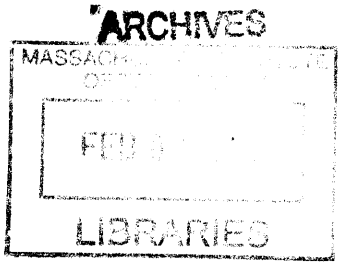
Theoretical and Simulations-Based Modeling of Micellization in Linear and Branched Surfactant Systems



by

Jonathan D. Mendenhall

B.S., Pennsylvania State University (2004)
M.S.CEP, Massachusetts Institute of Technology (2006)

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2012

Signature of Author .....
Department of Chemical Engineering
14 August 2012

Certified by .....
Daniel Blankschtein
Professor of Chemical Engineering
Thesis Supervisor

Accepted by .....
Patrick S. Doyle
Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Theoretical and Simulations-Based Modeling of Micellization in Linear and Branched Surfactant Systems

by

Jonathan D. Mendenhall

Submitted to the Department of Chemical Engineering
on 14 August 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Surfactants are chemically-heterogeneous molecules possessing hydrophilic (head) and hydrophobic (tail) moieties. This dual nature of surfactants leads to interesting phase behavior in aqueous solution as a function of surfactant concentration, including: (i) formation of surfactant monolayers at surfaces and interfaces, and (ii) self-assembly into finite aggregates (micelles) in the bulk solution beyond the critical micelle concentration (cmc). This concentration-dependent phase behavior induces changes in solution properties. For example, the surface activity of surfactants can decrease the surface tension, and self-assembly in bulk solution can lead to changes in viscosity, equivalent conductivity, solubilization capacity, and other bulk properties. These effects make surfactants quite attractive and unique for use in product formulations, where they are utilized as detergents, dispersants, emulsifiers, solubilizers, surface and interfacial tension modifiers, and in other contexts.

The specific chemical structure of the surfactant head and tail is essential in determining the overall performance properties of a surfactant in aqueous media. The surfactant tail drives the self-assembly process through the hydrophobic effect, while the surfactant head imparts a certain extent of solubility to the surfactant in aqueous solution through preferential interactions with the hydrogen-bonding network of water. The interplay between these two effects gives rise to the particular phase diagram of a surfactant, including the specific cmc at which micelles begin to form. In addition to serving as a quantitative indicator of micelle formation, the cmc represents a limit to surface monolayer formation, and hence to surface and interfacial tension reduction, because surfactant adsorption at interfaces remains approximately constant beyond the cmc. In addition, the cmc represents the onset of changes in bulk solution properties.

This Thesis is concerned with the prediction of cmc's and other micellization properties for a variety of linear and branched surfactant chemical architectures which are commonly encountered in practice. Single-component surfactant solutions are investigated, in order to clarify the specific contributions of the surfactant head and tail to the free energy of micellization, a quantity which determines the cmc and all other

aspects of micellization. First, a molecular-thermodynamic (MT) theory is presented which makes use of bulk-phase thermodynamics and a phenomenological thought process to describe the energetics related to the formation of a micelle from its constituent surfactant monomers. Second, a combined computer-simulation/molecular-thermodynamic (CSMT) framework is discussed which provides a more detailed quantification of the hydrophobic effect using molecular dynamics simulations. A novel computational strategy to identify surfactant head and tail using an iterative dividing surface approach, along with simulated micelle results, is proposed. Force-field development for novel surfactant structures is also discussed. Third, a statistical-thermodynamic, single-chain, mean-field theory for linear and branched tail packing is formulated, which enables quantification of the specific energetic penalties related to confinement and constraint of surfactant tails within micelles. Finally, these theoretical and simulations-based strategies are used to predict the micellization behavior of 55 linear surfactants and 28 branched surfactants. Critical micelle concentration and optimal micelle properties are reported and compared with experiment, demonstrating good agreement across a range of surfactant head and tail types. In particular, the CSMT framework is found to provide improved agreement with experimental cmc's for the branched surfactants considered.

Thesis Supervisor: Daniel Blankschtein
Title: Professor of Chemical Engineering

Contents

List of Figures

List of Tables

Acknowledgments

I will always remember my time at MIT as a period of great personal and professional growth, during which I was able to balance work with travel (to Europe, China, Canada, and the Western US), athletics (squash, floorball, and summer softball), and music (guitar). My work and interests have brought me in contact with many inspiring individuals over the years – professors, colleagues, friends, and even those met briefly in chance encounters on campus or on the road. I am grateful for all the people who have played some role in shaping me into the person I am today, and I will remember each fondly.

