technique, minimizing relative entropy to derive chemically sensitive coarse-grained interaction parameters,

eliminating the need for experimental input and enabling high-throughput screening of unexplored parameter space. We summarize the procedure to construct such molecularly informed field theory model in Chapter 2. Overall, the molecularly informed field theory offers a promising approach for understanding complex formulations in a computationally efficient and chemically informed manner.

In Chapter 3, we present phase diagrams for a well-characterized model coacervate system consisting of poly(acrylic acid) (PAA) and poly(allylamine hydrochloride) (PAH) in a sodium chloride (NaCl) aqueous solution that capture the effects of charged group stoichiometry, electrostatic screening of salt, and chemistry-specific effects due to polyelectrolyte molecular details (PAA-PAH asymmetries) on coacervation. Coacervation propensity decreases as the charged group stoichiometry deviates from stoichiometric conditions, resulting in a more dilute coacervate phase and a smaller two-phase region. The two-phase region shrinks at higher salt concentrations, and the model correctly predicts that small ions preferentially partition in the dilute phase. This field-theoretic model of polyelectrolyte coacervate captures key qualitative trends relevant to formulation design, which are often challenging to observe using particle-based approaches.

Building on top of the simplified coacervate model, in Chapter 4, we investigate a more realistic system involving 5 components: cationic polyelectrolyte polydiallyldimethylammonium (PDADMA), anionic surfactant sodium dodecyl sulfate (SDS), nonionic ethoxylated surfactants (C_mEO_n), NaCl salt, and water. The molecularly informed field theory allows us to systematically explore the self-assembly behavior of these multi-component

systems and investigate how equilibrium self-assemblies of mixed micelles change under various solution mixing conditions including the ratio of SDS to C_mEO_n, charge stoichiometry between polyelectrolyte and micelle, salt concentration, and the overall concentration of solids. Moreover, we establish a correlation between the degree of polyelectrolyte adsorption and the onset of polyelectrolyte-micelle complexation, as characterized by the solution turbidity observed in experiments. Remarkably, our model correctly predicts that nonionic ethoxylated C_mEO_n surfactants with shorter hydrophilic groups undergo complexation transitions at lower mixing ratios of anionic to nonionic surfactants, showing semi-quantitative agreement with experimental observations.

In Chapter 5, we utilized the molecularly informed field-theoretic approach to calculate the critical micelle concentration (CMC) of intrinsically disordered protein (IDP) surfactants. Our results showcase the predictive power of this approach, yielding a CMC value within a factor of 2 of experimental data for strongly micellizing systems with CMCs in the μM range, a challenging regime for traditional particle-based methods. The success of our approach in capturing the self-assembly behavior of IDP surfactants highlights its ability to handle the complex interactions arising from diverse amino acids in bio-based macromolecules. This study underscores the potential of the molecularly informed field-theoretic workflow in designing novel sustainable materials, thereby minimizing environmental impacts through the replacement of petroleum-based ingredients.

Cellulose acetate (CA) is another promising sustainable ingredient as alternative to synthetic polymers in formulations. It has been established in the literature that water

solubility is achieved for cellulose when the alcohols are substituted with acetate. In Chapter 6, we investigate the factors contributing to CA's water solubility. Remarkably, our molecularly informed field-theoretic model predicts a miscible window at intermediate degree-of-substitution (DS), in qualitative agreement with experimental observations without any experimental input. We attribute CA's solution behavior to the relative DS at three alcohol sites, C2, C3, and C6. Selective acetylation of C3 alcohols leads to water-soluble CA at low DS values (DS $\lesssim 0.5$), while at intermediate DS values ($\sim$ 1-1.5), water solubility is achieved when C2 and C3 alcohols are selectively acetylated. At high DS values (DS $\gtrsim 2$), water-soluble CA with evenly distributed acetate, having a slightly higher propensity for C2 and C6 sites, is predicted. Our model provides valuable insights into CA's solubility behavior, enabling the design and synthesis of tailored water-soluble CA for various applications.

In summary, we have presented a comprehensive investigation into the solution phase behavior of various soft matter formulations with increasing complexity. Chapters 3, 4, 5, and 6 demonstrate the application of molecularly informed field theories to study polyelectrolyte coacervation, polyelectrolyte-micelle complexation, self-assembly of IDP surfactants, and cellulose acetate miscibility, respectively. Notably, our predictions are in good agreement with experimental observations without any experimental input. The success of this workflow highlights its potential in designing sustainable materials, reducing reliance on petroleum-based ingredients in formulations, and optimizing soft matter systems for various applications.

Future work can focus on several potential directions to enhance the accuracy and applicability of the molecularly informed field-theoretic approach. Firstly, refining the atomistic force field used in the parameterization process can help resolve discrepancies between the model predictions and experimental data. Additionally, increasing the complexity of the coarse-grained model by introducing higher-resolution coarse-grained species and incorporating dipole solvents such that the electrostatic screening depends on the local environment can provide more accurate representations of complex mixtures. Moreover, incorporating field fluctuations in the field theory via complex Langevin samplings[75,77] can further enhance the model's accuracy, enabling better predictions of thermodynamic properties and self-assembly behaviors. As the workflow is highly adaptable to different chemistries, future research can explore more complex systems, encompassing a broader range of soft matter formulations. Overall, the work we present here opens up exciting avenues for the development and application of the molecularly informed field-theoretic workflow in *de novo* soft materials design.

Appendix A

Analytical Field Theory for Multi-Species Systems

In the following, we derive the analytical approximation for the field theory of a system consisting of S bead species and M molecule types. Each molecule type m has n_m copies, each has the chain length of N_m. A molecule m consists of $\rho_{\alpha,m}V/n_m$ number of any bead species α that carries charge σ_α in the unit of the electronic charge e. The total bead concentration of species α is $\rho_\alpha = \sum_m^M \rho_{\alpha,m}$. The interaction potentials for the bonded and non-bonded terms are

$$\beta U_0(\boldsymbol{r}) = \sum_m^M \sum_{i=1}^{N_m} \frac{3}{2b_m^2} \left|\boldsymbol{r}_{m,i} - \boldsymbol{r}_{m,i-1}\right|^2 \tag{A.1}$$

$$\beta U_1(\boldsymbol{r}) = \int d\boldsymbol{r} \int d\boldsymbol{r}' \; \hat{\boldsymbol{\rho}}(\boldsymbol{r}) \cdot \beta \boldsymbol{U}(|\boldsymbol{r} - \boldsymbol{r}'|) \cdot \hat{\boldsymbol{\rho}}(\boldsymbol{r}') \tag{A.2}$$

where the microscopic particle number density, $\hat{\boldsymbol{\rho}}(\boldsymbol{r})$, is a vector of length S with its element $(\hat{\boldsymbol{\rho}}(\boldsymbol{r}))_\alpha = \hat{\rho}_\alpha(\boldsymbol{r}) = \sum_m^M \sum_{i=1}^{N_m} \delta(\boldsymbol{r} - \boldsymbol{r}_{m,i})\delta_{\psi_{i,m},\alpha}$. This is equivalent to counting all beads i on all chains m that are of species α, i.e., $\psi_{i,m} = \alpha$, at position $\boldsymbol{r}$. $\beta\boldsymbol{U}$ is a SxS pair interaction matrix and can be a sum of the excluded volume term and the electrostatic term

$$(\beta\boldsymbol{U}_{ev})_{\alpha\gamma}(r) = \frac{u_{\alpha\gamma}}{(2\pi(a_\alpha^2 + a_\gamma^2))^{3/2}} e^{-r^2/2(a_\alpha^2+a_\gamma^2)} \tag{A.3}$$

$$(\beta\boldsymbol{U}_{el})_{\alpha\gamma}(r) = \frac{l_b\sigma_\alpha\sigma_\gamma}{r} \operatorname{erf}\left(\frac{r}{2\sqrt{a_\alpha^2/2 + a_\gamma^2/2}}\right) \tag{A.4}$$

Here, we include both terms while deriving the mean-field solutions, $\beta\boldsymbol{U} = \beta\boldsymbol{U}_{ev} + \beta\boldsymbol{U}_{el}$, but only consider the electrostatic contribution in the Gaussian approximation about the mean-field solutions, $\beta\boldsymbol{U} = \beta\boldsymbol{U}_{el}$.

Upon the particle-to-field transformation described in[74], we introduce the chemical potential fields, $\boldsymbol{w} = [w_1, w_2, ..., w_S]$, in addition to the bead density fields, $\boldsymbol{\rho} = [\rho_1, \rho_2, ..., \rho_S]$, and the canonical partition function of this model can be expressed as a complex-valued statistical field theory

$$\mathcal{Z} = \mathcal{Z}_0 \int \mathcal{D}\boldsymbol{\rho} \int \mathcal{D}\boldsymbol{w}\; e^{-H[\boldsymbol{\rho},\boldsymbol{w}]} \tag{A.5}$$

where $\mathcal{Z}_0 = e^{\beta U_{SI}} \prod_m^M \frac{1}{n_m!} \left(\frac{Z_{0,m}}{\lambda_T^{3N_m}}\right)^{n_m}$ is the partition function for an ideal gas of discrete Gaussian chains, and $Z_{0,m}$ is the path integral of molecule m in zero field. The first term in $\mathcal{Z}_0$ subtracts the self-interactions present in $\beta\boldsymbol{U}$,

$$\beta U_{SI} = \frac{\beta}{2} \sum_{\alpha}^{S} \rho_\alpha V \left(\frac{u_{\alpha\alpha}}{(4\pi a_\alpha^2)^{3/2}} + \frac{l_b}{\sqrt{\pi} a_\alpha} \right) \tag{A.6}$$

Lastly, the effective Hamiltonian in Eq. A.5 is

$$\begin{aligned} H[\boldsymbol{\rho}, \boldsymbol{w}] = & -i \int d\boldsymbol{r} \; (\boldsymbol{w} \cdot \boldsymbol{\rho}) \\ & + \frac{1}{2} \int d\boldsymbol{r} \int d\boldsymbol{r}' \; \boldsymbol{\rho}(\boldsymbol{r}) \cdot \beta \boldsymbol{U}(|\boldsymbol{r} - \boldsymbol{r}'|) \cdot \boldsymbol{\rho}(\boldsymbol{r}') \\ & - \sum_{m}^{M} n_m \ln Q_m[i\boldsymbol{w}] \end{aligned} \tag{A.7}$$

A.1 Mean-field approximation

The mean-field (MF) approximation is the most important analytical approximation technique that assumes a single field configuration dominates the functional integral in Eq. A.5, i.e., $\mathcal{Z} = \mathcal{Z}_0 e^{-H[\boldsymbol{\rho}^*, \boldsymbol{w}^*]}$. The MF solution of field f is obtained by solving for $\frac{\delta H[f]}{\delta f(\boldsymbol{r})}_{f^*} = 0$.

$$\begin{aligned} \frac{\delta H}{\delta w_\alpha(\boldsymbol{r})}_{\boldsymbol{\rho}^*, \boldsymbol{w}^*} &= -i\rho_\alpha^* - \sum_{m|\alpha \in m}^{M} n_m \frac{\delta \ln Q[i\boldsymbol{w}]}{\delta w_\alpha} \\ &= -i\rho_\alpha^* + i \sum_{m|\alpha \in m}^{M} \rho_{\alpha,m} = 0 \\ \rho_\alpha^* &= \sum_{m|\alpha \in m}^{M} \rho_{\alpha,m} = \rho_\alpha \end{aligned} \tag{A.8}$$

$$\begin{aligned}
\frac{\delta H}{\delta \rho_\alpha(\boldsymbol{r})}_{\boldsymbol{\rho}^*,\boldsymbol{w}^*} &= -iw_\alpha^* + \frac{1}{2}\int d\boldsymbol{r}\ \sum_\gamma^S \rho_\gamma^*(\beta \boldsymbol{U})_{\alpha\gamma} \\
&= -iw_\alpha^* + \frac{1}{2}\sum_\gamma^S \int d\boldsymbol{r}\ \rho_\gamma^*(\beta \boldsymbol{U}_{ev})_{\alpha\gamma} + \frac{1}{2}\sum_\gamma^S \int d\boldsymbol{r}\ \rho_\gamma^*(\beta \boldsymbol{U}_{el})_{\alpha\gamma} \\
&= -iw_\alpha^* + \sum_\gamma^S u_{\alpha\gamma}\rho_\gamma = 0 \\
w_\alpha^* &= -i\sum_\gamma^S u_{\alpha\gamma}\rho_\gamma
\end{aligned} \tag{A.9}$$

Here, the $\int d\boldsymbol{r}$ integrals are reduced to a simple summation by recognizing that the excluded volume term is a Gaussian integral and the electrostatic term is zero due to the charge neutrality constraint. Finally, the last term in Eq. A.7 involving the MF single chain partition functions is

$$-\sum_m^M n_m \ln Q_m[i\boldsymbol{w}^*] = iV\boldsymbol{w}^* \cdot \boldsymbol{\rho}^*$$

The MF Hamiltonian is

$$\begin{aligned}
H[\boldsymbol{\rho}^*, \boldsymbol{w}^*] &= -iV\boldsymbol{w}^* \cdot \boldsymbol{\rho}^* \\
&\quad + \int d\boldsymbol{r} \int d\boldsymbol{r}' \boldsymbol{\rho}^* \cdot \beta \boldsymbol{U} \cdot \boldsymbol{\rho}^* \\
&\quad - \sum_m^M n_m \ln Q_m[i\boldsymbol{w}^*] \\
&= \frac{V}{2}\boldsymbol{\rho}^* \cdot \boldsymbol{u} \cdot \boldsymbol{\rho}^*
\end{aligned} \tag{A.10}$$

where $\boldsymbol{u}$ is a SxS matrix of the excluded volume parameters defined in Eq. A.3.

A.2 Gaussian approximation

The Gaussian approximation is derived by applying small perturbations to each fields about their MF solutions, i.e., $\boldsymbol{w} = \boldsymbol{w}^* + \boldsymbol{\delta w}$, $\boldsymbol{\rho} = \boldsymbol{\rho}^* + \boldsymbol{\delta \rho}$. The effective Hamiltonian of Eq. A.7 is re-written with perturbation terms up to second order

$$\begin{aligned} H = &-iV\boldsymbol{w}^* \cdot \boldsymbol{\rho}^* - i\int d\boldsymbol{r}\ \boldsymbol{\delta\rho}\cdot\boldsymbol{\delta w} + \frac{V}{2}\boldsymbol{\rho}^*\cdot\boldsymbol{u}\cdot\boldsymbol{\rho}^* \\ &+ \frac{1}{2}\int d\boldsymbol{r}\int d\boldsymbol{r}'\ \boldsymbol{\delta\rho}(\boldsymbol{r})\cdot\beta\boldsymbol{U}\cdot\boldsymbol{\delta\rho}(\boldsymbol{r}') - \sum_m^M n_m \ln Q_m[i\boldsymbol{w}] \end{aligned} \tag{A.11}$$

We follow the derivation in ref. 74 to get the Gaussian approximation of the single chain structure factor,

$$Q_m = e^{-i\sum_{\alpha|\alpha\in m}\frac{\rho_{\alpha,m}V}{n_m}w^*_\alpha}\left(1 - \frac{1}{2V^2}\sum_{l=1}^{N_m}\sum_{j=1}^{N_m}\sum_k \delta\hat{w}_{\psi_{l,m}}(k)\delta\hat{w}_{\psi_{j,m}}(-k)e^{-\frac{k^2b_m^2|l-j|}{6}}\right) \tag{A.12}$$

where the $\delta\hat{w}$ is the Fourier transforn of δw field, b_m is the root-mean-squared bond length of a homogeneous polymer m and the factor $\rho_{\alpha,m}V/n_m$ is the number of bead α on one molecule m.

We introduce the intramolecular structure factor matrix

$$\hat{\mathbf{G}}(k) = \sum_m^M \hat{\mathbf{G}}_m(k) \tag{A.13}$$

where $\hat{\mathbf{G}}_m$ is the form factor of molecule type m and its element is

$$\left(\hat{\mathbf{G}}_m(k)\right)_{\alpha\gamma} = \frac{n_m}{V}\sum_l^{N_m}\sum_j^{N_m} e^{-\frac{k^2b_m^2}{6}|i-j|}\delta_{\psi_{l,m},\alpha}\delta_{\psi_{j,m},\gamma} \tag{A.14}$$

With these definitions, the $-\sum_m^M n_m \ln Q_m$ term reduces to

$$
\begin{aligned}
-\sum_{m}^{M} n_m \ln Q_m &= iV\boldsymbol{w}^* \cdot \boldsymbol{\rho}^* + \frac{1}{2V}\sum_{k} \boldsymbol{\delta\hat{w}}(k) \cdot \hat{\mathbf{G}} \cdot \boldsymbol{\delta\hat{w}}(-k) \\
&= iV\boldsymbol{w}^* \cdot \boldsymbol{\rho}^* + \frac{1}{2}\int d\mathbf{r} \int d\mathbf{r}' \delta\boldsymbol{w}(\mathbf{r}) \cdot \mathbf{G} \cdot \delta\boldsymbol{w}(\mathbf{r}')
\end{aligned}
\tag{A.15}
$$

and the Hamiltonian in Eq. A.11 becomes

$$
\begin{aligned}
H = &\frac{V}{2}\boldsymbol{\rho}^* \cdot \boldsymbol{u} \cdot \boldsymbol{\rho}^* - i\int d\boldsymbol{r}\ \boldsymbol{\delta\rho} \cdot \boldsymbol{\delta w} \\
&+ \frac{1}{2}\int d\boldsymbol{r} \int d\boldsymbol{r}'\ \boldsymbol{\delta\rho}(\boldsymbol{r}) \cdot \beta\boldsymbol{U} \cdot \boldsymbol{\delta\rho}(\boldsymbol{r}') \\
&+ \frac{1}{2}\int d\mathbf{r} \int d\mathbf{r}' \delta\boldsymbol{w}(\mathbf{r}) \cdot \mathbf{G} \cdot \delta\boldsymbol{w}(\mathbf{r}')
\end{aligned}
\tag{A.16}
$$

where the first term is just the MF Hamiltonian defined in Eq. A.10. The partition function is

$$
\begin{aligned}
\mathcal{Z} = &\mathcal{Z}_0\, e^{-H[\boldsymbol{\rho}^*, \boldsymbol{w}^*]} \\
&\int \mathcal{D}\boldsymbol{\delta\rho} \exp\left(-\frac{1}{2}\int d\boldsymbol{r} \int d\boldsymbol{r}'\ \boldsymbol{\delta\rho}(\boldsymbol{r}) \cdot \beta\boldsymbol{U} \cdot \boldsymbol{\delta\rho}(\boldsymbol{r}')\right) \\
&\int \mathcal{D}\boldsymbol{\delta w}\ \exp\left(i\int d\boldsymbol{r}\ \boldsymbol{\delta\rho} \cdot \boldsymbol{\delta w} - \frac{1}{2}\int d\mathbf{r} \int d\mathbf{r}' \delta\boldsymbol{w}(\mathbf{r}) \cdot \mathbf{G} \cdot \delta\boldsymbol{w}(\mathbf{r}')\right) \\
&= \mathcal{Z}_0\, e^{-H[\boldsymbol{\rho}^*, \boldsymbol{w}^*]} \\
&\int \mathcal{D}\boldsymbol{\delta\rho} \exp\left(-\frac{1}{2}\int d\boldsymbol{r} \int d\boldsymbol{r}'\ \boldsymbol{\delta\rho}(\boldsymbol{r}) \cdot (\beta\boldsymbol{U} + \boldsymbol{G}^{-1}) \cdot \boldsymbol{\delta\rho}(\boldsymbol{r}')\right) \\
&\int \mathcal{D}\boldsymbol{\delta w}\ \exp\left(-\frac{1}{2}\int d\mathbf{r} \int d\mathbf{r}' \delta\boldsymbol{w}(\mathbf{r}) \cdot \mathbf{G} \cdot \delta\boldsymbol{w}(\mathbf{r}')\right)
\end{aligned}
\tag{A.17}
$$

where in the last line we have evaluated $\int \mathcal{D}\boldsymbol{\delta w}$ per the Gaussian integral formula. Lastly, we continue simplify H by re-expressing and evaluating the Gaussian integrals in k-space,

$$
\begin{aligned}
\mathcal{Z} &= \mathcal{Z}_0\, e^{-H[\boldsymbol{\rho}^*,\boldsymbol{w}^*]} \\
&\prod_k \int d\boldsymbol{\delta\hat{\rho}}_k \exp\left(-\frac{1}{2}\boldsymbol{\delta\hat{\rho}}(k)\cdot\frac{\beta}{V}\left(\hat{\boldsymbol{U}}+\hat{\boldsymbol{G}}^{-1}\right)\cdot\boldsymbol{\delta\hat{\rho}}(k)\right) \\
&\prod_k \int d\boldsymbol{\delta\hat{w}}_k \exp\left(-\frac{1}{2}\boldsymbol{\delta\hat{w}}(k)\cdot\frac{1}{V}\hat{\boldsymbol{G}}\cdot\boldsymbol{\delta\hat{w}}(k)\right) \\
&= \mathcal{Z}_0\, e^{-H[\boldsymbol{\rho}^*,\boldsymbol{w}^*]}\prod_k \frac{(2\pi V)^S}{\left|\beta\hat{\boldsymbol{U}}+\hat{\boldsymbol{G}}^{-1}\right|^{1/2}\left|\hat{\boldsymbol{G}}\right|^{1/2}} \\
&= \mathcal{Z}_0\, e^{-H[\boldsymbol{\rho}^*,\boldsymbol{w}^*]}\prod_k \frac{(2\pi V)^S}{\left|\beta\hat{\boldsymbol{U}}(k)\cdot\hat{\boldsymbol{G}}(k)+\boldsymbol{I}\right|^{1/2}} \\
&= \mathcal{Z}_0\, e^{-H[\boldsymbol{\rho}^*,\boldsymbol{w}^*]}\prod_k \frac{(2\pi V)^S}{\left|\beta\hat{\boldsymbol{U}}_{el}(k)\cdot\hat{\boldsymbol{G}}(k)+\boldsymbol{I}\right|^{1/2}}
\end{aligned}
\tag{A.18}
$$

This is the partition function that is used to derive the free energy, pressure and chemical potentials.

A.3 Thermodynamics from the Gaussian approximation

The Helmholtz free energy is

$$
\begin{aligned}
\beta F &= -\ln\mathcal{Z} \\
&= -\beta U_{SI} + \sum_m^M\left(-n_m\ln\left(\frac{Z_{0,m}}{V\lambda_T^{3N_m}}\right)+n_m\ln\frac{n_m}{V}-n_m\right)+\frac{V}{2}\boldsymbol{\rho}^*\cdot\boldsymbol{u}\cdot\boldsymbol{\rho}^* \\
&+\frac{V}{4\pi^2}\int dk\; k^2\ln\left(\left|\beta\hat{\boldsymbol{U}}_{el}(k)\cdot\hat{\boldsymbol{G}}(k)+\boldsymbol{I}\right|\right)
\end{aligned}
\tag{A.19}
$$

For phase diagram calculation in the Gibbs ensemble, the first two terms can be neglected. The pressure and chemical potential are derived from the following free energy

expression,

$$\beta F' = \sum_{m}^{M} \left(n_m \ln \frac{n_m}{V} - n_m \right) + \frac{V}{2} \boldsymbol{\rho}^* \cdot \boldsymbol{u} \cdot \boldsymbol{\rho}^* + \frac{V}{4\pi^2} \int dk\ k^2 \ln \left(\left| \beta \hat{\boldsymbol{U}}_{el}(k) \cdot \hat{\boldsymbol{G}}(k) + \boldsymbol{I} \right| \right) \tag{A.20}$$

The pressure of the system and chemical potential of chain m are

$$\begin{aligned} \beta P &= -\left(\frac{\partial \beta F'}{\partial V} \right)_{T, n_m} \\ &= \sum_{m}^{M} \frac{n_m}{V} + \frac{1}{2} \boldsymbol{\rho}^* \cdot \boldsymbol{u} \cdot \boldsymbol{\rho}^* \\ &+ \frac{V}{4\pi^2} \int dk\ k^2 \left(\mathrm{Tr} \left[(\beta \hat{\boldsymbol{U}}_{el}(k) \cdot \hat{\boldsymbol{G}}(k) + \boldsymbol{I})^{-1} (\beta \hat{\boldsymbol{U}}_{el}(k) \cdot \hat{\boldsymbol{G}}(k)) \right] + \ln \left| (\beta \hat{\boldsymbol{U}}_{el}(k) \cdot \hat{\boldsymbol{G}}(k) + \boldsymbol{I} \right| \right) \end{aligned} \tag{A.21}$$

$$\begin{aligned} \beta \mu_m &= \left(\frac{\partial \beta F'}{\partial n_m} \right)_{V, T, n_{n \neq m}} \\ &= \ln \frac{n_m}{V} + \sum_{\alpha \in m, \gamma \notin m}^{S} \frac{\rho_{\alpha,m} V}{n_m} \rho^*_\gamma u_{\alpha\gamma} + \frac{1}{2} \sum_{\alpha \in m, \gamma \in m} \left(\frac{\rho_{\alpha,m} V}{n_m} \rho^*_\gamma + \frac{\rho_{\gamma,m} V}{n_m} \rho^*_\alpha \right) u_{\alpha\gamma} \\ &+ \frac{N_m}{4\pi^2} \int dk\ k^2\, \mathrm{Tr} \left[(\beta \hat{\boldsymbol{U}}_{el}(k) \cdot \hat{\boldsymbol{G}}(k) + \boldsymbol{I})^{-1} \frac{\beta \hat{\boldsymbol{U}}_{el}(k) \cdot \hat{\boldsymbol{G}}_m(k)}{n_m N_m / V} \right] \end{aligned} \tag{A.22}$$

Appendix B

Coarse-Graining NaCl from the External Potential Ensemble

In a recent publication,[100] we proposed an approach to enhance the fidelity of a coarse-grained (CG) model by employing a more intelligent choice of simulation ensemble. Our findings revealed that coarse-graining in the external potential ensemble offers improved thermodynamic accuracy and transferability of CG models compared to the commonly used single, uniform-composition ensembles. As explained in our previous work, the coarse-graining is performed at a state of inhomogeneous response in the composition due to an applied spatially-varying external potential on different species. The resulting CG model is then optimized to reproduce the same response as the all-atom (AA) reference system while capturing the dependence on the locally varying composition. Here, we adopt the same strategy to assess the quality of a CG model for an ionic NaCl salt

solution.

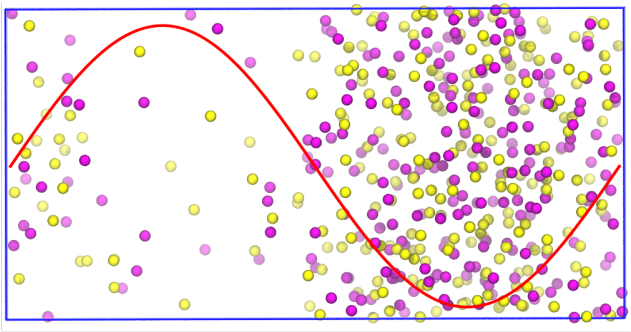

Figure B.1: Simulation snapshot with the sinusoidal external potential (red line) applied on Na^+ (pink) and Cl^- (yellow) ions along the longest box dimension. The ions partition into the region where the potential amplitude is lower, corresponding to smaller difference in the free energy. Water molecules are not shown for clarity.

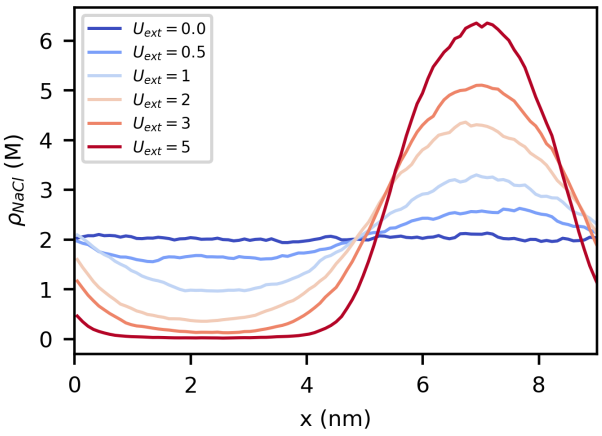

Figure B.2: One-dimensional density profiles of NaCl in the reference simulations at varying external potential amplitude, U_{ext}, from 0 to 5 k_BT. $U_{ext} = 0\ k_BT$ is equivalent to simulating in the NVT ensemble without the applied external potential.

In reference simulations of 2 M NaCl (246 NaCl molecules, and 6432 water molecules), we apply an sinusoidal external potential, $\beta U_{\sin}(x) = \beta U_{ext} \sin(2\pi x/L_x)$, on both Na^+ and Cl^- ion species at varied amplitudes U_{ext} $(0 - 5k_BT)$. L_x is the box dimension along which the potential is applied. We first perform a NPT (1 atm, 298.15 K) simulation without the external potential for 40 ns and record the average box dimensions: 4.65

nm × 4.65 nm × 9.30 nm. We then conduct the final NVT simulations initiating from the last frame of the preceding NPT run and apply the external potentials on Na^+ and Cl^- ions along the longest box dimension. After a 2 ns warm-up period, we run for an additional 40 ns and collect the trajectories for coarse-graining. The response to the applied external potential and the density profile are visualized in Fig. B.1 and Fig. B.2, respectively.

We investigate the thermodynamic fidelity of the CG model of NaCl solution by comparing the predicted mean ionic activity coefficient curve to that of the AA model. We follow the strategy of Mester et al. to calculate the activity coefficient in the AA model: we insert a pair of non-interacting Na^+ and Cl^- ions then slowly turn on the van der Waals and Coulombic potentials to determine the excess chemical potential.[237] The mean ionic activity coefficient, γ, is given by

$$\ln \gamma(m) = \beta \, \frac{\mu_{ig} + \mu_{ex} - \mu_0}{2} - \ln \; m \tag{B.1}$$

where μ_{ig} and μ_{ex} are the ideal gas and excess parts of the chemical potential, μ_0 is the chemical potential at infinite dilution, and m is the molality of NaCl. For the CG model, we have direct access to the chemical potential from conducting field-theoretic simulations on its field representation, discussed in section 2.3. The activity coefficient curve, Fig. B.3, from the $U_{ext} = 2k_BT$ best matches that of the AA system. The trace of the Fisher information matrix (the Hessian of S_{rel} with respect to interaction parameters) has a maximum near $U_{ext} = 2k_BT$ and is consistent with our earlier findings that maximizing the Fisher information improves the thermodynamics of CG models.[100]

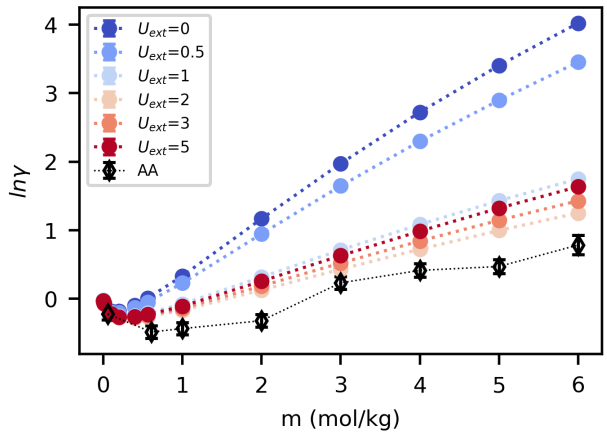

Figure B.3: Natural logarithm of the mean ionic activity coefficient $\ln\gamma$ as a function of the molality m for different U_{ext} ensembles. $U_{ext} = 2k_BT$ ensemble shows the best match to the AA model's activity curve. Dotted lines are a guide to the eye.

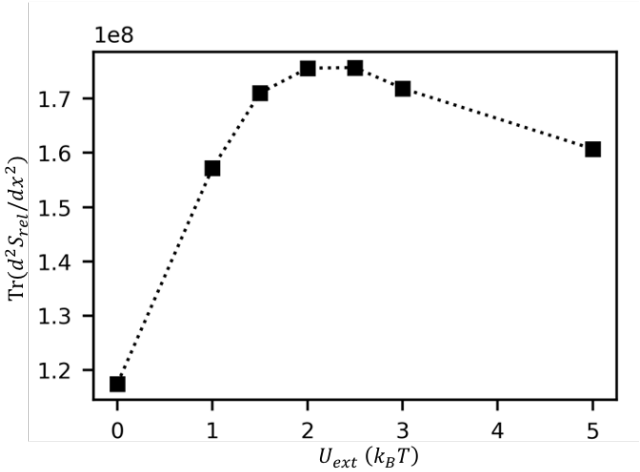

Figure B.4: The trace of the Hessian of S_{rel} with respect to interaction parameters plotted against the external potential amplitude. The maximum is found at $U_{ext} = 2k_BT$

Tables B.1 and B.2 summarize the CG parameters for NaCl in water obtained minimizing the relative entropy in the $U_{ext} = 2k_BT$ ensemble, while keeping the like-interaction of water fixed.

Table B.1: Optimized parameters derived from S_{rel} minimization

Parameter	Value	
v_{w,Na^+}	0.0	k_BT
v_{w,Cl^-}	0.51301	k_BT
v_{Na^+,Na^+}	0.59897	k_BT
v_{Cl^-,Cl^-}	1.44760	k_BT
v_{Na^+,Cl^-}	0.0	k_BT

Table B.2: Fixed parameters

Parameter	Value	
$v_{w,w}$	0.33897	k_BT
a_w	0.31	nm
a_{Na^+}	0.31	nm
a_{Cl^-}	0.31	nm
σ_w	0	e
σ_{Na^+}	+1	e
σ_{Cl^-}	-1	e

Appendix C

Coarse-Grained Models of Water

Despite the significant improvement in computational efficiency due to the reduced resolution in CG models, studying long length and time-scale phenomena in explicit-solvent simulations is still challenging. It is beneficial to develop computationally efficient models of CG solvents, specifically water, that can be utilized in studying complex formulations. In bottom-up coarse-graining approaches, reference trajectories used in deriving CG parameters are obtained by mapping each water molecule to a CG site. For multiple water models, one straight forward strategy is mapping a cluster of more than one water molecules into one CG bead, with the bead's coordinate determined by the center-of-mass of the cluster. Here, we use the k-means clustering algorithm[233–235] to identify clusters of water molecules in each trajectory frame by minimizing the within-cluster sum of variances of coordinates. While the number of molecules in each cluster cannot be directly specified, the total number of clusters k, or the total number of CG water molecules,

determines the average number of water molecules in each cluster.

Prior to clustering, we need to transform the center-of-mass Cartesian coordinates of individual water molecules into new coordinates that account for periodic boundary conditions. For a Cartesian coordinate $\boldsymbol{r} = (x, y, z)$ and a simulation box of lengths $\boldsymbol{L} = (L_x, L_y, L_z)$ in 3D, the corresponding transformed coordinates are:

$$\boldsymbol{\theta} = \cos \frac{2\pi \boldsymbol{r}}{\boldsymbol{L}} \tag{C.1}$$

$$\boldsymbol{\phi} = \sin \frac{2\pi \boldsymbol{r}}{\boldsymbol{L}} \tag{C.2}$$

Upon the transformation, each Cartesian coordinate, e.g. x, is represented by two new coordinates θ and ϕ. This allows molecules that are in close in distance but located across a periodic boundary to be included in the same cluster, as illustrated in Fig. C.1 in 2D.

We perform separate k-means clustering that groups, on average, 2, 3, and 6 water molecules. This results in mapped trajectories that are used as references to derived the CG parameters for the 2-, 3-, and 6-water models, respectively. The Scikit-learn library[238] is employed to carry out the clustering process. We set the maximum number of iterations to 500 and repeat the k-means algorithm with different centroid seeds 20 times. We ensure that these parameters are sufficiently large by monitoring the convergence of the average number of water molecules per cluster and the average cluster sphericity index (Fig. C.2). The latter is a metric used to evaluate the shape of a cluster of molecules and is calculated as follows[239]:

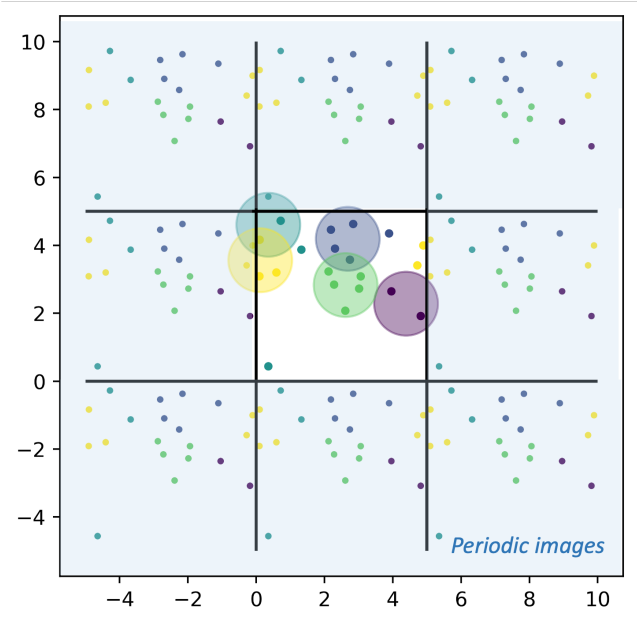

Figure C.1: Example of k-means clustering in 2D that groups neighboring molecules located across a periodic boundary. The shaded regions depict periodic boundary images. Molecules are represented by filled dots and color-coded according to their clusters. The transparent circles represent CG sites with coordinates determined by the center-of-mass of the member molecules.

$$\Psi = \left(\frac{c^2}{ab}\right)^{1/3} \tag{C.3}$$

Here, the sphericity index is calculated based on the lengths of the three representative axes of an object, denoted as a, b, and c, arranged from the longest to the shortest axis. The sphericity index ranges from 0 to 1, with 1 indicating a perfect sphere. In our case, we represent the shape of a cluster by the minimum volume enclosing ellipsoid of the molecules belonging to the cluster.

The parameters for the excluded volume interaction (Eq. 2.7) are derived following

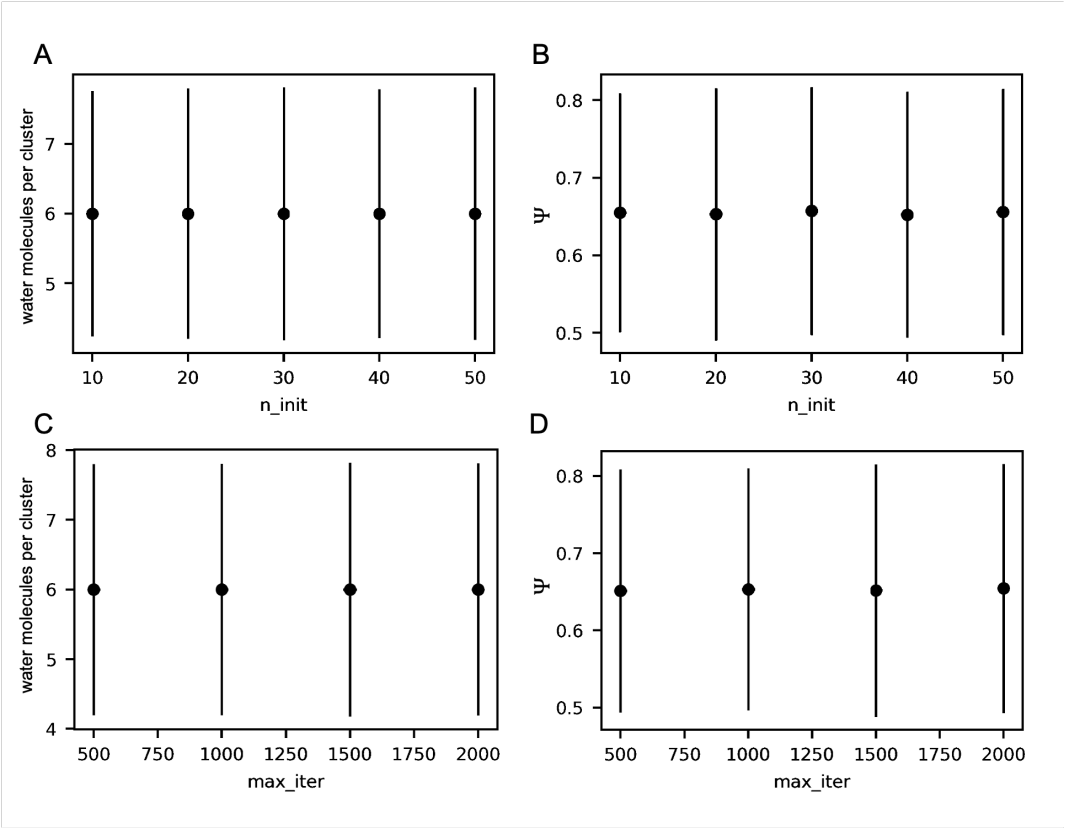

Figure C.2: Average number of water molecules per cluster and cluster sphericity index, Ψ, with (A, B) varying number of repetition (n_init) at the maximum number of iterations (max_iter) of 500 and (C, D) varying maximum number of iterations at the number of repetition of 20 for the 6-water model. Solid bars represent the standard deviations.

the same procedure in Chapter 3. We enforce the interaction radius to approximately the cube-root of the specific volume of n water molecules for n-water model. The water-water interaction strength, $v_{w,w}$, is determined by minimizing the relative entropy in the NPT ensemble such that the CG model has a compressibility $\kappa_T \approx 4.51 \times 10^{-10}\,\mathrm{Pa}^{-1}$, near that of OPC water. Matching κ_T between AA and CG simulations uniquely determines a CG pressure of $P_{CG} \approx 285.99\ k_B T/\,\mathrm{nm}^3$. Table C.1 summarizes the parameters for different water models, including the conventional 1-water model.