The person who has had the greatest impact on me over the years is Professor Daniel Blankschtein. Professor Blankschtein, as my thesis advisor, helped guide me in my research, inspired in me a great interest in colloid and interface science and thermodynamics, and demonstrated through his own example the important role of teaching and mentorship in a chemical engineering education. Professor Blankschtein was also a good friend to me, someone with whom I could chat regarding personal matters and interests, and who was always willing to provide advice and guidance. I am proud to have had him as my advisor and mentor, and this Thesis would certainly not have been possible without his tireless effort and support.

My thesis committee members, Professors Alan Hatton and Greg Rutledge, also played an important role in providing guidance during my research, challenging me to test various aspects of my theories, and providing suggestions for further improvement. I am thankful to have had the opportunity to benefit from their knowledge regarding theoretical modeling and computer simulations.

I would also like to thank Professors Karen Gleason and Narendra Maheshri, for whom I worked as a teaching assistant in 10.213 (Undergraduate Chemical Engineering Thermodynamics). As my first experience in teaching, this course provided me with the opportunity to observe the classroom process from the other side, and the challenges associated with communicating important engineering concepts in a manner which students from diverse backgrounds can all appreciate and understand.

Another important aspect of my education at MIT was my participation in the David H. Koch School of Chemical Engineering Practice, a rewarding four-month program involving intensive projects at General Mills and Novartis. I would like to thank Professor Hatton, as the director of this program, and Bob Fisher and Steve Martin (now a professor at Virginia Tech), who served as station directors. This was an important experience for me, providing me with the opportunity to work on experimental and theoretical projects of direct relevance to the food science and pharmaceutical industries. I also learned a lot from the other students participating in the program – Curtiss Schneider, Arman Haidari, Saurabh Tejwani, Jie Chen, Fei Chen, Aruna Mohan, and Terence Sio. Our teamwork allowed us to tackle some nice problems.

One important aspect of any research effort involves the people with whom one is surrounded. I consider myself quite fortunate to have benefitted from my many interactions with labmates in the Blankschtein Group: Henry Lam, Priscila Mazzola, Joe Kushner, Saswata Ghosh, Arthur Goldsipe, Brian Stephenson, Srinivas Moorkanikkara, Baris Polat, Jen Seto, Shangchao Lin, Chih-Jen Shih, Jaisree Iyer, Diviya Sinha, Carl Schoellhammer, and Vishnu Sresht. Arthur and Brian, in particular, were great senior labmates to me, and I recall fondly our time together. Arthur imparted to me a great love for learning languages and an appreciation for live music and supporting local musicians. Brian and I shared the experience of exploring Canada (in particular, Nova Scotia, New Brunswick, and Prince Edward Island), and I consider him a close friend. Baris and Shangchao have also been great friends over the years: Baris has opened up his home to me on several occasions and been a confidant, and Shangchao has been an excellent collaborator and roommate. In addition, Shangchao provided the initial push for me to visit him and his family in China, which was an exciting experience. I wish all labmates and their families the best as their careers develop.

It has been an additional pleasure to interact with visitors to the Blankschtein lab over the years, including visiting Professors (and former Blankschtein group students) Leo Lue and Pak Yuet, and other visitors Isaac Reif, Carlota Yagui, Renata

Lopez, and Lingling Zhao. Leo was instrumental in fostering my interest in Monte Carlo techniques. Lingling provided me with an opportunity for collaboration on a simulations-based project, and she and her family later hosted me when I visited their home city of Nanjing, China. It has also been nice to meet former students during their brief return visits to MIT, including Sudhakar Puvvada, Anat Shiloach, and Ayal Naor. In particular, I have enjoyed having several conversations about music and recording with Ayal.

There are many others to thank, too numerous to mention here, including MIT faculty members, the Student Services Office staff, the ASO staff, classmates, roommates, teammates, squash partners, collaborators, and fellow musicians.

I would like to acknowledge in closing my family: my dad Dave, mom Joanne, sister Renee, and brother Dan. Their support, and the support of grandparents, aunts, uncles, and cousins, has been very important to me, and has enabled me to reach this point.

Finally, I would like to dedicate this Thesis to the memory of a good friend from the past, David Pierce.

Chapter 1

Introduction

1.1 Motivation and Background

1.1.1 Overview

In this Thesis, I will focus on the theoretical and simulations-based prediction of surfactant micellization behavior. In the Sections which follow, I will provide a summary of the general motivation and background for this study. Throughout this Thesis, I will provide additional motivation and in-depth background material in the relevant Chapters and Appendices. Chapters 2 and 4 include additional theoretical background, while Chapter 3 includes additional theoretical and simulations background. Chapters 5 and 6 provide additional motivation for the study of specific surfactant systems. Appendices A through E together provide a substantial simulations background, while Appendix F provides a background for molecular property prediction. Please refer to Section 1.2 for an overview of the contents of these Chapters and Appendices. As a result of this Thesis structure, I have kept the contents of this Introduction brief.