Table C.1: Parameters for different coarse-grained water models

Model	a_w (nm)	$v_{w,w}$ (k_BT)
1-water	0.31	0.33897
2-water	0.40	0.67589
4-water	0.50	1.42035
6-water	0.60	1.88049

Appendix D

Supplemental Information for Chapter 3

D.1 All-atom simulations

D.1.1 Forcefield validation

We show the solution densities for one component PE aqueous solutions at full ionization and zero added salt across PE composition. Densities of PAA and PAH aqueous solutions are used as a metric to quantify the reliability of the AA forcefield. We use $N = 45$ for PAA and $N = 187$ for PAH to match the average chain lengths used in experiments. Experimental densities for PAA ($M_w = 4.31$ kg/mol, $M_n = 2.4$ kg/mol) are provided by BASF. PAH densities are collected in the Helgeson lab with an Anton Paar DMA 4101 density meter, using PAH ($M_w = 17.5$ kg/mol) purchased from Sigma-Aldrich without

further purification. The simulation box is at least 7 nm and we conduct all simulations for at least 30 ns (preceded by a 10 ns equilibration period) in the NPT ensemble at 1 atm and 298.15 K using the Monte Carlo barostat with the update frequency of 25 timesteps. We note that solution density has a short correlation time ($\sim$ 10 ps) and 30 ns is more than sufficient to sample uncorrelated states.

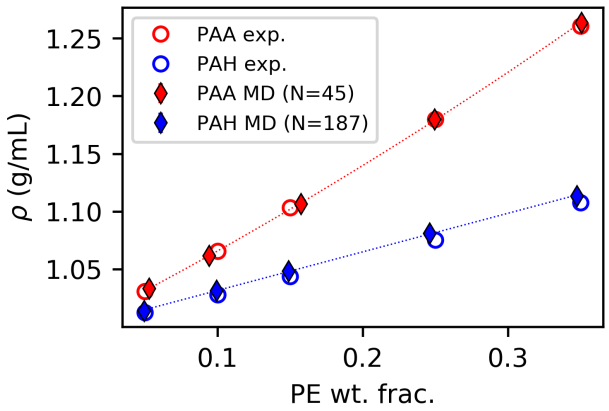

Figure D.1: Comparison between solution densities from experiments (open circles) and simulations (closed diamonds) for fully ionized PAA and PAH solutions at 1 atm and 298.15 K for various PE weight fractions (counterion included). The degrees of polymerization are 45 for PAA and 187 for PAH in the simulations. Dotted lines are a guide to the eye.

D.1.2 All-atom reference simulations

Pure water

The pure water simulation box is composed of 3305 water molecules. We perform the simulation in the NPT ensemble at 1 atm and 298.15 K with a 4 ns equilibration followed by a 40 ns production run. The average box side length is 4.7 nm.

NaCl solution

See Appendix B for more details.

Polyelectrolytes-NaCl mixture

The AA reference system consists of 10 of each 24-mer PE species, 325 NaCl molecules, and 12500 water molecules in a cubic box. We build atactic PEs of random stereochemistries (Fig. D.2) with target dyad fractions of 0.44 meso and 0.56 racemic. The resulting average meso fractions are 0.42 for PAA and 0.46 for PAH. After the energy minimization step, we relax the initial configuration at 400 K for 4 ns and cool down to 298.15 K for 10 ns. Lastly, we collect the trajectory for coarse-graining in the last 400 ns. We perform all simulation steps in the NPT ensemble at 1 atm. The average box side length from the production run is $\sim$ 7.46 nm.

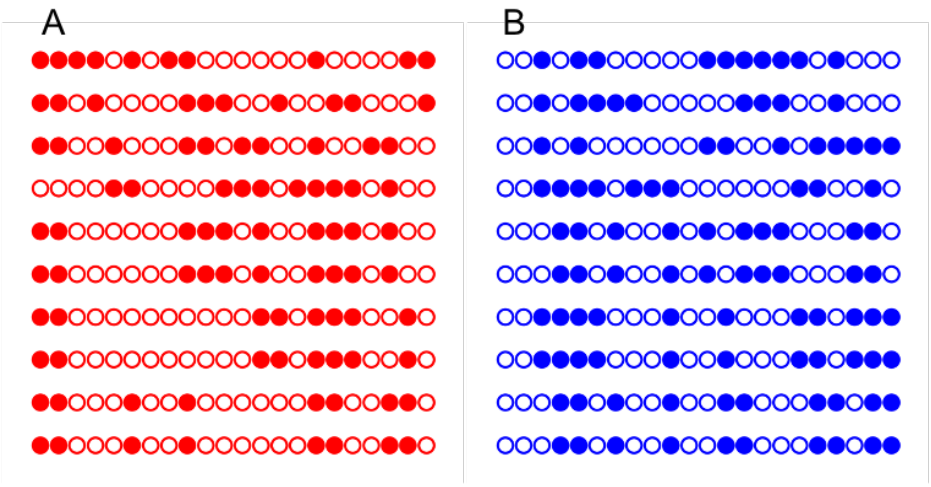

Figure D.2: Dyad compositions of (A) PAA and (B) PAH chains in the reference simulation. A 24-mer has 23 chiral centers which results in a total of 22 dyads per chain. Full symbols indicate meso dyads while open symbols are racemic dyads.

D.2 Coarse-grained parameters

Table D.1: Fixed parameters

Parameter	Value	
a_w	0.31	nm
a_{Na^+}	0.31	nm
a_{Cl^-}	0.31	nm
a_{p-}	0.45	nm
a_{p+}	0.45	nm
σ_w	0	e
σ_{Na^+}	+1	e
σ_{Cl^-}	-1	e
σ_{p-}	-1	e
σ_{p+}	+1	e

Table D.2: Optimized parameters derived from S_{rel} minimization

Parameter	Value	
$v_{w,w}$	0.33897	k_BT
v_{w,Na^+}	0.0	k_BT
v_{w,Cl^-}	0.51301	k_BT
v_{Na^+,Na^+}	0.59897	k_BT
v_{Cl^-,Cl^-}	1.44760	k_BT
v_{Na^+,Cl^-}	0.0	k_BT
$v_{p-,p-}$	0.89475	k_BT
$v_{p-,p+}$	0.68023	k_BT
v_{p-,Na^+}	0.0	k_BT
v_{p-,Cl^-}	0.86294	k_BT
$v_{p-,w}$	0.40998	k_BT
$v_{p+,p+}$	0.87466	k_BT
v_{p+,Na^+}	0.15057	k_BT
v_{p+,Cl^-}	0.64160	k_BT
$v_{p+,w}$	0.47948	k_BT
b_{p-}	0.36434	nm
b_{p+}	0.35817	nm

D.3 Additional figures from the Gaussian approximation

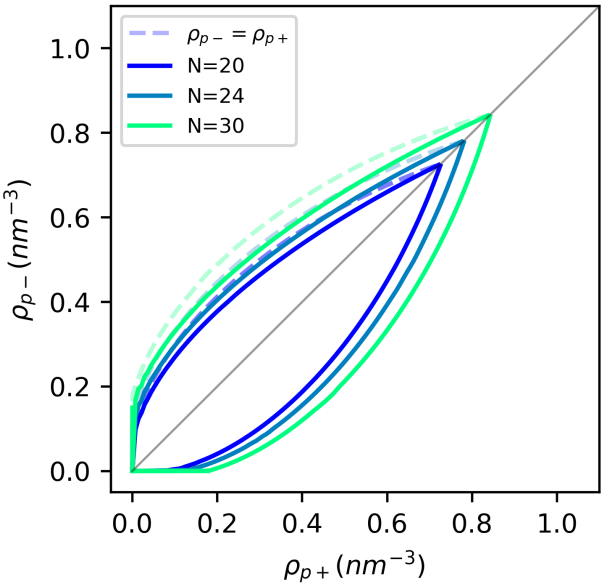

Figure D.3: Phase diagram for salt-free conditions at varying chain length, N. Faded dashed lines are hypothetical binodals for excess PAA conditions if the phase diagram is symmetric ($\rho_{p-} = \rho_{p+}$).

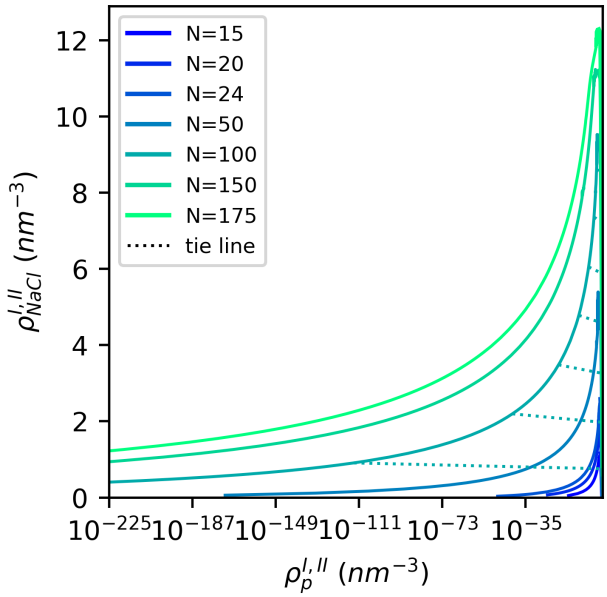

Figure D.4: Binodals at $f_{p-} = 0.5$ for varying N plotted in the log scale for the polymer concentration. Dotted lines are example tie lines for $N = 100$.

We calculate the complexation free energy, ΔF, defined as the difference between the free energy of the final phase separated system F_f, and of the initial PE mixtures (aqueous PAA and aqueous PAH) before mixing, F_i. For each set of coexisting phases found from the Gibbs ensemble simulation, we pair it with an initial state consisting of an aqueous PAA and aqueous PAH solutions. Although there are many routes to get to the same final coacervate mixture, we choose the initial PE mixtures to each have the same PE, salt, and water concentrations as the overall composition used in determining the coexisting phases, Fig. D.5A. With this choice, the initial volumes, V_{p-} and V_{p+}, are related to the overall stoichiometry by $V_{p-}/(V_{p-} + V_{p+}) = f_{p-}$. As expected, the complexation free energy becomes less negative with added salt and finally becomes zero, Fig. D.5B. This indicates the reduced coacervation driving forces as salt concentration increases.

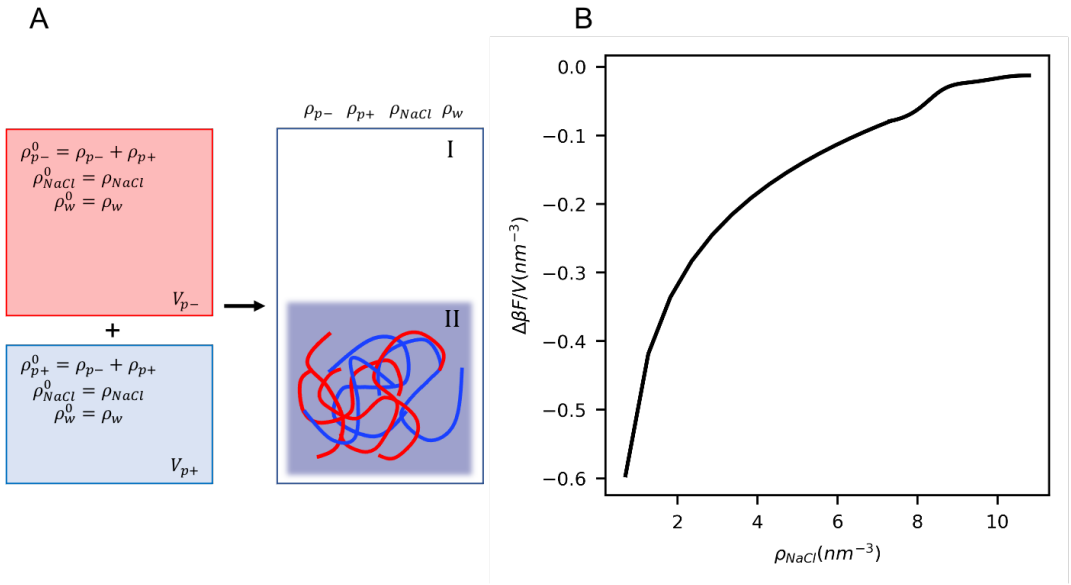

Figure D.5: (A) Schematic showing an experiment of mixing two aqueous PE solutions at the same PE, salt and water densities, resulting in a phase separated system. The initial volumes, V_{p-} and V_{p+}, are related to the overall stoichiometry in the final solution by $V_{p-}/(V_{p-}+V_{p+}) = f_{p-}$. (B) Complexation free energy density with added salt. This value becomes less negative with salt concentration, indicating the reduced driving force to coacervate.

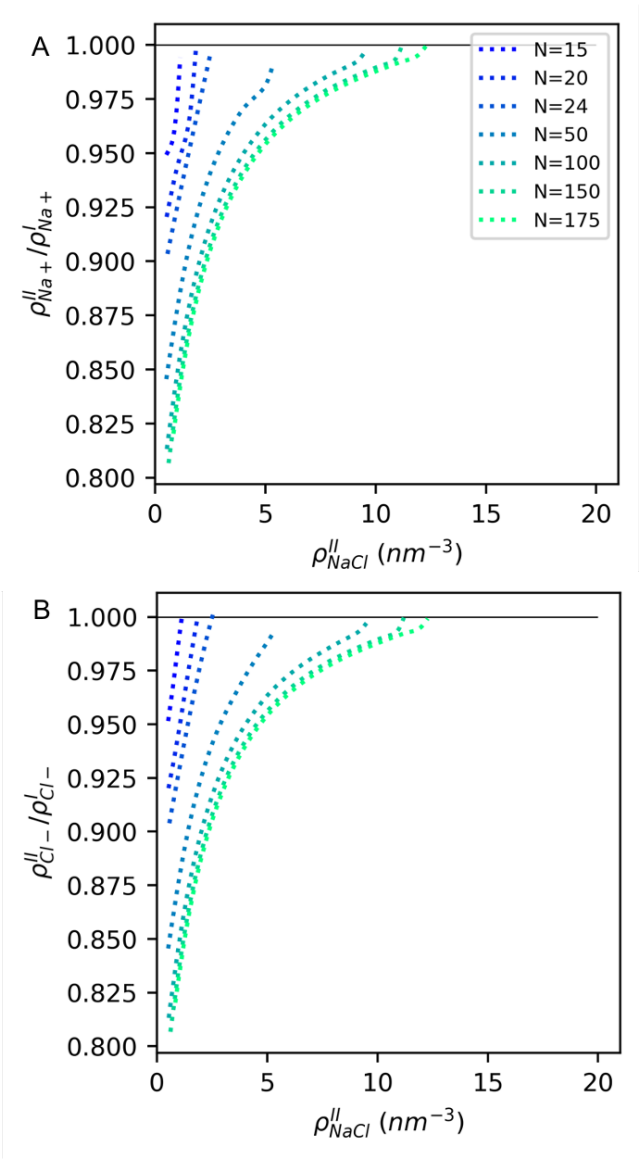

Figure D.6: (A) Na^+ and (B) Cl^- concentration in the coacervate phase relative to the dilute phase for different N at $f_{p-} = 0.5$. Both the anions and cations preferentially partition in the dilute phase.

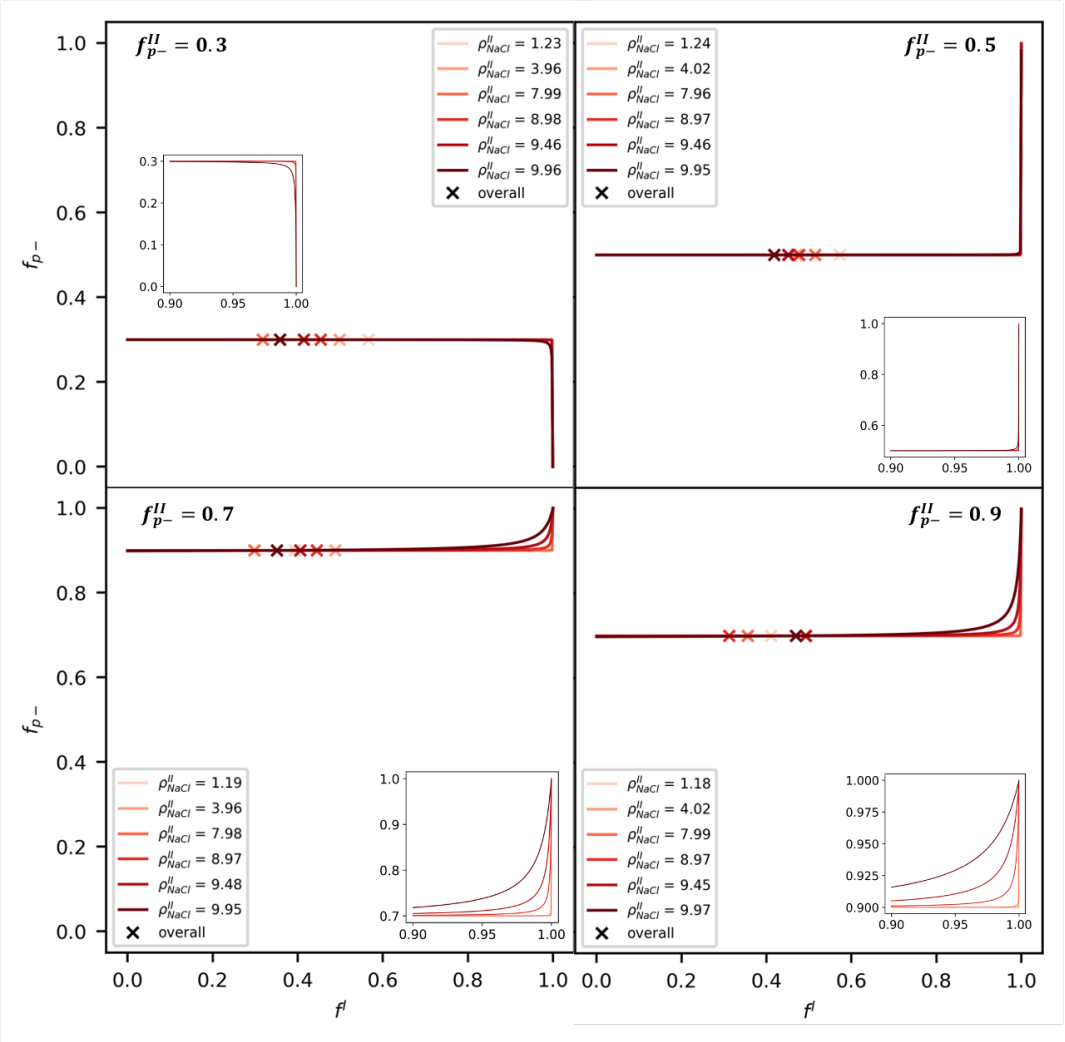

Figure D.7: The overall PE stoichiometry f_{p-} as marching along the tie line by varying the dilute phase volume fraction f^I for coacervate stoichiometries f^{II}_{p-} of 0.3, 0.5, 0.7, and 0.9. Different series represent tie lines corresponding to different excess salt concentrations in the coacervate phase. Cross symbols are the overall compositions used in Gibbs ensemble calculations.

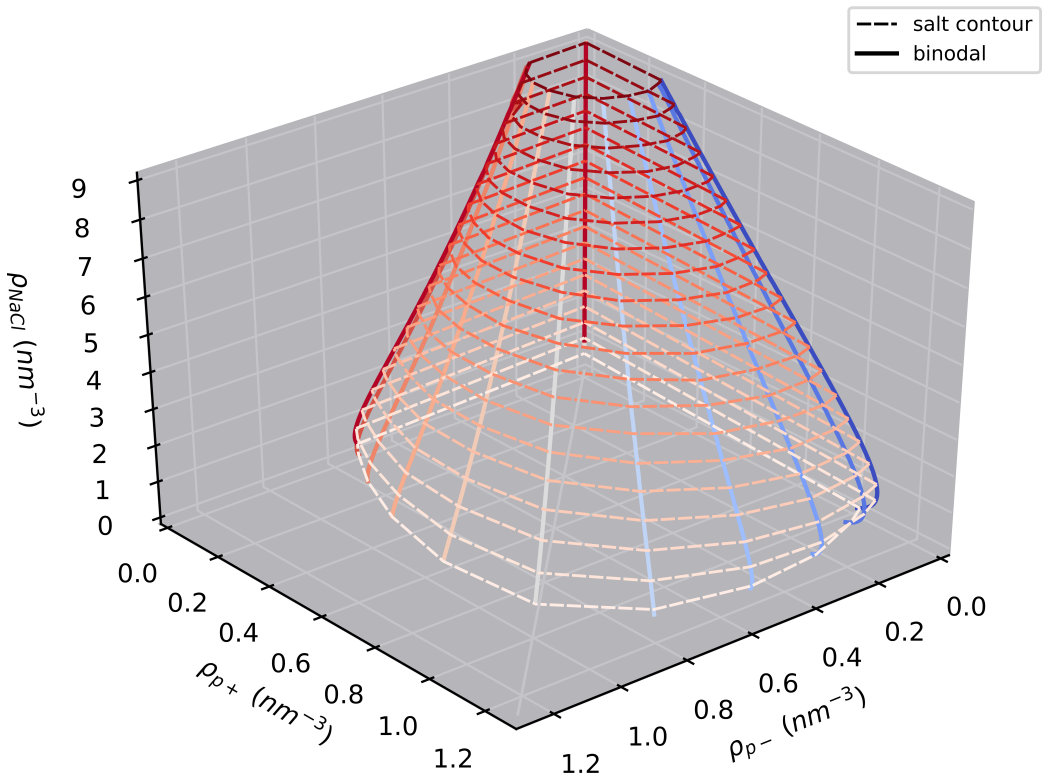

Figure D.8: Three-dimensional plot of the binodals (solid lines and color coded based on the PE stoichiometry in the coacervate, f_{p-}^{II}) for $N = 150$. Horizontal dashed lines are contours at various excess salt concentrations. Note that these are not tie lines and simply connect points of the same excess salt concentration.

Appendix E

Supplemental Information for

Chapter 4

Table E.1: Reference all-atom system sizes. Each row represents a simulation shown in Fig. 6.3.

	n_w	n_{Na^+}	n_{Cl-}	n_{DS}	n_{PDADMA}^{6mer}	$n_{C_{13}EO_7}$	n_{PEO}^{20mer}	$n_{dodecane}$	box(nm)	ensemble
1	3305								$4.7 \times 4.7 \times 4.7$	NPT
2							140		$5.7 \times 5.7 \times 5.7$	NPT
3			5880		980				$11.4 \times 11.4 \times 11.4$	NPT
4	3300	96		96				179	$4.5 \times 4.5 \times 10.2$	NVT
5	6432	245	245						$4.7 \times 4.7 \times 9.3$	NVT
6	14500					90			$8.0 \times 8.0 \times 8.0$	NPT
7	15000			126	21				$8.1 \times 8.1 \times 8.1$	NPT
8	15000	280	280	126	21				$8.1 \times 8.1 \times 8.1$	NPT
9	31500	345	345	42	7	137			$10.3 \times 10.3 \times 10.3$	NPT

Table E.2: Optimized excluded volume interaction parameters, $v_{\alpha,\gamma}$ (k_BT)

α \ γ	w	Na^+	Cl^-	SO_4	C_2	p^+	COH	EO
w	0.33897							
Na^+	0.0	0.61704						
Cl^-	0.52264	0.0	1.47764					
SO_4	0.26396	0.0	0.0	2.55051				
C_2	0.99291	0.65940	3.20991	1.51615	1.67953			
p^+	0.47581	0.14099	0.36890	0.0	1.02659	1.70031		
COH	0.47273	0.0	1.31378	0.64582	1.02802	0.39340	0.0	
EO	0.56090	0.0	0.81420	0.46855	1.26004	0.94159	0.95735	1.07767

Table E.3: Remaining coarse-grained interaction parameters: optimized root-mean-square bond length $b_{\alpha\gamma}$ and fixed parameters which include charge σ_α and excluded volume interaction range a_α.

Parameter	Value	
$b_{C_2\ C_2}$	0.30157	nm
$b_{C_2\ SO_4^-}$	0.15422	nm
$b_{p^+\ p^+}$	0.57983	nm
$b_{C_2\ EO}$	0.24657	nm
$b_{EO\ COH}$	0.26315	nm
$b_{EO\ EO}$	0.31829	nm
σ_w	0.0	e
σ_{Na^+}	+1.0	e
σ_{Cl^-}	-1.0	e
$\sigma_{SO_4^-}$	-1.0	e
σ_{C_2}	0.0	e
σ_{p^+}	+1.0	e
σ_{COH}	0.0	e
σ_{EO}	0.0	e
a_w	0.31074	nm
a_{Na^+}	0.31074	nm
a_{Cl^-}	0.31074	nm
$a_{SO_4^-}$	0.31074	nm
a_{C_2}	0.31074	nm
a_{p^+}	0.63	nm
a_{COH}	0.31074	nm
a_{EO}	0.375	nm

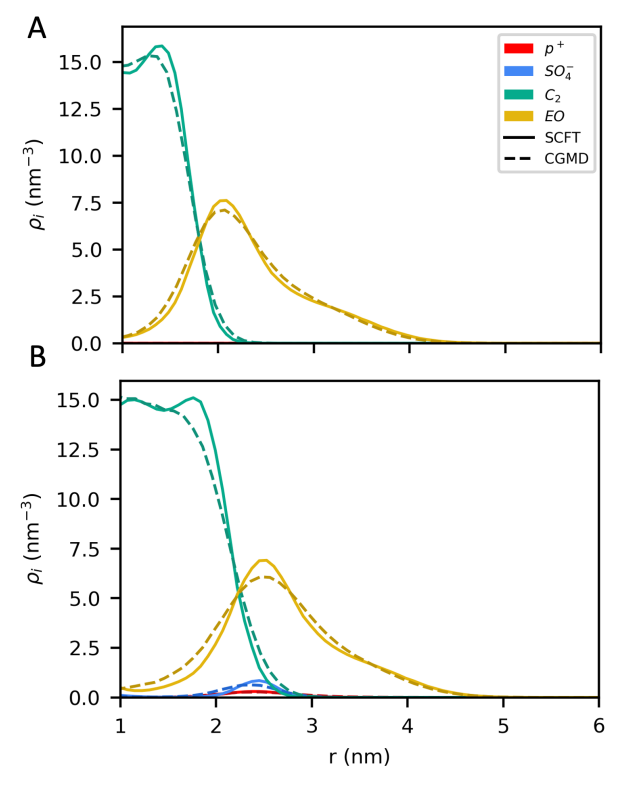

Figure E.1: Comparison of micelle density profiles between CGMD and SCFT for $SDS/C_{13}EO_{11}$ micelle and 24-mer PDADMA in 0.4 M NaCl at (A) Y = 0 (no SDS) and (B) Y = 0.31, respectively. Since it is difficult to obtain equilibrium micelle sizes from CGMD, we use the cell size and molecule numbers from equilibrated field theory to run its CGMD counterpart.

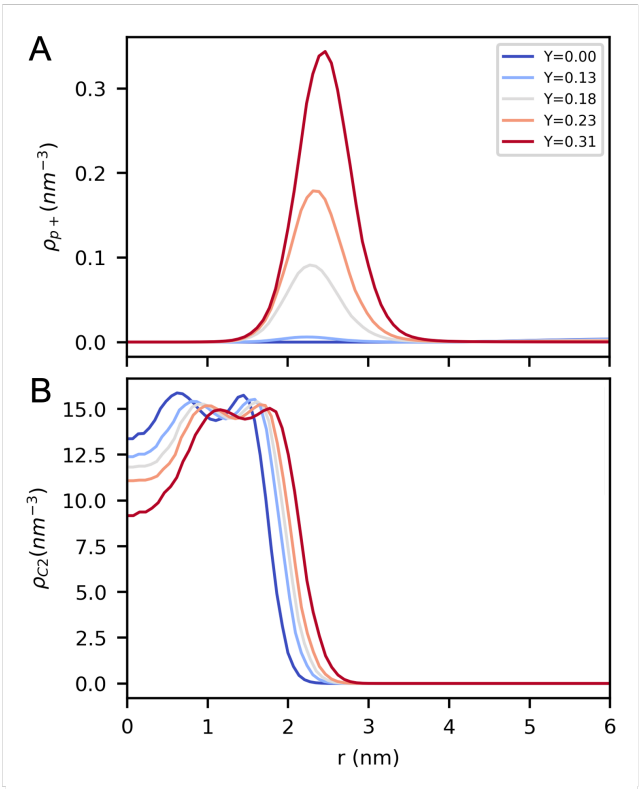

Figure E.2: Radial density profiles of (A) polycation monomer p^+ and (B) alkyl bead C_2 for various SDS fractions Y at 0.4 M NaCl with $C_{13}EO_{11}$ as the nonionic surfactant.

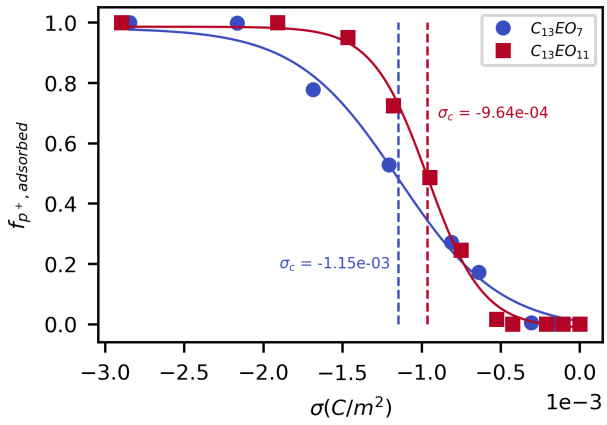

Figure E.3: Fraction of adsorbed PDADMA against micelle surface charge density for mix micelles of SDS with nonionic $C_{13}EO_n$ surfactants at 0.4 M NaCl. σ_c is the micelle surface charge density at the Y_c transition.

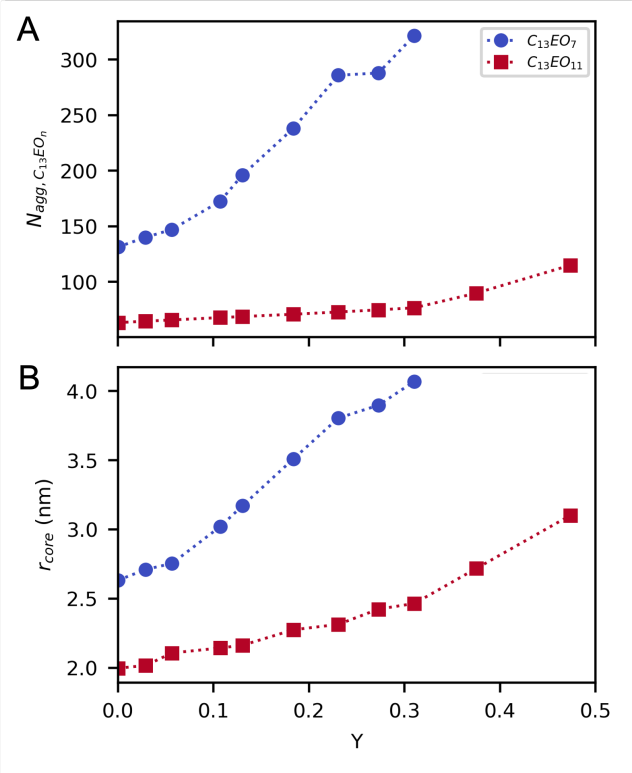

Figure E.4: (A) Aggregation number of ethoxylated nonionic surfactant $C_{13}EO_n$ and (B) core radius (measured at 0.1 C_2 locus) of the mixed micelle as a function of Y at 0.4 M NaCl.

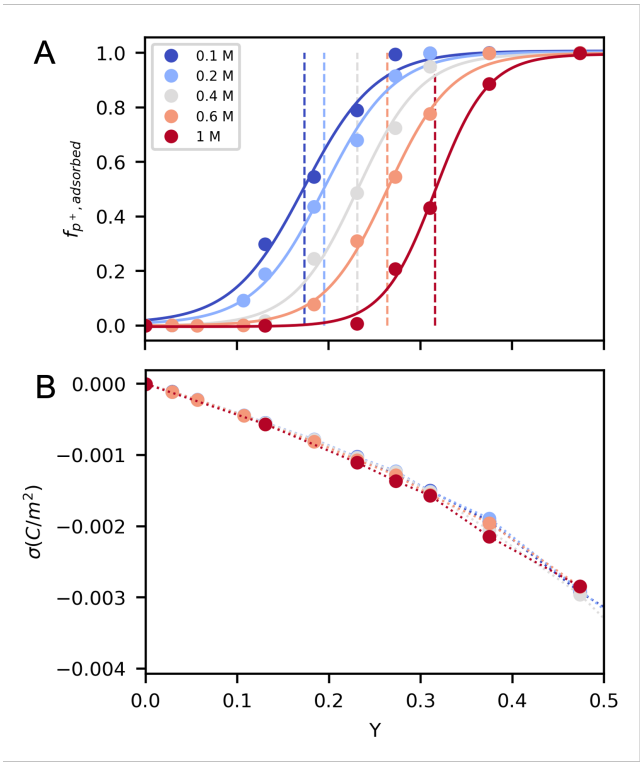

Figure E.5: SCFT-predicted (A) binding isotherms and (B) micelle surface charge density as the anionic surfactant mole fraction, Y, increases during the course of the isoionic titration for SDS/$C_{13}EO_{11}$ micelles at salt concentrations from 0.1 to 1.0 M. Dashed lines indicate the composition Y_c at the inflection.

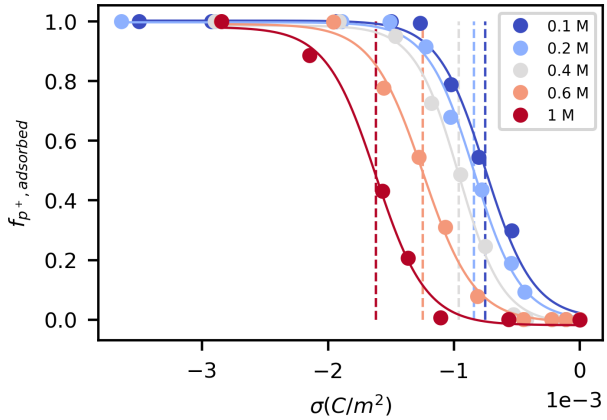

Figure E.6: Fraction of adsorbed PDADMA against micelle surface charge density for SDS/$C_{13}EO_{11}$ micelles at salt concentrations from 0.1 to 1.0 M. Dashed lines indicate the surface charge density σ_c at the inflection.

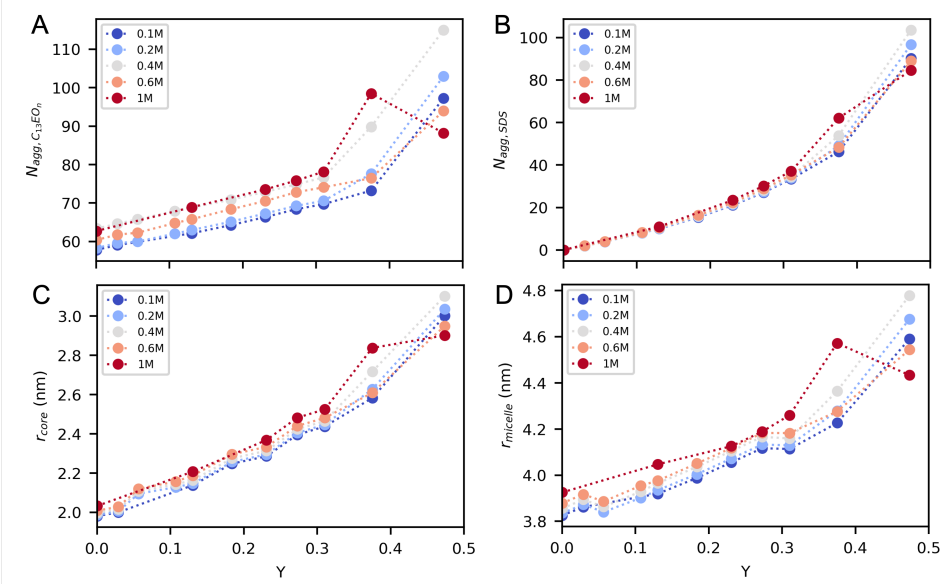

Figure E.7: SCFT-predicted aggregation numbers of (A) SDS and (B) $C_{13}EO_{11}$ as a function of Y for various salt concentrations. (C) Core and (D) micelle radii at the same conditions.

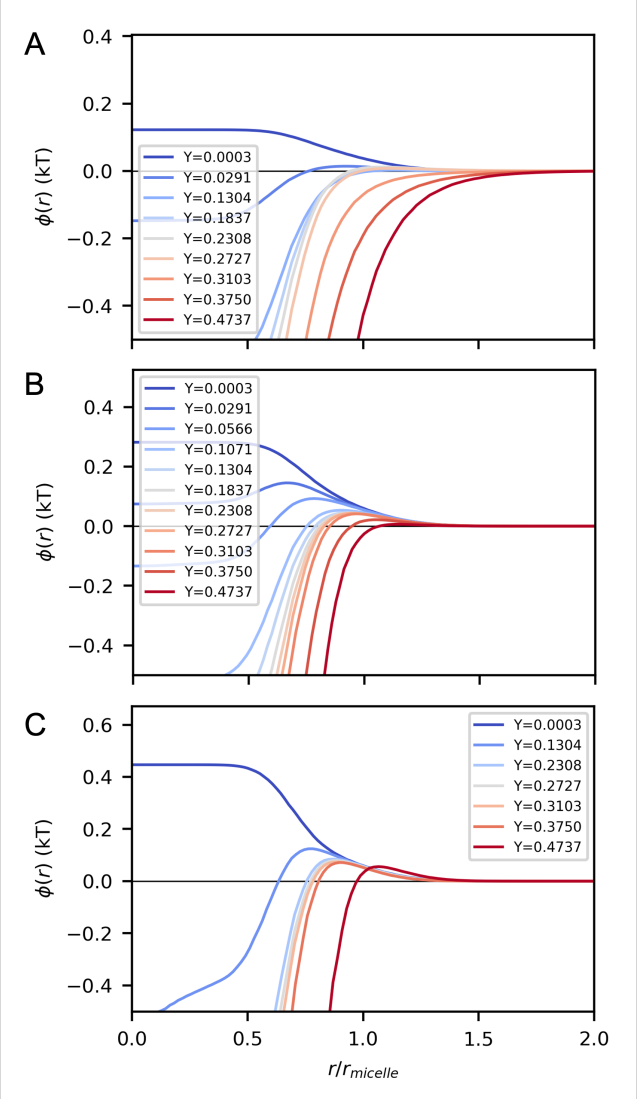

Figure E.8: SCFT-predicted electrostatic potential as a function of the radius from the micelle center for various anionic surfactant fractions, Y, at (A) 0.1 (B) 0.4 and (C) 1.0 M NaCl. The x-axis is normalized by the micelle total radii at same conditions reported in Fig. E.7D.