In Section 1.1.2, I provide a definition for a surfactant (surface active agent) and discuss the origin of the equilibrium solution behavior of surfactants in aqueous solution, including self-assembly of surfactants at certain concentrations into aggregates called micelles. In Section 1.1.3, I provide a discussion of the types of theories

that have been introduced to help elucidate self-assembly behavior (in particular, micellization behavior) in aqueous media. Finally in Section 1.1.4, I provide a brief selection and description of simulation studies aimed at providing a more microscopic description of the self-assembly (specifically, micellization) behavior of surfactants. The literature on surfactants is voluminous, due to their industrial importance, and I will only attempt to provide a brief overview here of what has been reported in the literature.

1.1.2 Surfactants and Their Practical Use

Surfactants are low-molecular weight, surface-active molecules that exhibit an interesting phase behavior in aqueous solution, due to their chemical heterogeneity.[12] A conventional surfactant possesses one or several contiguous hydrophilic moieties, often referred to as the surfactant head, and one or several contiguous hydrophobic moieties, often referred to as the surfactant tail.[13] Surfactants possessing multiple heads separated by one or more tails (bolaform surfactants[14,15]) or possessing multiple tails separated by one or more heads (gemini surfactants[16,17]) have been synthesized and characterized. However, these more complex surfactants are less commonly encountered in practice relative to surfactants possessing a single head connected to a single tail,[12] and are considered specialty surfactants, due to their expense.[18] Note that higher-molecular weight, polymeric materials, such as diblock copolymers, can also behave like surfactants, depending on the nature of the polymer blocks selected.[19] However, these materials are outside the scope of this Thesis.

Surfactants are typically classified by the functional nature of their head groups, which are either charged (specifically, anionic,[20] cationic,[21] or zwitterionic/amphoteric[22]) or nonionic (e.g., the alkyl ethoxylates or sugar-based surfactants such as alkyl glucoside or alkyl maltoside).[23] Certain surfactants, such as the alkyl dimethyl amine oxides, may be either charged or nonionic, depending on the solution pH.[24,25] Similarly, the charges of certain zwitterionic surfactants may also be pH-dependent (e.g., the alkyl aminopropionic acids), although surfactants such as alkyl sulfobetaine do not exhibit such pH-sensitivity.[12]

Surfactant tail groups typically comprise linear or branched, hydrocarbon-based, aliphatic chains.[12,26] Aromatic groups, including benzene[27] and naphthalene[28] rings, may also be present. Alternative tail chemistries include the fluorocarbons[29] and silicon oxides.[30] Hybrid surfactants have also been synthesized, possessing both hydrocarbon and fluorocarbon regions.[31]

The surfactant tail usually lacks any capability for hydrogen-bonding. In aqueous solution, then, the hydrogen-bonding network of water is disrupted by the presence of the tail, in a size-dependent manner.[32,33] This results in an entropic penalty, since fewer hydrogen bonds can be accommodated for water molecules which are located very close to the hydrophobic material.[32] This penalty can be reduced only if the surfactant tail is somehow partially removed from contact with the aqueous solution.[34] To this end: (i) the surfactant can migrate from the bulk to an available surface or interface, whereupon the exposure of the tail to the bulk solution can be limited upon adsorption through the formation of a surfactant monolayer,[34] (ii) the surfactant can aggregate with other surfactants to form finite aggregates dispersed in bulk solution called micelles, whereupon the surfactant tail is partially shielded in the interior or core of the micelle, and the remaining hydrophobic surface area exposed at the micelle core-water interface is reduced relative to the surface area exposed in the case of the freely-dispersed surfactant,[35] (iii) at high concentrations, the surfactant can associate with other surfactants to form bicontinuous, liquid crystalline, or other types of mesophases,[36] and (iv) the surfactant can precipitate out of solution.[37,38] Option (i) will occur even at low surfactant concentrations, while options (ii) requires that a certain threshold concentration, known as the critical micelle concentration (cmc) be reached.[35] Option (iii) typically can be viewed as a manifestation of intermicellar interactions which occur as the density of micelles in solution increase with increasing total surfactant concentration. For example, initially spherical micelles may begin to elongate to form rods, which then become close-packed and ordered in a hexagonal arrangement, which then form lamellar structures alternating with regions containing solvent.[36] Option (iv) can occur at high concentrations as the solution becomes saturated with aggregates, but will also sometimes occur at low concentrations, below

that necessary to observe options (ii) or (iii), for temperatures beneath a certain threshold known as the Krafft point.[13] The Krafft point is strongly dependent on the nature of the specific chemical structure of the surfactant under consideration, being generally reduced for branched surfactants,[39,40] which do not pack as readily within the crystalline form of the precipitate.[41]

This overall propensity for a surfactant to self-associate in aqueous media due to entropic considerations is generally referred to as the hydrophobic effect.[34] Surfactant self-association due to the hydrophobic effect has important consequences on aqueous solution behavior and, as a result, formulation performance.