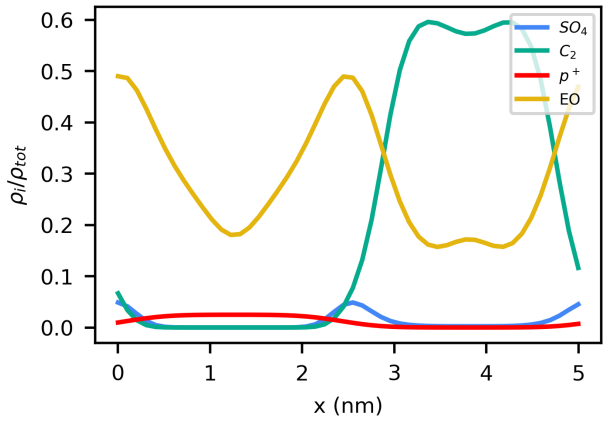

Figure E.9: SCFT-predicted density profile of the lamellar structure at 0.75 weight fraction for the polyelectrolyte-micelle mixture corresponding to Fig. 9. The equilibrium spacing from SCFT is 5.1 nm.

Appendix F

Supplemental Information for Chapter 5

Table F.1: Water excluded volume strength v_{ww} and fixed parameters.

Parameter	Value	
v_{ww}	0.1	k_BT
σ_w	0.0	e
σ_1	0.0	e
σ_2	0.0	e
σ_3	0.0	e
σ_{3+}	+1.0	e
σ_{3-}	-1.0	e
a_w	0.31	nm
a_1	0.5	nm
a_2	0.5	nm
a_3	0.5	nm
a_{3+}	0.5	nm
a_{3-}	0.5	nm

Fig. F.2 suggests that including the tail-tail simulation as one of the reference simulations increases the sampling frequency of the aggregated state. This can be quantified

Table F.2: Model Ia - Optimized root-mean-square bond length b and excluded volume strength $v_{\alpha\gamma}$ of replicates.

Parameter		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
b	(nm)	0.41404	0.41192	0.41189	0.41129	0.41337	0.41255	0.41280	0.41233	0.41404	0.41268	0.41115	0.41386	0.41301	0.41285	0.41272	0.41246	0.41180	0.41301	0.41408	0.41100
v_{11}	(k_BT)	2.73787	2.83906	2.82551	2.86648	2.88342	2.85235	2.84729	2.77189	2.79962	2.81338	2.87534	2.80244	2.85305	2.83588	2.83536	2.84609	2.82070	2.78056	2.82827	2.85910
v_{12}	(k_BT)	1.70924	1.73794	1.66928	1.66263	1.64941	1.74187	1.71843	1.68367	1.72961	1.70928	1.69286	1.69700	1.65142	1.76568	1.63628	1.70409	1.68473	1.77111	1.72730	1.67673
v_{13}	(k_BT)	0.67281	0.96047	0.98897	0.95627	1.02508	0.93133	0.95629	1.01900	0.90869	0.94185	0.92496	0.93818	0.93660	0.88312	0.90921	0.93532	1.01833	0.90925	0.85121	0.93767
v_{1w}	(k_BT)	0.49203	0.49986	0.49640	0.49816	0.49764	0.49761	0.49699	0.49424	0.49370	0.49501	0.49770	0.49527	0.49692	0.49672	0.49429	0.49570	0.49397	0.49551	0.49459	0.49488
v_{22}	(k_BT)	0.03034	0.01106	-0.01266	0.09596	0.04889	0.04549	0.02429	0.10683	0.05100	-0.01319	0.05897	0.02811	0.04987	-0.04442	0.03318	0.11903	0.04641	0.08129	0.05097	0.05784
v_{23}	(k_BT)	0.40474	0.38959	0.52529	0.47645	0.49378	0.38801	0.42453	0.41955	0.44131	0.42932	0.38907	0.41929	0.43086	0.43205	0.41497	0.37401	0.45773	0.45680	0.44184	0.46862
v_{2w}	(k_BT)	0.24978	0.24529	0.24865	0.24826	0.25258	0.24958	0.24810	0.24924	0.25629	0.24624	0.24808	0.24921	0.24742	0.24521	0.23953	0.24980	0.24607	0.25723	0.24706	0.24590
v_{33}	(k_BT)	2.92536	3.04738	2.82124	2.92811	2.78617	2.97987	2.90049	2.83204	2.96774	3.04416	2.99622	2.96416	2.96303	3.05710	3.04496	2.96112	2.84812	2.84117	2.99455	2.95457
v_{3w}	(k_BT)	0.29902	0.32566	0.32042	0.32054	0.31649	0.31854	0.32118	0.31599	0.31684	0.32608	0.31653	0.32144	0.32202	0.32340	0.32862	0.31843	0.32210	0.31110	0.32290	0.32298

Table F.3: Model Ib - Optimized root-mean-square bond length b and excluded volume strength $v_{\alpha\gamma}$ of replicates.

Parameter		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
b	(nm)	0.40854	0.40876	0.40762	0.40741	0.40701	0.40819	0.40976	0.40880	0.40666	0.40900	0.40892	0.40914	0.40947	0.40911	0.40837	0.40829	0.40904	0.40750	0.40712	0.40882
v_{11}	(k_BT)	4.42845	4.32203	4.53392	4.51929	4.34306	4.43908	4.42836	4.43136	4.45100	4.46892	4.28534	4.47855	4.35988	4.36472	4.34902	4.40848	4.49980	4.32262	4.46235	4.47475
v_{12}	(k_BT)	1.62920	1.73553	1.61769	1.57743	1.62968	1.64721	1.60594	1.55880	1.65054	1.61908	1.72756	1.64861	1.65690	1.69107	1.61460	1.69663	1.54695	1.71394	1.61792	1.64051
v_{13}	(k_BT)	1.73189	1.79614	1.74193	1.87055	1.86963	1.91232	1.80964	1.86670	1.63487	1.83960	1.78818	1.78800	1.87253	1.82829	1.89391	1.91066	2.01537	1.76831	1.82836	1.86992
v_{1w}	(k_BT)	0.55571	0.55465	0.55893	0.55776	0.55600	0.56184	0.55628	0.55673	0.55974	0.55927	0.55499	0.56069	0.55865	0.55918	0.55446	0.56674	0.56148	0.55736	0.56112	0.56193
v_{22}	(k_BT)	0.32744	0.28480	0.22874	0.32513	0.30278	0.33216	0.25634	0.40398	0.25009	0.29400	0.25807	0.25594	0.28742	0.28170	0.33994	0.26777	0.35240	0.30624	0.27512	0.27474
v_{23}	(k_BT)	0.69936	0.78558	0.70611	0.73893	0.68389	0.66914	0.74493	0.66011	0.71600	0.70946	0.81160	0.75358	0.73457	0.75888	0.78812	0.75125	0.72006	0.72798	0.74973	0.74659
v_{2w}	(k_BT)	0.26751	0.27684	0.26418	0.26911	0.26669	0.26515	0.26364	0.26948	0.26291	0.26638	0.27272	0.26667	0.26912	0.26682	0.26864	0.26312	0.26659	0.27008	0.26230	0.26817
v_{33}	(k_BT)	3.41798	3.26563	3.48471	3.40275	3.45841	3.48721	3.33186	3.37125	3.48506	3.40822	3.24674	3.35207	3.36718	3.33973	3.18910	3.34716	3.39032	3.37895	3.39613	3.34823
v_{3w}	(k_BT)	0.37342	0.36373	0.37732	0.37408	0.37793	0.37542	0.37690	0.36868	0.37777	0.37648	0.37195	0.37093	0.37277	0.37297	0.37239	0.37695	0.37688	0.37027	0.37734	0.37407

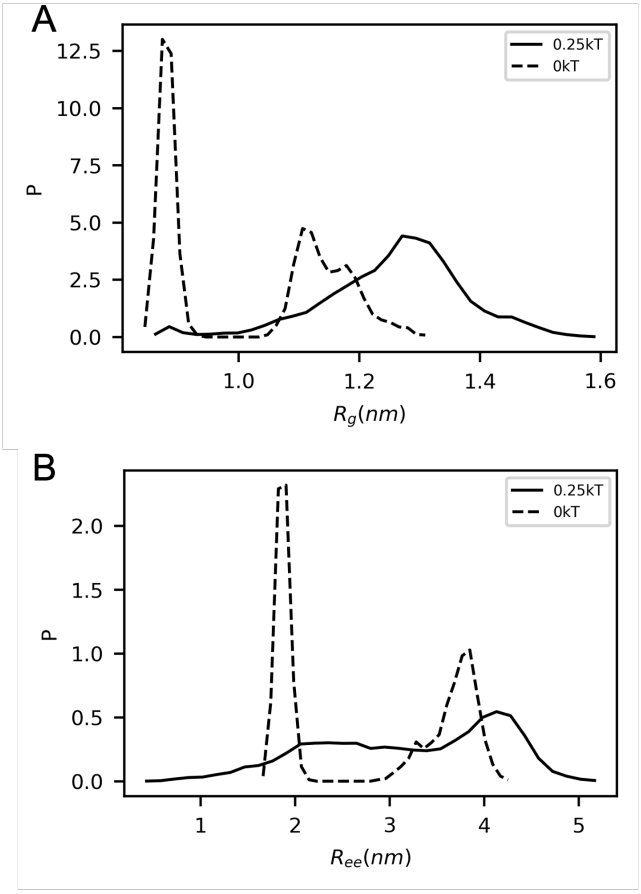

Figure F.1: (A) Radius of gyration and (B) end-to-end distance distribution of the hydrophobic tails AA simulation with (solid line, $v_{bias} = 0.25\ k_BT$) and without the repulsive potential (dashed line, $v_{bias} = 0\ k_BT$).

by the Gaussian weighted number of contacts between any residues i and j of the IDP in the mapped reference simulations:

$$X = \sum_{i,j} \frac{\exp(-r_{ij})^2/2(a_i^2 + a_j^2)}{\left(2\pi(a_i^2 + a_j^2)\right)^{3/2}}, \tag{F.1}$$

where $a_i = 0.5$ nm is the Gaussian regularization length of the amino acid CG beads. Eq. F.1 calculates the number of contact between any two residues weighted by their distance, r_{ij}. Conceptually, X is the structural variable corresponding to the employed Gaussian interaction potential that the relative entropy minimization tries to match when

Table F.4: Model IIa - Optimized root-mean-square bond length b and excluded volume strength $v_{\alpha\gamma}$ of replicates.

Parameter		1	2	3	4	5	6
b	(nm)	0.42703	0.42861	0.42953	0.42979	0.43028	0.42997
v_{11}	(k_BT)	2.14475	2.14194	2.17733	2.13583	2.15696	2.12751
v_{12}	(k_BT)	1.31010	1.33362	1.32762	1.30144	1.30557	1.35289
v_{13}	(k_BT)	0.98507	0.99424	0.85433	0.99116	0.96647	0.95341
v_{1w}	(k_BT)	0.48517	0.48677	0.48498	0.48429	0.48281	0.48456
v_{22}	(k_BT)	-0.14156	-0.08206	-0.13154	-0.10071	-0.04072	-0.10613
v_{23}	(k_BT)	0.58323	0.53698	0.46693	0.57534	0.54284	0.54677
v_{2w}	(k_BT)	0.23278	0.23440	0.22839	0.23471	0.23839	0.23766
v_{33}	(k_BT)	2.72227	2.68802	2.92339	2.68625	2.68400	2.68386
v_{3w}	(k_BT)	0.32302	0.31986	0.32742	0.32162	0.31887	0.31915

Table F.5: Model IIb - Optimized root-mean-square bond length b and excluded volume strength $v_{\alpha\gamma}$ of replicates.

Parameter		1	2	3	4	5	6
b	(nm)	0.43532	0.43653	0.43520	0.43483	0.43356	0.43575
v_{11}	(k_BT)	2.62957	2.64275	2.66497	2.67106	2.64449	2.70540
v_{12}	(k_BT)	1.14924	1.18237	1.14619	1.15980	1.21354	1.14914
v_{13}	(k_BT)	1.67453	1.55666	1.68777	1.62585	1.53982	1.55289
v_{1w}	(k_BT)	0.52428	0.52061	0.52262	0.52157	0.52638	0.52436
v_{22}	(k_BT)	0.14899	0.13910	0.14036	0.10959	0.07190	0.13550
v_{23}	(k_BT)	0.85076	0.87990	0.90642	0.87477	0.95256	0.91057
v_{2w}	(k_BT)	0.27176	0.27807	0.27446	0.27179	0.27633	0.27444
v_{33}	(k_BT)	2.85672	2.81779	2.77738	2.92479	2.75814	2.75883
v_{3w}	(k_BT)	0.36290	0.35852	0.36058	0.36486	0.35715	0.36044

deriving CG parameters. As listed in Table F.6, reference system I (a single simulation of the head and tail IDP fragments) with a small number of amino acid contacts will likely produce a CG model that is less hydrophobic than reference system II (expanded ensemble of three simulations).

Table F.6: Gaussian weighted number of contacts between any IDP residues.

Reference system	X	X/n_{pairs} [a]
I	61	0.065
II [b]	79	0.088

[a] Weighted by the total number of possible contacts

[b] Average over three reference simulations

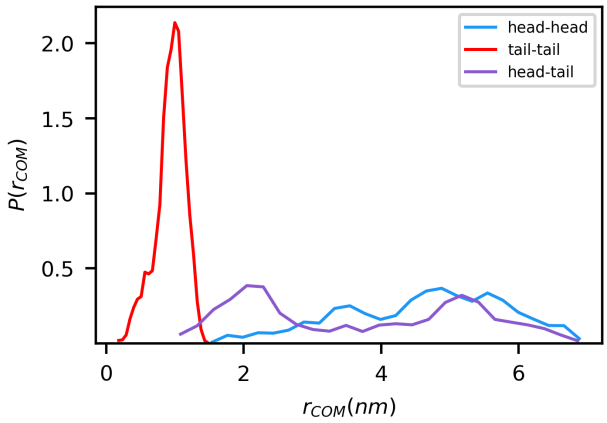

Figure F.2: Center-of-mass distance between the IDP fragments in three reference simulations. The simulation of two tail fragments (tail-tail) observes higher frequency of the aggregated state than the other two simulations.

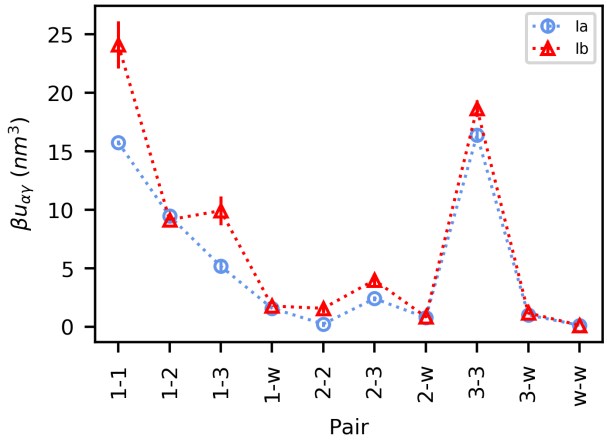

Figure F.3: Average integrated excluded volume interactions, $u_{\alpha\gamma} = v_{\alpha\gamma}(2\pi(a_\alpha^2 + a_\gamma^2))^{3/2}$, for models Ia and Ib. Overall, model Ib exhibits higher repulsion, evidenced by higher values of $u_{\alpha\gamma}$, between the amino acids. This results in less hydrophobic IDP, thus higher CMC, using model Ib.

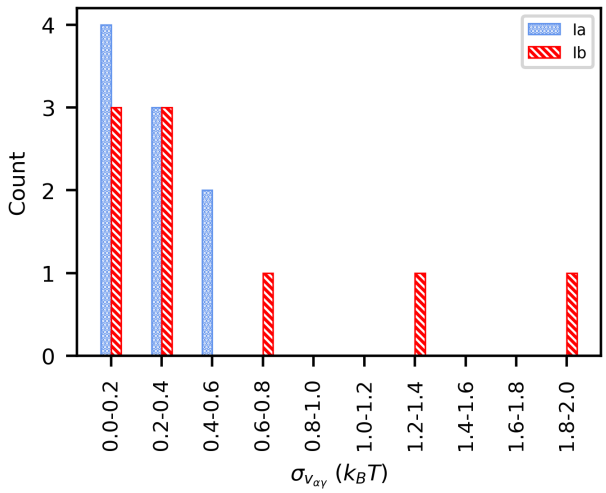

Figure F.4: Histogram for standard deviations of 9 excluded volume parameters, $v_{\alpha\gamma}$, across 20 replicates for models Ia and Ib.

Appendix G

Supplemental Information for

Chapter 6

Table G.1: Reference all-atom system sizes. Each row represents a simulation shown in Fig. 6.3 in main text.

	n_w	$n_{cellulose}$ (8-mer)	DS	n_{G0}	n_{G1_2}	n_{G1_3}	n_{G1_6}	$n_{G1_{23}}$	$n_{G1_{26}}$	$n_{G1_{36}}$	n_{G3}	box(nm)	ensemble
1	3305											$4.7 \times 4.7 \times 4.7$	NPT
2	15278	33	0.0	264								$8.0 \times 8.0 \times 8.0$	NPT
3	15278	33	1.0		88	88	88					$8.1 \times 8.1 \times 8.1$	NPT
4	15278	33	2.0					88	88	88		$8.1 \times 8.1 \times 8.1$	NPT
5	15278	33	3.0								264	$8.1 \times 8.1 \times 8.1$	NPT
6	15278	33	0.5	66	22	22	22	22	22	22	66	$8.1 \times 8.1 \times 8.1$	NPT

Table G.2: Optimized excluded volume interaction parameters, $v_{\alpha,\gamma}$ (k_BT)

α \ γ	w	G0	$G1_2$	$G1_3$	$G1_6$	$G2_{23}$	$G2_{26}$	$G2_{36}$	G3
w	0.33897								
G0	1.91172	1.87274							
$G1_2$	2.62044	2.09653	3.45210						
$G1_3$	2.51129	2.86779	3.43468	3.39547					
$G1_6$	2.49281	2.35662	3.56470	3.05431	3.21460				
$G2_{23}$	3.23102	2.81463	5.48032	4.54609	4.97849	5.38668			
$G2_{26}$	3.18075	3.43011	3.70639	3.89624	4.19072	5.46667	5.23067		
$G2_{36}$	3.14279	3.04176	5.78957	3.60610	3.33025	5.34429	5.21189	5.14820	
G3	3.83704	4.04721	4.88777	4.59731	5.38941	6.30660	6.88359	6.30443	7.74686

Table G.3: Remaining coarse-grained interaction parameters: optimized root-mean-square bond length b and fixed excluded volume interaction range a_α.

Parameter	Value	
b	0.46953	nm
a_w	0.6	nm
a_{G0}	0.6	nm
a_{G1_2}	0.6	nm
a_{G1_3}	0.6	nm
a_{G1_6}	0.6	nm
$a_{G2_{23}}$	0.6	nm
$a_{G2_{26}}$	0.6	nm
$a_{G2_{36}}$	0.6	nm
a_{G3}	0.6	nm

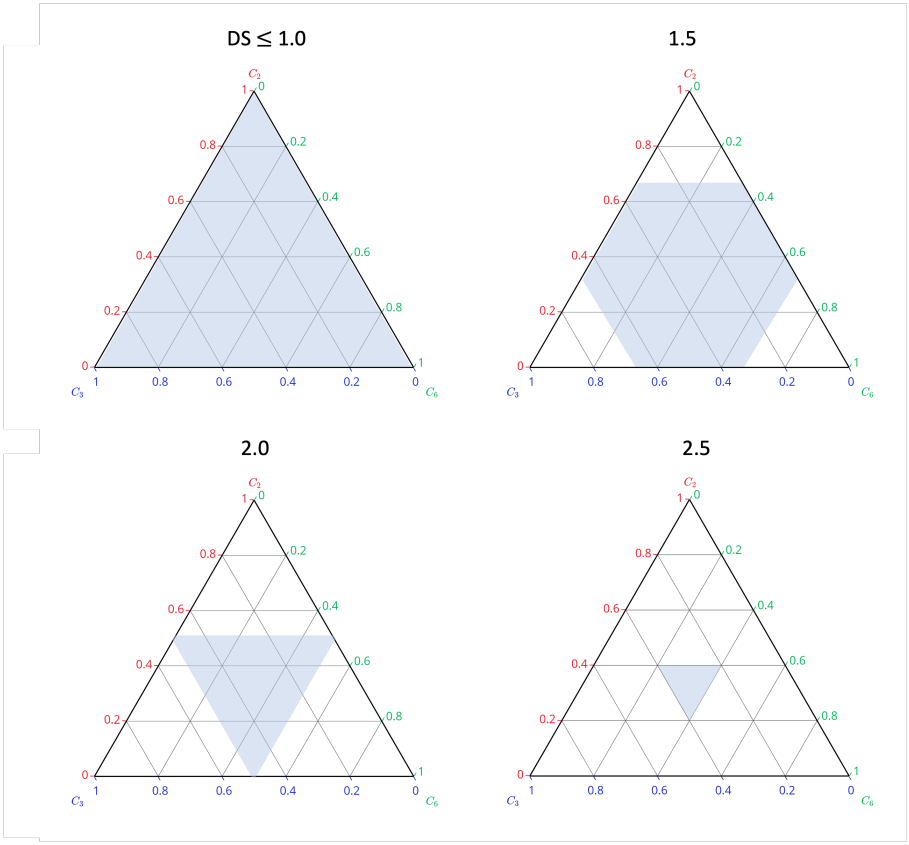

Figure G.1: Composition space specified by the relative acetylation ratio C2:C3:C6 for various DS values. Each shaded region corresponds to the combination of C2:C3:C6 ratios that are achievable within the constraints of the specific DS value.

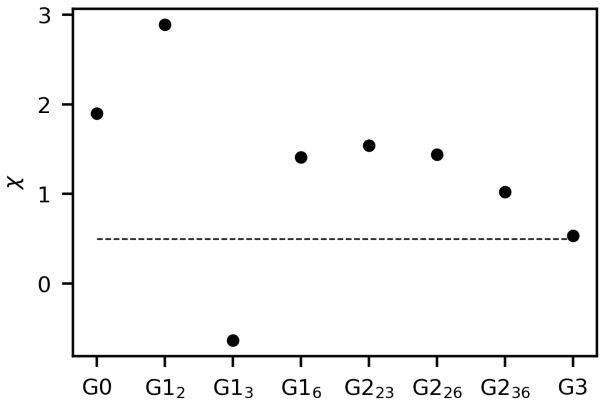

Figure G.2: Effective χ for 8 CG bead types. Dashed line denotes $\chi = 0.5$

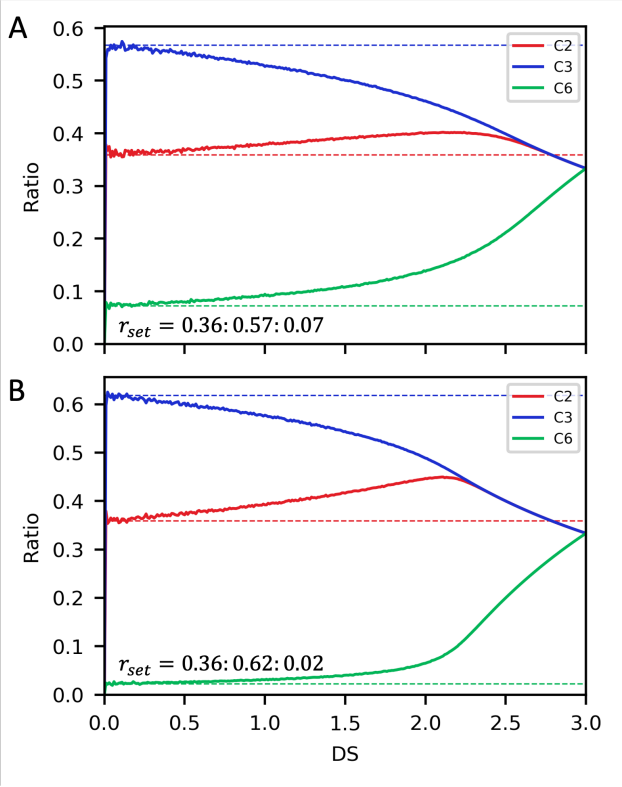

Figure G.3: Relative substituent ratios at C2, C3, and C6 corresponding to (A) $r_{set} = 0.36 : 0.57 : 0.07$ and (B) $r_{set} = 0.36 : 0.62 : 0.02$. Dashed lines represent r_{set}.

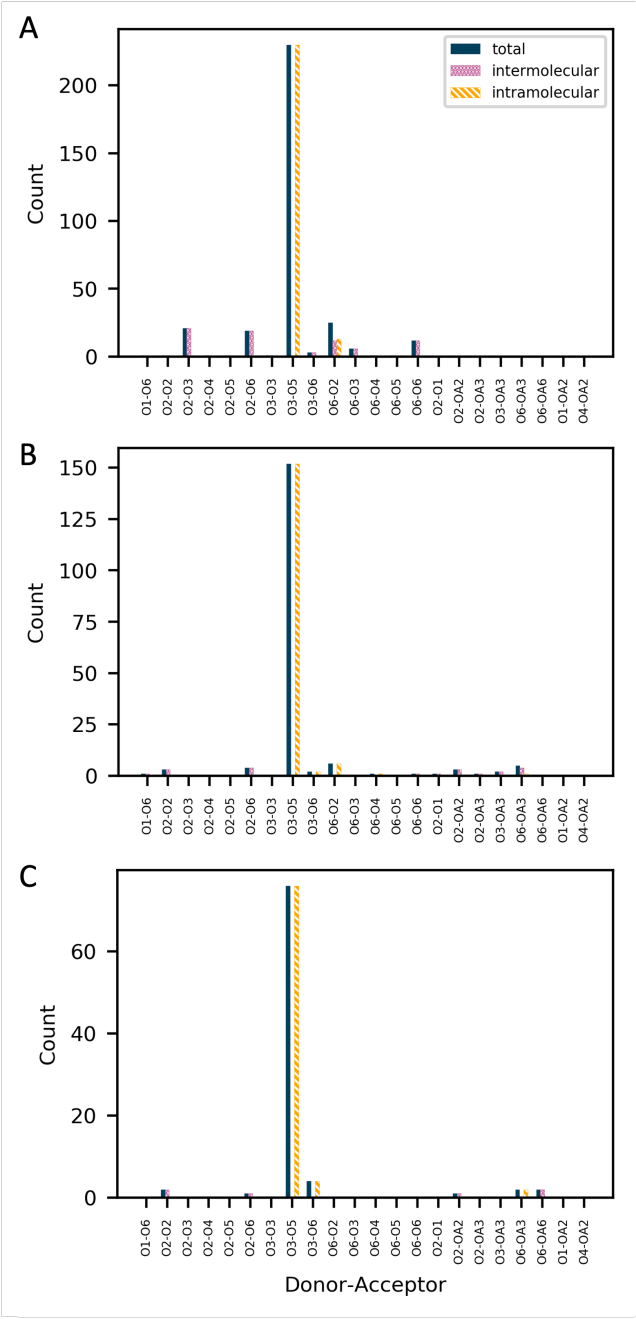

Figure G.4: Number of total, inter-, and intramolecular hydrogen bonds between all monomers of cellulose categorized by the hydrogen donor-acceptor pairs calculated from AA simulations of 8-mers at (A) DS=0, (B) DS=1, and (C) DS=2, corresponding to simulations 2, 3, and 4 in Fig. 6.3.

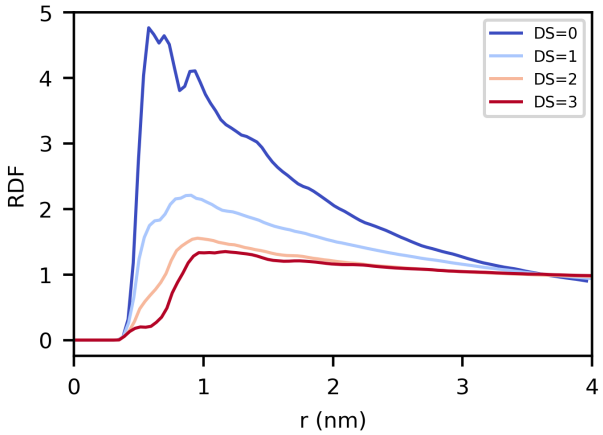

Figure G.5: Interchain radial distribution functions (RDFs) between C1 carbons of the glucose monomers as a function of DS. DS=0, 1, 2, and 3 correspond to reference AA simulations 2-5, respectively.

References

[1] Lu Li, Samanvaya Srivastava, Siqi Meng, Jeffrey M Ting, and Matthew V Tirrell. Effects of non-electrostatic intermolecular interactions on the phase behavior of ph-sensitive polyelectrolyte complexes. *Macromolecules*, 53(18):7835–7844, 2020.

[2] Yimin Luo, Mengyang Gu, Chelsea ER Edwards, Megan T Valentine, and Matthew E Helgeson. High-throughput microscopy to determine morphology, microrheology, and phase boundaries applied to phase separating coacervates. *Soft Matter*, 18(15):3063–3075, 2022.

[3] Cornelus G De Kruif, Fanny Weinbreck, and Renko de Vries. Complex coacervation of proteins and anionic polysaccharides. *Current opinion in colloid & interface science*, 9(5):340–349, 2004.

[4] VB Tolstoguzov. Some physico-chemical aspects of protein processing in foods. multicomponent gels. *Food Hydrocolloids*, 9(4):317–332, 1995.

[5] F Weinbreck, R De Vries, P Schrooyen, and CG De Kruif. Complex coacervation of whey proteins and gum arabic. *Biomacromolecules*, 4(2):293–303, 2003.

[6] ED Goddard. Polymer/surfactant interaction—its relevance to detergent systems. *Journal of the American Oil Chemists' Society*, 71(1):1–16, 1994.

[7] Sara Llamas, Eduardo Guzman, Francisco Ortega, Nawel Baghdadli, Colette Cazeneuve, Ramon G Rubio, and Gustavo S Luengo. Adsorption of polyelectrolytes and polyelectrolytes-surfactant mixtures at surfaces: A physico-chemical approach to a cosmetic challenge. *Advances in colloid and interface science*, 222:461–487, 2015.

[8] Q Wang, Y Dan, and XG Wang. A new polymer flooding agent prepared through intermacromolecular complexation. *Journal of Macromolecular Science, Part A: Pure and Applied Chemistry*, 34(7):1155–1169, 1997.

[9] Fanny Weinbreck, M Minor, and CG De Kruif. Microencapsulation of oils using whey protein/gum arabic coacervates. *Journal of microencapsulation*, 21(6):667–679, 2004.

[10] Hua Zhao, Chengjun Sun, Russell J Stewart, and J Herbert Waite. Cement proteins of the tube-building polychaete phragmatopoma californica. *Journal of Biological Chemistry*, 280(52):42938–42944, 2005.

[11] Hui Shao, Kent N Bachus, and Russell J Stewart. A water-borne adhesive modeled after the sandcastle glue of p. californica. *Macromolecular bioscience*, 9(5):464–471, 2009.

[12] Russell J Stewart, Ching Shuen Wang, and Hui Shao. Complex coacervates as a foundation for synthetic underwater adhesives. *Advances in colloid and interface science*, 167(1-2):85–93, 2011.

[13] Wei Wei, Yerpeng Tan, Nadine R Martinez Rodriguez, Jing Yu, Jacob N Israelachvili, and J Herbert Waite. A mussel-derived one component adhesive coacervate. *Acta biomaterialia*, 10(4):1663–1670, 2014.

[14] Louis A Luzzi. Microencapsulation. *Journal of pharmaceutical sciences*, 59(10):1367–1376, 1970.

[15] Sarah H Klass, Matthew J Smith, Tahoe A Fiala, Jess P Lee, Anthony O Omole, Bong-Gyoon Han, Kenneth H Downing, Sanjay Kumar, and Matthew B Francis. Self-assembling micelles based on an intrinsically disordered protein domain. *Journal of the American Chemical Society*, 141(10):4291–4299, 2019.

[16] Mihaela Delcea, Helmuth Möhwald, and André G Skirtach. Stimuli-responsive lbl capsules and nanoshells for drug delivery. *Advanced drug delivery reviews*, 63(9):730–747, 2011.

[17] Whitney C Blocher McTigue and Sarah L Perry. Design rules for encapsulating proteins into complex coacervates. *Soft Matter*, 15(15):3089–3103, 2019.

[18] Benjamin J Coscia, John C Shelley, Andrea R Browning, Jeffrey M Sanders, Robin Chaudret, Roger Rozot, Fabien Léonforte, Mathew D Halls, and Gustavo S Luengo. Shearing friction behaviour of synthetic polymers compared to a functionalized polysaccharide on biomimetic surfaces: models for the prediction of performance of eco-designed formulations. *Physical Chemistry Chemical Physics*, 25(3):1768–1780, 2023.

[19] Dominique Langevin. Complexation of oppositely charged polyelectrolytes and surfactants in aqueous solutions. a review. *Advances in colloid and interface science*, 147:170–177, 2009.

[20] Rungsima Chollakup, Wirasak Smitthipong, Claus D Eisenbach, and Matthew Tirrell. Phase behavior and coacervation of aqueous poly (acrylic acid)- poly (allylamine) solutions. *Macromolecules*, 43(5):2518–2528, 2010.

[21] Lu Li, Samanvaya Srivastava, Marat Andreev, Amanda B Marciel, Juan J de Pablo, and Matthew V Tirrell. Phase behavior and salt partitioning in polyelectrolyte complex coacervates. *Macromolecules*, 51(8):2988–2995, 2018.

[22] Yalin Liu, Brian Momani, H Henning Winter, and Sarah L Perry. Rheological characterization of liquid-to-solid transitions in bulk polyelectrolyte complexes. *Soft Matter*, 13(40):7332–7340, 2017.

[23] Amy Y Xu, Ebru Kizilay, Slawomir P Madro, Justin Z Vadenais, Kianan W McDonald, and Paul L Dubin. Dilution induced coacervation in polyelectrolyte–micelle and polyelectrolyte–protein systems. *Soft Matter*, 14(12):2391–2399, 2018.

[24] Xueming Tang, Weizhong Zou, Peter H Koenig, Shawn D McConaughy, Mike R Weaver, David M Eike, Michael J Schmidt, and Ronald G Larson. Multiscale modeling of the effects of salt and perfume raw materials on the rheological properties of commercial threadlike micellar solutions. *The Journal of Physical Chemistry B*, 121(11):2468–2485, 2017.

[25] Jason M Swails, Darrin M York, and Adrian E Roitberg. Constant ph replica exchange molecular dynamics in explicit solvent using discrete protonation states: implementation, testing, and validation. *Journal of chemical theory and computation*, 10(3):1341–1352, 2014.

[26] Torsten Wallin and Per Linse. Monte carlo simulations of polyelectrolytes at charged micelles. 1. effects of chain flexibility. *Langmuir*, 12(2):305–314, 1996.

[27] Scott PO Danielsen, James McCarty, Joan-Emma Shea, Kris T Delaney, and Glenn H Fredrickson. Molecular design of self-coacervation phenomena in block polyampholytes. *Proceedings of the National Academy of Sciences*, 116(17):8224–8232, 2019.

[28] Nazish Hoda and Ronald G Larson. Explicit-and implicit-solvent molecular dynamics simulations of complex formation between polycations and polyanions. *Macromolecules*, 42(22):8851–8863, 2009.

[29] Muralidharan S Sulatha and Upendra Natarajan. Molecular dynamics simulations of adsorption of poly (acrylic acid) and poly (methacrylic acid) on dodecyltrimethylammonium chloride micelle in water: effect of charge density. *The Journal of Physical Chemistry B*, 119(38):12526–12539, 2015.

[30] Barry Z Shang, Zuowei Wang, and Ronald G Larson. Effect of headgroup size, charge, and solvent structure on polymer- micelle interactions, studied by molecular

dynamics simulations. *The Journal of Physical Chemistry B*, 113(46):15170–15180, 2009.

[31] Zhenhai Liu, Yazhuo Shang, Jian Feng, Changjun Peng, Honglai Liu, and Ying Hu. Effect of hydrophilicity or hydrophobicity of polyelectrolyte on the interaction between polyelectrolyte and surfactants: molecular dynamics simulations. *The Journal of Physical Chemistry B*, 116(18):5516–5526, 2012.

[32] Heyi Liang and Juan J de Pablo. A coarse-grained molecular dynamics study of strongly charged polyelectrolyte coacervates: Interfacial, structural, and dynamical properties. *Macromolecules*, 55(10):4146–4158, 2022.

[33] Mithun Radhakrishna, Kush Basu, Yalin Liu, Rasmia Shamsi, Sarah L Perry, and Charles E Sing. Molecular connectivity and correlation effects on polymer coacervation. *Macromolecules*, 50(7):3030–3037, 2017.

[34] Junzhe Lou, Sean Friedowitz, Jian Qin, and Yan Xia. Tunable coacervation of well-defined homologous polyanions and polycations by local polarity. *ACS central science*, 5(3):549–557, 2019.

[35] Evan Spruijt, Adrie H Westphal, Jan Willem Borst, Martien A Cohen Stuart, and Jasper van der Gucht. Binodal compositions of polyelectrolyte complexes. *Macromolecules*, 43(15):6476–6484, 2010.

[36] Hadi M Fares, Yara E Ghoussoub, Jose D Delgado, Jingcheng Fu, Volker S Urban, and Joseph B Schlenoff. Scattering neutrons along the polyelectrolyte complex/coacervate continuum. *Macromolecules*, 51(13):4945–4955, 2018.

[37] HG Bungenberg de Jong and HR Kruyt. Coacervation (partial miscibility in colloid systems). In *Proc. K. Ned. Akad. Wet*, volume 32, pages 849–856, 1929.

[38] Evan Spruijt et al. Strength, structure and stability of polyelectrolyte complex coacervates. *Wageningen University, Wageningen*, 2012.

[39] Sarah L Perry, Yue Li, Dimitrios Priftis, Lorraine Leon, and Matthew Tirrell. The effect of salt on the complex coacervation of vinyl polyelectrolytes. *Polymers*, 6(6):1756–1772, 2014.

[40] Jasper Van der Gucht, Evan Spruijt, Marc Lemmers, and Martien A Cohen Stuart. Polyelectrolyte complexes: Bulk phases and colloidal systems. *Journal of colloid and interface science*, 361(2):407–422, 2011.

[41] Dimitrios Priftis, Nicolas Laugel, and Matthew Tirrell. Thermodynamic characterization of polypeptide complex coacervation. *Langmuir*, 28(45):15947–15957, 2012.

[42] Suvasree Mukherjee, Abhijit Dan, Subhash C Bhattacharya, Amiya K Panda, and Satya P Moulik. Physicochemistry of interaction between the cationic polymer poly (diallyldimethylammonium chloride) and the anionic surfactants sodium dodecyl sulfate, sodium dodecylbenzenesulfonate, and sodium n-dodecanoylsarcosinate in water and isopropyl alcohol- water media. *Langmuir*, 27(9):5222–5233, 2011.

[43] Guillaume Sudre, Yvette Tran, Costantino Creton, and Dominique Hourdet. ph/temperature control of interpolymer complexation between poly (acrylic acid) and weak polybases in aqueous solutions. *Polymer*, 53(2):379–385, 2012.

[44] Qifeng Wang and Joseph B Schlenoff. The polyelectrolyte complex/coacervate continuum. *Macromolecules*, 47(9):3108–3116, 2014.

[45] Pilar C Suarez-Martinez, Piotr Batys, Maria Sammalkorpi, and Jodie L Lutkenhaus. Time–temperature and time–water superposition principles applied to poly (allylamine)/poly (acrylic acid) complexes. *Macromolecules*, 52(8):3066–3074, 2019.

[46] Dimitrios Priftis, Xiaoxing Xia, Khatcher O Margossian, Sarah L Perry, Lorraine Leon, Jian Qin, Juan J de Pablo, and Matthew Tirrell. Ternary, tunable polyelectrolyte complex fluids driven by complex coacervation. *Macromolecules*, 47(9):3076–3085, 2014.

[47] John H Clint. Micellization of mixed nonionic surface active agents. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 71:1327–1334, 1975.

[48] Walter C Presto and Walter Preston. Some correlating principles of detergent action. *The Journal of Physical Chemistry*, 52(1):84–97, 1948.

[49] Milton J Rosen and Joy T Kunjappu. *Surfactants and interfacial phenomena*. John Wiley & Sons, 2012.

[50] Hanne M van der Kooij, Evan Spruijt, Ilja K Voets, Remco Fokkink, Martien A Cohen Stuart, and Jasper van der Gucht. On the stability and morphology of complex coacervate core micelles: From spherical to wormlike micelles. *Langmuir*, 28(40):14180–14191, 2012.

[51] Jose Rodrigo Magana, Christian CM Sproncken, and Ilja K Voets. On complex coacervate core micelles: Structure-function perspectives. *Polymers*, 12(9):1953, 2020.

[52] Alan Parker and Wolfgang Fieber. Viscoelasticity of anionic wormlike micelles: effects of ionic strength and small hydrophobic molecules. *Soft Matter*, 9(4):1203–1213, 2013.

[53] Nasreen Khan and Blair Brettmann. Intermolecular interactions in polyelectrolyte and surfactant complexes in solution. *Polymers*, 11(1):51, 2018.

[54] J Th G Overbeek and MJ Voorn. Phase separation in polyelectrolyte solutions. theory of complex coacervation. *Journal of Cellular and Comparative Physiology*, 49(S1):7–26, 1957.

[55] I Michaeli, J Th G Overbeek, and MJ Voorn. Phase separation of polyelectrolyte solutions. *Journal of Polymer Science*, 23(103):443–450, 1957.

[56] DA Mcquarrie. Statistical mechanics, 1965.

[57] Masao Doi, Samuel Frederick Edwards, and Samuel Frederick Edwards. *The theory of polymer dynamics*, volume 73. oxford university press, 1988.

[58] V Yu Borue and I Ya Erukhimovich. A statistical theory of globular polyelectrolyte complexes. *Macromolecules*, 23(15):3625–3632, 1990.

[59] V Yu Borue and I Ya Erukhimovich. A statistical theory of weakly charged polyelectrolytes: fluctuations, equation of state and microphase separation. *Macromolecules*, 21(11):3240–3249, 1988.

[60] Alexander Kudlay and Monica Olvera de la Cruz. Precipitation of oppositely charged polyelectrolytes in salt solutions. *The Journal of chemical physics*, 120(1):404–412, 2004.

[61] Alexander Kudlay, Alexander V Ermoshkin, and Monica Olvera de La Cruz. Complexation of oppositely charged polyelectrolytes: effect of ion pair formation. *Macromolecules*, 37(24):9231–9241, 2004.

[62] Martin Castelnovo and J-F Joanny. Complexation between oppositely charged polyelectrolytes: Beyond the random phase approximation. *The European Physical Journal E*, 6(1):377–386, 2001.

[63] Robert A Riggleman, Rajeev Kumar, and Glenn H Fredrickson. Investigation of the interfacial tension of complex coacervates using field-theoretic simulations. *The Journal of chemical physics*, 136(2):024903, 2012.

[64] Jian Qin and Juan J de Pablo. Criticality and connectivity in macromolecular charge complexation. *Macromolecules*, 49(22):8789–8800, 2016.

[65] Kris T Delaney and Glenn H Fredrickson. Theory of polyelectrolyte complexation—complex coacervates are self-coacervates. *The Journal of Chemical Physics*, 146(22):224902, 2017.

[66] James McCarty, Kris T Delaney, Scott PO Danielsen, Glenn H Fredrickson, and Joan-Emma Shea. Complete phase diagram for liquid–liquid phase separation of intrinsically disordered proteins. *The journal of physical chemistry letters*, 10(8):1644–1652, 2019.

[67] Scott PO Danielsen, James McCarty, Joan-Emma Shea, Kris T Delaney, and Glenn H Fredrickson. Small ion effects on self-coacervation phenomena in block polyampholytes. *The Journal of chemical physics*, 151(3):034904, 2019.

[68] Jonghoon Lee, Yuri O Popov, and Glenn H Fredrickson. Complex coacervation: A field theoretic simulation study of polyelectrolyte complexation. *The Journal of chemical physics*, 128(22):224908, 2008.

[69] Yuri O Popov, Jonghoon Lee, and Glenn H Fredrickson. Field-theoretic simulations of polyelectrolyte complexation. *arXiv preprint arXiv:0705.0718*, 2007.

[70] Kevin Shen and Zhen-Gang Wang. Polyelectrolyte chain structure and solution phase behavior. *Macromolecules*, 51(5):1706–1717, 2018.

[71] Maria Panoukidou, Charlie Ray Wand, Annalaura Del Regno, Richard L Anderson, and Paola Carbone. Constructing the phase diagram of sodium laurylethoxysulfate using dissipative particle dynamics. *Journal of colloid and interface science*, 557:34–44, 2019.

[72] Pep Espanol and Patrick B Warren. Perspective: Dissipative particle dynamics. *The Journal of chemical physics*, 146(15), 2017.

[73] Karel Šindelka, Adam Kowalski, Michael Cooke, César Mendoza, and Martin Lísal. Interactions of cationic surfactant-fatty alcohol monolayers with natural human hair surface: Insights from dissipative particle dynamics. *Journal of Molecular Liquids*, 375:121385, 2023.

[74] Glenn Fredrickson et al. *The equilibrium theory of inhomogeneous polymers*, volume 134. Oxford University Press on Demand, 2006.

[75] Glenn H Fredrickson, Venkat Ganesan, and François Drolet. Field-theoretic computer simulation methods for polymers and complex fluids. *Macromolecules*, 35(1):16–39, 2002.

[76] Erin M Lennon, George O Mohler, Hector D Ceniceros, Carlos J García-Cervera, and Glenn H Fredrickson. Numerical solutions of the complex langevin equations in polymer field theory. *Multiscale Modeling & Simulation*, 6(4):1347–1370, 2008.

[77] Glenn H Fredrickson and Kris 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):e2201804119, 2022.

[78] Debra J Audus, Jeffrey D Gopez, Daniel V Krogstad, Nathaniel A Lynd, Edward J Kramer, Craig J Hawker, and Glenn H Fredrickson. Phase behavior of electrostatically complexed polyelectrolyte gels using an embedded fluctuation model. *Soft Matter*, 11(6):1214–1225, 2015.

[79] Pengfei Zhang, Nayef M Alsaifi, Jianzhong Wu, and Zhen-Gang Wang. Polyelectrolyte complex coacervation: Effects of concentration asymmetry. *The Journal of chemical physics*, 149(16):163303, 2018.

[80] Sarah L Perry and Charles E Sing. Prism-based theory of complex coacervation: Excluded volume versus chain correlation. *Macromolecules*, 48(14):5040–5053, 2015.

[81] M Scott Shell. Coarse-graining with the relative entropy. *Advances in chemical physics*, 161:395–441, 2016.

[82] My Nguyen, Nicholas Sherck, Kevin Shen, Chelsea ER Edwards, Brian Yoo, Stephan Köhler, Joshua C Speros, Matthew E Helgeson, Kris T Delaney, M Scott Shell, et al. Predicting polyelectrolyte coacervation from a molecularly informed field-theoretic model. *Macromolecules*, 55(21):9868–9879, 2022.

[83] My Nguyen, Kevin Shen, Nicholas Sherck, Stephan Köhler, Rohini Gupta, Kris T Delaney, M Scott Shell, and Glenn H Fredrickson. A molecularly informed field-theoretic study of the complexation of polycation pdadma with mixed micelles of sodium dodecyl sulfate and ethoxylated surfactant. *The European Physical Journal E*, Accepted.

[84] Nicholas Sherck, Kevin Shen, My Nguyen, Brian Yoo, Stephan Kohler, Joshua C Speros, Kris T Delaney, M Scott Shell, and Glenn H Fredrickson. Molecularly informed field theories from bottom-up coarse-graining. *ACS Macro Letters*, 10(5):576–583, 2021.

[85] Scott J Weiner, Peter A Kollman, David A Case, U Chandra Singh, Caterina Ghio, Guliano Alagona, Salvatore Profeta, and Paul Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106(3):765–784, 1984.

[86] Robert B Best, Xiao Zhu, Jihyun Shim, Pedro EM Lopes, Jeetain Mittal, Michael Feig, and Alexander D MacKerell Jr. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *Journal of chemical theory and computation*, 8(9):3257–3273, 2012.

[87] Wilfred F Van Gunsteren and Martin Karplus. Effect of constraints on the dynamics of macromolecules. *Macromolecules*, 15(6):1528–1544, 1982.

[88] William L Jorgensen and Julian Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[89] M Scott Shell. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of chemical physics*, 129(14):144108, 2008.

[90] Aviel Chaimovich and M Scott Shell. Relative entropy as a universal metric for multiscale errors. *Physical Review E*, 81(6):060104, 2010.

[91] Tanmoy Sanyal and M Scott Shell. Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation. *The Journal of chemical physics*, 145(3):034109, 2016.

[92] Saeed Izadi, Ramu Anandakrishnan, and Alexey V Onufriev. Building water models: a different approach. *J. Phys. Chem. Lett.*, 5(21):3863–3871, 2014.

[93] Kris T Delaney and Glenn H Fredrickson. Recent developments in fully fluctuating field-theoretic simulations of polymer melts and solutions. *The Journal of Physical Chemistry B*, 120(31):7615–7634, 2016.

[94] Joshua Lequieu, Trenton Koeper, Kris T Delaney, and Glenn H Fredrickson. Extreme deflection of phase boundaries and chain bridging in a (ba') n miktoarm star polymers. *Macromolecules*, 53(2):513–522, 2020.

[95] Alexander Weyman, Vlasis G Mavrantzas, and Hans Christian Öttinger. Field-theoretic simulations beyond δ-interactions: Overcoming the inverse potential problem in auxiliary field models. *The Journal of Chemical Physics*, 155(2):024106, 2021.

[96] Alexander Weyman, Vlasis G Mavrantzas, and Hans Christian Öttinger. Direct calculation of the functional inverse of realistic interatomic potentials in field-theoretic simulations. *The Journal of Chemical Physics*, 156(22):224115, 2022.

[97] Nicholas JH Dunn and WG Noid. Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids. *The Journal of chemical physics*, 143(24):243148, 2015.

[98] Thomas T Foley, M Scott Shell, and William George Noid. The impact of resolution upon entropy and information in coarse-grained models. *The Journal of chemical physics*, 143(24):12B601_1, 2015.

[99] JW Mullinax and WG Noid. Extended ensemble approach for deriving transferable coarse-grained potentials. *The Journal of Chemical Physics*, 131(10):104110, 2009.

[100] Kevin Shen, Nicholas Sherck, My Nguyen, Brian Yoo, Stephan Köhler, Joshua Speros, Kris T Delaney, Glenn H Fredrickson, and M Scott Shell. Learning composition-transferable coarse-grained models: Designing external potential ensembles to maximize thermodynamic information. *The Journal of Chemical Physics*, 153(15):154116, 2020.

[101] Rungsima Chollakup, John B. Beck, Klaus Dirnberger, Matthew Tirrell, and Claus D. Eisenbach. Polyelectrolyte molecular weight and salt effects on the phase behavior and coacervation of aqueous solutions of poly(acrylic acid) sodium salt and poly(allylamine) hydrochloride. *Macromolecules*, 46(6):2376–2390, 2013.

[102] Mitsuhiro Suda and Tsuneo Okubo. The influence of ph on the alternate multilayered adsorption of macrocations and macroanions on colloidal spheres. *Colloid and Polymer Science*, 282(5):518–523, 2004.

[103] Surita R Bhatia, Sarwat F Khattak, and Susan C Roberts. Polyelectrolytes for cell encapsulation. *Current opinion in colloid & interface science*, 10(1-2):45–51, 2005.

[104] Vikramjit S Rathee, Aristotle J Zervoudakis, Hythem Sidky, Benjamin J Sikora, and Jonathan K Whitmer. Weak polyelectrolyte complexation driven by associative charging. *The Journal of Chemical Physics*, 148(11):114901, 2018.

[105] In Suk Joung and Thomas 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):9020–9041, 2008.

[106] Junmei Wang, M. Wolf Romain, W. Caldwell James, A. Kollman Peter, and A. Case David. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 7 2004.

[107] Junmei Wang, Piotr Cieplak, and Peter A Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *Journal of computational chemistry*, 21(12):1049–1074, 2000.

[108] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N.

Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian˜16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

[109] Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*, 13(7):e1005659, 2017.

[110] Ulrich W Suter. Epimerization of vinyl polymers to stereochemical equilibrium. 1. theory. *Macromolecules*, 14(3):523–528, 1981.

[111] Abhishek K Gupta and Upendra Natarajan. Tacticity effects on conformational structure and hydration of poly-(methacrylic acid) in aqueous solutions-a molecular dynamics simulation study. *Molecular Simulation*, 42(9):725–736, 2016.

[112] Dimitri P Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.

[113] Michael JD Powell. Algorithms for nonlinear constraints that use lagrangian functions. *Mathematical programming*, 14(1):224–248, 1978.

[114] Nicholas Ian Mark Gould. On the convergence of a sequential penalty function method for constrained minimization. *SIAM Journal on Numerical Analysis*, 26(1):107–128, 1989.

[115] Steven W Cranford and Markus J Buehler. Variation of weak polyelectrolyte persistence length through an electrostatic contour length. *Macromolecules*, 45(19):8067–8082, 2012.

[116] Dimitris G Mintis and Vlasis G Mavrantzas. Effect of ph and molecular length on the structure and dynamics of short poly (acrylic acid) in dilute solution: Detailed molecular dynamics study. *The Journal of Physical Chemistry B*, 123(19):4204–4219, 2019.

[117] Athanassios Z Panagiotopoulos, N Quirke, M Stapleton, and DJ Tildesley. Phase equilibria by simulation in the gibbs ensemble: alternative derivation, generalization and application to mixture and membrane equilibria. *Molecular Physics*, 63(4):527–545, 1988.

[118] Sean Friedowitz, Junzhe Lou, Kayla Patricia Barker, Karis Will, Yan Xia, and Jian Qin. Looping-in complexation and ion partitioning in nonstoichiometric polyelectrolyte mixtures. *Science Advances*, 7(31):eabg8654, 2021.

[119] Angelika E Neitzel, Yan N Fang, Boyuan Yu, Artem M Rumyantsev, Juan J de Pablo, and Matthew V Tirrell. Polyelectrolyte complex coacervation across a broad range of charge densities. *Macromolecules*, 54(14):6878–6890, 2021.

[120] Pengfei Zhang, Nayef M Alsaifi, Jianzhong Wu, and Zhen-Gang Wang. Salting-out and salting-in of polyelectrolyte solutions: A liquid-state theory study. *Macromolecules*, 49(24):9720–9730, 2016.

[121] Douglas J Grzetic, Kris T Delaney, and Glenn H Fredrickson. Contrasting dielectric properties of electrolyte solutions with polar and polarizable solvents. *Physical review letters*, 122(12):128007, 2019.

[122] Krisztina Bali, Zsófia Varga, Attila Kardos, Imre Varga, Tibor Gilányi, Attila Domján, András Wacha, Attila Bóta, Judith Mihály, and Róbert Mészáros. Effect of dilution on the nonequilibrium polyelectrolyte/surfactant association. *Langmuir*, 34(48):14652–14660, 2018.

[123] Robert Bradbury, Jeffrey Penfold, Robert K Thomas, Ian M Tucker, Jordan T Petkov, and Craig Jones. Manipulating perfume delivery to the interface using polymer–surfactant interactions. *Journal of colloid and interface science*, 466:220–226, 2016.

[124] Nora Kristen and Regine von Klitzing. Effect of polyelectrolyte/surfactant combinations on the stability of foam films. *Soft Matter*, 6(5):849–861, 2010.

[125] Cornelus G De Kruif, Fanny Weinbreck, and Renko de Vries. Complex coacervation of proteins and anionic polysaccharides. *Current opinion in colloid & interface science*, 9(5):340–349, 2004.

[126] K S Mayya, Amitabha Bhattacharyya, and J-F Argillier. Micro-encapsulation by complex coacervation: Influence of surfactant. *Polymer International*, 52(4):644–647, 2003.

[127] Yulia V Shulevich, Thuy Huu Nguyen, Dmitry S Tutaev, Alexander V Navrotskii, and Ivan A Novakov. Purification of fat-containing wastewater using polyelectrolyte–surfactant complexes. *Separation and Purification Technology*, 113:18–23, 2013.

[128] Manmohan Singh, Maylene Briones, Gary Ott, and Derek O'Hagan. Cationic microparticles: a potent delivery system for dna vaccines. *Proceedings of the National Academy of Sciences*, 97(2):811–816, 2000.

[129] Katumitu Hayakawa and Jan CT Kwak. Surfactant-polyelectrolyte interactions. 1. binding of dodecyltrimethylammonium ions by sodium dextransulfate and sodium poly (styrenesulfonate) in aqueous solution in the presence of sodium chloride. *The Journal of Physical Chemistry*, 86(19):3866–3870, 1982.

[130] Katumitu Hayakawa and Jan CT Kwak. Study of surfactant-polyelectrolyte interactions. 2. effect of multivalent counterions on the binding of dodecyltrimethylammonium ions by sodium dextran sulfate and sodium poly (styrene sulfonate) in aqueous solution. *The Journal of Physical Chemistry*, 87(3):506–509, 1983.

[131] Manuel López-López, Pilar López-Cornejo, Victoria Isabel Martín, Francisco José Ostos, Cintia Checa-Rodríguez, Rosario Prados-Carvajal, José Antonio Lebrón, Pablo Huertas, and María Luisa Moyá. Importance of hydrophobic interactions in the single-chained cationic surfactant-dna complexation. *Journal of colloid and interface science*, 521:197–205, 2018.

[132] Kyrre Thalberg, Bjoern Lindman, and Karin Bergfeldt. Phase behavior of systems of polyacrylate and cationic surfactants. *Langmuir*, 7(12):2893–2898, 1991.

[133] Yilin Wang, Kozue Kimura, Qingrong Huang, Paul L Dubin, and Werner Jaeger. Effects of salt on polyelectrolyte- micelle coacervation. *Macromolecules*, 32(21):7128–7134, 1999.

[134] Leonardo Chiappisi, Ingo Hoffmann, and Michael Gradzielski. Complexes of oppositely charged polyelectrolytes and surfactants–recent developments in the field of biologically derived polyelectrolytes. *Soft Matter*, 9(15):3896–3909, 2013.

[135] Adeline Maria Benhur, Jangelis Diaz, and Samiul Amin. Impact of polyelectrolyte-surfactant interactions on the rheology and wet lubrication performance of conditioning shampoo. *International Journal of Cosmetic Science*, 43(2):246–253, 2021.

[136] Ioannis S Chronakis and Paschalis Alexandridis. Rheological properties of oppositely charged polyelectrolyte- surfactant mixtures: effect of polymer molecular weight and surfactant architecture. *Macromolecules*, 34(14):5005–5018, 2001.

[137] Lutz Maibaum, Aaron R Dinner, and David Chandler. Micelle formation and the hydrophobic effect. *The Journal of Physical Chemistry B*, 108(21):6778–6781, 2004.

[138] Nirmesh Jain, Siwar Trabelsi, Samuel Guillot, Daragh McLoughlin, Dominique Langevin, Pierre Letellier, and Mireille Turmine. Critical aggregation concentration in mixed solutions of anionic polyelectrolytes and cationic surfactants. *Langmuir*, 20(20):8496–8503, 2004.

[139] Jason J Madinya and Charles E Sing. Hybrid field theory and particle simulation model of polyelectrolyte–surfactant coacervation. *Macromolecules*, 55(6):2358–2373, 2022.

[140] Torsten Wallin and Per Linse. Polyelectrolyte-induced micellization of charged surfactants. calculations based on a self-consistent field lattice model. *Langmuir*, 14(11):2940–2949, 1998.

[141] Yingjie Li, Paul L Dubin, Henry A Havel, Shun L Edwards, and Herbert Dautzenberg. Complex formation between polyelectrolyte and oppositely charged mixed micelles: soluble complexes vs coacervation. *Langmuir*, 11(7):2486–2492, 1995.

[142] Markus Antonietti, Juergen Conrad, and Andreas Thuenemann. Polyelectrolyte-surfactant complexes: a new type of solid, mesomorphous material. *Macromolecules*, 27(21):6007–6011, 1994.

[143] Yakov Lapitsky and Eric W Kaler. Surfactant and polyelectrolyte gel particles for encapsulation and release of aromatic oils. *Soft Matter*, 2(9):779–784, 2006.

[144] L Magnus Bergström, UR Mikael Kjellin, Per M Claesson, and Isabelle Grillo. Small-angle neutron scattering study of mixtures of cationic polyelectrolyte and anionic surfactant: Effect of polyelectrolyte charge density. *The Journal of Physical Chemistry B*, 108(6):1874–1881, 2004.

[145] Michael Gradzielski and Ingo Hoffmann. Polyelectrolyte-surfactant complexes (pescs) composed of oppositely charged components. *Current Opinion in Colloid & Interface Science*, 35:124–141, 2018.

[146] Monojoy Goswami, Jose M Borreguero, Philip A Pincus, and Bobby G Sumpter. Surfactant-mediated polyelectrolyte self-assembly in a polyelectrolyte–surfactant complex. *Macromolecules*, 48(24):9050–9059, 2015.

[147] Yaxun Fan, Matthias Kellermeier, Amy Y Xu, Volodymyr Boyko, Sebastian Mirtschin, and Paul L Dubin. Modulation of polyelectrolyte–micelle interactions via zeta potentials. *Macromolecules*, 50(14):5518–5526, 2017.

[148] Dongcui Li, Manish S Kelkar, and Norman J Wagner. Phase behavior and molecular thermodynamics of coacervation in oppositely charged polyelectrolyte/surfactant systems: A cationic polymer jr 400 and anionic surfactant sds mixture. *Langmuir*, 28(28):10348–10362, 2012.

[149] Ghazi Ben Messaoud, Lyndsay Promeneur, Martha Brennich, Sophie LKW Roelants, Patrick Le Griel, and Niki Baccile. Complex coacervation of natural sophorolipid bolaamphiphile micelles with cationic polyelectrolytes. *Green Chemistry*, 20(14):3371–3385, 2018.

[150] Francois Guillemet and Lennart Piculell. Interactions in aqueous mixtures of hydrophobically modified polyelectrolyte and oppositely charged surfactant. mixed micelle formation and associative phase separation. *The Journal of Physical Chemistry*, 99(22):9201–9209, 1995.

[151] Anna Svensson, Lennart Piculell, Bernard Cabane, and Philippe Ilekti. A new approach to the phase behavior of oppositely charged polymers and surfactants. *The Journal of Physical Chemistry B*, 106(5):1013–1018, 2002.

[152] M Muthukumar. 50th anniversary perspective: A perspective on polyelectrolyte solutions. *Macromolecules*, 50(24):9528–9560, 2017.

[153] Ebru Kizilay, A Basak Kayitmazer, and Paul L Dubin. Complexation and coacervation of polyelectrolytes with oppositely charged colloids. *Advances in colloid and interface science*, 167(1-2):24–37, 2011.

[154] Gunja Pandav, Victor Pryamitsyn, Jeffrey Errington, and Venkat Ganesan. Multibody interactions, phase behavior, and clustering in nanoparticle–polyelectrolyte mixtures. *The Journal of Physical Chemistry B*, 119(45):14536–14550, 2015.

[155] Margarita Antonov, Malek Mazzawi, and Paul L Dubin. Entering and exiting the protein- polyelectrolyte coacervate phase via nonmonotonic salt dependence of critical conditions. *Biomacromolecules*, 11(1):51–59, 2010.

[156] Allie C Obermeyer, Carolyn E Mills, Xue-Hui Dong, Romeo J Flores, and Bradley D Olsen. Complex coacervation of supercharged proteins with polyelectrolytes. *Soft Matter*, 12(15):3570–3581, 2016.

[157] Fatih Comert and Paul L Dubin. Liquid-liquid and liquid-solid phase separation in protein-polyelectrolyte systems. *Advances in Colloid and Interface Science*, 239:213–217, 2017.

[158] Paul L Dubin, Mary E Curran, and Jiadong Hua. Critical linear charge density for binding of a weak polycation to an anionic/nonionic mixed micelle. *Langmuir*, 6(3):707–709, 1990.

[159] Kevin Shen and Zhen-Gang Wang. Electrostatic correlations and the polyelectrolyte self energy. *The Journal of chemical physics*, 146(8):084901, 2017.

[160] Pengfei Zhang, Kevin Shen, Nayef M Alsaifi, and Zhen-Gang Wang. Salt partitioning in complex coacervation of symmetric polyelectrolytes. *Macromolecules*, 51(15):5586–5593, 2018.

[161] Tyler K Lytle and Charles E Sing. Transfer matrix theory of polymer complex coacervation. *Soft Matter*, 13(39):7001–7012, 2017.

[162] Li-Wei Chang, Tyler K Lytle, Mithun Radhakrishna, Jason J Madinya, Jon Vélez, Charles E Sing, and Sarah L Perry. Sequence and entropy-based control of complex coacervates. *Nature communications*, 8(1):1273, 2017.

[163] Artem M Rumyantsev, Nicholas E Jackson, and Juan J De Pablo. Polyelectrolyte complex coacervates: Recent developments and new frontiers. *Annual Review of Condensed Matter Physics*, 12:155–176, 2021.

[164] Artem M Rumyantsev, Albert Johner, Matthew V Tirrell, and Juan J de Pablo. Unifying weak and strong charge correlations within the random phase approximation: Polyampholytes of various sequences. *Macromolecules*, 55(14):6260–6274, 2022.

[165] Charles E Sing and Sarah L Perry. Recent progress in the science of complex coacervation. *Soft Matter*, 16(12):2885–2914, 2020.

[166] Tyler K Lytle and Charles E Sing. Tuning chain interaction entropy in complex coacervation using polymer stiffness, architecture, and salt valency. *Molecular Systems Design & Engineering*, 3(1):183–196, 2018.

[167] Sean Friedowitz and Jian Qin. Reversible ion binding for polyelectrolytes with adaptive conformations. *AIChE Journal*, 67(12):e17426, 2021.

[168] Sean Friedowitz, Ali Salehi, Ronald G Larson, and Jian Qin. Role of electrostatic correlations in polyelectrolyte charge association. *The Journal of chemical physics*, 149(16):163335, 2018.

[169] Artem M Rumyantsev, Oleg V Borisov, and Juan J de Pablo. Structure and dynamics of hybrid colloid–polyelectrolyte coacervates. *Macromolecules*, 2023.

[170] Robert D Groot. Electrostatic interactions in dissipative particle dynamics—simulation of polyelectrolytes and anionic surfactants. *The Journal of chemical physics*, 118(24):11265–11277, 2003.

[171] Jose M Borreguero, Philip A Pincus, Bobby G Sumpter, and Monojoy Goswami. Unraveling the agglomeration mechanism in charged block copolymer and surfactant complexes. *Macromolecules*, 50(3):1193–1205, 2017.

[172] Per Hansson. Phase behavior of aqueous polyion–surfactant ion complex salts: A theoretical analysis. *Journal of colloid and interface science*, 332(1):183–193, 2009.

[173] Boyao Wen, Bofeng Bai, and Ronald 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*, 599:773–784, 2021.

[174] Richard Becker and Werner Döring. Kinetische behandlung der keimbildung in übersättigten dämpfen. *Annalen der physik*, 416(8):719–752, 1935.

[175] Joshua A Mysona, Alon V McCormick, and David C Morse. Mechanism of micelle birth and death. *Physical Review Letters*, 123(3):038003, 2019.

[176] EAG Aniansson and S N_ Wall. Kinetics of step-wise micelle association. *The Journal of Physical Chemistry*, 78(10):1024–1030, 1974.

[177] Benjamin G Levine, David N LeBard, Russell DeVane, Wataru Shinoda, Axel Kohlmeyer, and Michael L Klein. Micellization studied by gpu-accelerated coarse-grained molecular dynamics. *Journal of Chemical Theory and Computation*, 7(12):4135–4145, 2011.

[178] David N LeBard, Benjamin G Levine, Philipp Mertmann, Stephen A Barr, Arben Jusufi, Samantha Sanders, Michael L Klein, and Athanassios Z Panagiotopoulos. Self-assembly of coarse-grained ionic surfactants accelerated by graphics processing units. *Soft Matter*, 8(8):2385–2397, 2012.

[179] Marie Jonsson and Per Linse. Polyelectrolyte–macroion complexation. i. effect of linear charge density, chain length, and macroion charge. *The Journal of Chemical Physics*, 115(7):3406–3418, 2001.

[180] Torsten Wallin and Per Linse. Monte carlo simulations of polyelectrolytes at charged micelles. 3. effects of surfactant tail length. *The Journal of Physical Chemistry B*, 101(28):5506–5513, 1997.

[181] RJ Allen and PB Warren. Phase behaviour of oppositely charged polymer/surfactant mixtures. *Europhysics Letters*, 64(4):468, 2003.

[182] Kevin Shen, My Nguyen, Nicholas Sherck, Brian Yoo, Stephan Köhler, Joshua Speros, Kris T Delaney, M Scott Shell, and Glenn H Fredrickson. Predicting surfactant phase behavior with a molecularly informed field theory. *Journal of Colloid and Interface Science*, 638:84–98, 2023.

[183] Callum J Dickson, Benjamin D Madej, Åge A Skjevik, Robin M Betz, Knut Teigen, Ian R Gould, and Ross C Walker. Lipid14: the amber lipid force field. *Journal of chemical theory and computation*, 10(2):865–879, 2014.

[184] Hui Yan, Shi-Ling Yuan, Gui-Ying Xu, and Cheng-Bu Liu. Effect of ca2+ and mg2+ ions on surfactant solutions investigated by molecular dynamics simulation. *Langmuir*, 26(13):10448–10459, 2010.

[185] Leandro Martínez, Ricardo Andrade, Ernesto G Birgin, and José Mario Martínez. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry*, 30(13):2157–2164, 2009.

[186] Donald W McQuigg, Jerome I Kaplan, and Paul L Dubin. Critical conditions for the binding of polyelectrolytes to small oppositely charged micelles. *The Journal of Physical Chemistry*, 96(4):1973–1978, 1992.

[187] Jean-Louis Barrat, Glenn H Fredrickson, and Scott W Sides. Introducing variable cell shape methods in field theory simulations of polymers. *The Journal of Physical Chemistry B*, 109(14):6694–6700, 2005.

[188] D John Mitchell, Gordon JT Tiddy, Loraine Waring, Theresa Bostock, and Malcolm P McDonald. Phase behaviour of polyoxyethylene surfactants with water. mesophase structures and partial miscibility (cloud points). *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 79(4):975–1000, 1983.

[189] Ole G Mouritsen and Kent Jørgensen. A new look at lipid-membrane structure in relation to drug research. *Pharmaceutical research*, 15:1507–1519, 1998.

[190] Samantha A Sanders and Athanassios Z Panagiotopoulos. Micellization behavior of coarse grained surfactant models. *The Journal of chemical physics*, 132(11):114902, 2010.

[191] Ming-Tsung Lee, Aleksey Vishnyakov, and Alexander V Neimark. Calculations of critical micelle concentration by dissipative particle dynamics simulations: the role of chain rigidity. *The Journal of Physical Chemistry B*, 117(35):10304–10310, 2013.

[192] Aleksey Vishnyakov, Ming-Tsung Lee, and Alexander V Neimark. Prediction of the critical micelle concentration of nonionic surfactants by dissipative particle dynamics simulations. *The journal of physical chemistry letters*, 4(5):797–802, 2013.

[193] Arben Jusufi and Athanassios Z Panagiotopoulos. Explicit-and implicit-solvent simulations of micellization in surfactant solutions. *Langmuir*, 31(11):3283–3292, 2015.

[194] Andrew P Santos and Athanassios Z Panagiotopoulos. Determination of the critical micelle concentration in simulations of surfactant systems. *The Journal of chemical physics*, 144(4):044709, 2016.

[195] Jiajia Zhou and An-Chang Shi. Critical micelle concentration of micelles with different geometries in diblock copolymer/homopolymer blends. *Macromolecular theory and simulations*, 20(8):690–699, 2011.

[196] Shiyi Qin, Tianyi Jin, Reid C Van Lehn, and Victor M Zavala. Predicting critical micelle concentrations for surfactants using graph convolutional neural networks. *The Journal of Physical Chemistry B*, 125(37):10610–10620, 2021.

[197] Vishnu Sresht, Eric P Lewandowski, Daniel Blankschtein, and Arben Jusufi. Combined molecular dynamics simulation–molecular-thermodynamic theory framework for predicting surface tensions. *Langmuir*, 33(33):8319–8329, 2017.

[198] M Lechuga, M Fernández-Serrano, E Jurado, J Núñez-Olea, and F Ríos. Acute toxicity of anionic and non-ionic surfactants to aquatic organisms. *Ecotoxicology and environmental safety*, 125:1–8, 2016.

[199] Samantha A Sanders, Maria Sammalkorpi, and Athanassios Z Panagiotopoulos. Atomistic simulations of micellization of sodium hexyl, heptyl, octyl, and nonyl sulfates. *The Journal of Physical Chemistry B*, 116(8):2430–2437, 2012.

[200] Frank H Quina, Patricia M Nassar, Joao BS Bonilha, and Barney L Bales. Growth of sodium dodecyl sulfate micelles with detergent concentration. *The Journal of Physical Chemistry*, 99(46):17028–17031, 1995.

[201] Barney L Bales. A definition of the degree of ionization of a micelle based on its aggregation number. *The Journal of Physical Chemistry B*, 105(29):6798–6804, 2001.

[202] Annalaura Del Regno, Patrick B Warren, David J Bray, and Richard L Anderson. Critical micelle concentrations in surfactant mixtures and blends by simulation. *The Journal of Physical Chemistry B*, 125(22):5983–5990, 2021.

[203] Sarah H Klass, Jamie M Gleason, Anthony O Omole, Bibiana Onoa, Carlos J Bustamante, and Matthew B Francis. Preparation of bioderived and biodegradable surfactants based on an intrinsically disordered protein sequence. *Biomacromolecules*, 23(3):1462–1470, 2022.

[204] Jiayu Xie and An-Chang Shi. Formation of complex spherical packing phases in diblock copolymer/homopolymer blends. *Giant*, 5:100043, 2021.

[205] Yanyan Zhu, Bin Zheng, Liangshun Zhang, David Andelman, and Xingkun Man. Formation of diblock copolymer nanoparticles: Theoretical aspects. *Giant*, 10:100101, 2022.

[206] Tanmoy Sanyal, Jeetain Mittal, and M Scott Shell. A hybrid, bottom-up, structurally accurate, g o^{-}-like coarse-grained protein model. *The Journal of chemical physics*, 151(4):044111, 2019.

[207] Paul Robustelli, Stefano Piana, and David E Shaw. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences*, 115(21):E4758–E4766, 2018.

[208] R 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 c12e5and c12e8. *Journal of colloid and interface science*, 197(2):370–376, 1998.

[209] Diego Romano Perinelli, Marco Cespi, Nicola Lorusso, Giovanni Filippo Palmieri, Giulia Bonacucina, and Paolo Blasi. Surfactant self-assembling and critical micelle concentration: one approach fits all? *Langmuir*, 36(21):5745–5753, 2020.

[210] Paul J Flory. Thermodynamics of high polymer solutions. *The Journal of chemical physics*, 9(8):660–660, 1941.

[211] Maurice L Huggins. Solutions of long chain compounds. *The Journal of chemical physics*, 9(5):440–440, 1941.

[212] Pierre-Gilles De Gennes and Pierre-Gilles Gennes. *Scaling concepts in polymer physics*. Cornell university press, 1979.

[213] Ludwik Leibler, Henri Orland, and John C Wheeler. Theory of critical micelle concentration for solutions of block copolymers. *The Journal of chemical physics*, 79(7):3550–3557, 1983.

[214] A Ben-Naim and FH Stillinger. Critical micelle concentration and the size distribution of surfactant aggregates. *The Journal of Physical Chemistry*, 84(22):2872–2876, 1980.

[215] M Antonio Floriano, Eugenio Caponetti, and Athanassios Z Panagiotopoulos. Micellization in model surfactant systems. *Langmuir*, 15(9):3143–3151, 1999.

[216] Aniket Bhattacharya and SD Mahanti. Energy and size fluctuations of amphiphilic aggregates in a lattice model. *Journal of Physics: Condensed Matter*, 12(28):6141, 2000.

[217] P 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(16):6054–6058, 2002.

[218] David L Mobley and J Peter Guthrie. Freesolv: a database of experimental and calculated hydration free energies, with input files. *Journal of computer-aided molecular design*, 28:711–720, 2014.

[219] Joshua Lequieu. Combining particle and field-theoretic polymer models with multi-representation simulations. *The Journal of Chemical Physics*, 158(24), 2023.

[220] Dieter Klemm, Brigitte Heublein, Hans-Peter Fink, and Andreas Bohn. Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte chemie international edition*, 44(22):3358–3393, 2005.

[221] Jinhui Pang, Xin Liu, Jun Yang, Fachuang Lu, Bo Wang, Feng Xu, Mingguo Ma, and Xueming Zhang. Synthesis of highly polymerized water-soluble cellulose acetate by the side reaction in carboxylate ionic liquid 1-ethyl-3-methylimidazolium acetate. *Scientific reports*, 6(1):33725, 2016.

[222] Th Heinze and T Liebert. Celluloses and polyoses/hemicelluloses. 2012.

[223] Kenji Kamide, Kunihiko Okajima, Keisuke Kowsaka, and Toshihiko Matsui. Solubility of cellulose acetate prepared by different methods and its correlationships with average acetyl group distribution on glucopyranose units. *Polymer journal*, 19(12):1405–1412, 1987.

[224] Takeaki Miyamoto, Yukiko Sato, Toru Shibata, Mitsuhiko Tanahashi, and Hiroshi Inagaki. 13c-nmr spectral studies on the distribution of substituents in water-soluble cellulose acetate. *Journal of Polymer Science: Polymer Chemistry Edition*, 23(5):1373–1381, 1985.

[225] Dong Tian, Yangyang Han, Canhui Lu, Xinxing Zhang, and Guiping Yuan. Acidic ionic liquid as "quasi-homogeneous" catalyst for controllable synthesis of cellulose acetate. *Carbohydrate polymers*, 113:83–90, 2014.

[226] Silvia Gomez-Bujedo, Etienne Fleury, and Michel R Vignon. Preparation of cellouronic acids and partially acetylated cellouronic acids by tempo/naclo oxidation of water-soluble cellulose acetate. *Biomacromolecules*, 5(2):565–571, 2004.

[227] Jonas Wolfs and Michael AR Meier. A more sustainable synthesis approach for cellulose acetate using the dbu/co 2 switchable solvent system. *Green Chemistry*, 23(12):4410–4420, 2021.

[228] Kenji Kamide, Masatoshi Saito, and Tatsuyuki Abe. Dilute solution properties of water-soluble incompletely substituted cellulose acetate. *Polymer Journal*, 13(5):421–431, 1981.

[229] Carsten Deus, Horst Friebolin, and Egon Siefert. Partiell acetylierte cellulose—synthese und bestimmung der substituentenverteilung mit hilfe der 1h nmr-spektroskopie. *Die Makromolekulare Chemie: Macromolecular Chemistry and Physics*, 192(1):75–83, 1991.

[230] Carlton L Crane. Method of preparing far-hydrolyzed cellulose esters, August 24 1943. US Patent 2,327,770.

[231] Olgun Guvench, Sairam S Mallajosyula, E Prabhu Raman, Elizabeth Hatcher, Kenno Vanommeslaeghe, Theresa J Foster, Francis W Jamison, and Alexander D MacKerell Jr. Charmm additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling. *Journal of chemical theory and computation*, 7(10):3162–3180, 2011.

[232] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[233] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[234] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[235] Edward W Forgy. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *biometrics*, 21:768–769, 1965.

[236] Charles M Buchanan, Kevin J Edgar, and Alan K Wilson. Preparation and characterization of cellulose monoacetates: the relationship between structure and water solubility. *Macromolecules*, 24(11):3060–3064, 1991.

[237] Zoltan Mester and Athanassios Z Panagiotopoulos. Mean ionic activity coefficients in aqueous nacl solutions from molecular dynamics simulations. *The Journal of chemical physics*, 142(4):044507, 2015.

[238] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[239] Edmund D Sneed and Robert L Folk. Pebbles in the lower colorado river, texas a study in particle morphogenesis. *The Journal of Geology*, 66(2):114–150, 1958.