At surfaces (e.g., air/water), the formation of a surfactant monolayer generally leads to a reduced equilibrium surface tension for the surfactant-laden surface relative to the clean (i.e., surfactant-free) surface.[42,43] A similar phenomenon is observed at liquid-liquid (e.g., oil/water) interfaces.[44] Surfactant adsorption to solid surface also has consequences, with wetting and spreading behavior affected by a change in spreading coefficient due to surface and interfacial tension modification.[36] Interfacial tension modification has important applications in areas such as enhanced oil recovery,[45] the cleaning of surfaces,[46] and the spreading of ink-jet inks on print media.[47]

In the bulk, surfactant aggregation may lead to an increase in solution viscosity depending on the shape of the micelles formed,[48] an increase in the effective solubility of sparingly-soluble organic solutes,[49] and changes in solution properties such as equivalent conductivity, osmotic pressure, and, in some cases, turbidity.[50] Micelle formation in the bulk also sets an effective limit to the surface tension reduction performance of surfactants, since addition of more surfactant to the solution upon reaching the cmc leads to more aggregation, rather than any substantial change in the surface monolayer in a single-component surfactant solution.[12]

The amphiphilic nature of surfactants, due to the presence of hydrophilic head and hydrophobic tail groups, has proven to make them invaluable for use in formulations. Surfactants are used widely as detergents,[51] dispersants,[52] emulsifiers,[53] solubilizers,[49] surface and interfacial tension modifiers,[54] and foaming and anti-foaming agents,[55] as well as in other advanced technology applications.[36] However, the specific nature of

the surfactant head and tail is important in determining which surfactants are most suitable for a particular application. For example various studies have focused on the role of branching in changing aggregation properties[56–61] and interfacial behavior.[26,27,59,62,63]

1.1.3 Theoretical Models of Micellization

Based on the discussion in Section 1.1.2, it is clear that surfactants exhibit a rich and complex phase behavior that has practical ramifications in the development of industrial formulations. Most such formulations are multi-component in practice,[64] and may contain additives other than surfactants,[51] which can in turn affect the thermodynamics and aggregation behavior of the system.[36,49] However, it is important to first gain an understanding of surfactants in single-component systems in order to fully understand the structure-property relationships which govern their behavior.[65] One can note that even the most basic model of mixed surfactant micellization, the ideal solution theory, relies on knowledge of pure cmc's in order to make predictions,[66] as does the regular solution theory of Holland and Rubingh.[67]

Several statistical models aimed at establishing relationships between micellization properties and various surfactant molecular descriptors have been developed in the form of qualitative structure-property relationships (QSPR's).[68–70] Although the fitting using these approaches can be excellent, it is difficult to understand fundamentally the reason that a particular surfactant structure yielding a particular cmc. Further, it is not always clear how to extend the model to new surfactant systems without re-fitting, depending on the particular molecular descriptors chosen.

More promising are the theoretical models that have been developed. These include various lattice models[71] and off-lattice models.[72] An example of effective lattice models include self-consistent field (SCF) theories[73–75] based on the work of Scheutjens and Fleer,[76] which extends the Flory-Huggins approach[77] to inhomogeneous systems such as micelles. This approach requires the determination of Flory-Huggins χ parameters between beads representing groups in the surfactant head and tail and beads representing the solvent. In their use of this approach, Hurter and coworkers first de-

termined micelle aggregation numbers using small-system thermodynamics,[78,79] followed by application of SCF theory to obtain segment distributions.[74] Focusing on application to block copolymer micelles, they were able to use this approach to assess not only micellization properties such as micelle size, from the segment distributions, but also solubilization properties, such as solute partitioning, which could then be favorably compared with octanol-water partitioning data from experiment. In another application, Jódar-Reyes and coworkers applied the SCF approach to the prediction of micellization properties for branched ethoxylated surfactants,[75] utilizing both a simplified linear model and a more realistic branched model, in which they discriminated between CH_3 and CH_2 groups in defining beads for the tail regions. Although the linear model produced accurate cmc's, the more realistic branched model did not demonstrate agreement, although qualitative comparison to experimental adsorption data at surfaces is improved.

Another important category of models, which forms the basis for my computational framework in Chapter 2, are the molecular-thermodynamic (MT) models, which combine the equations of bulk phase thermodynamics with molecular details of the surfactants considered, based on an identification of the surfactant head and tail. The work of Israelachvili, Mitchell, and Ninham provides an early look at this type of theory,[72] in which they propose a simple model for the reference-state chemical potential of surfactant aggregates and monomers in order to compute the mole fractions of aggregates in solution as a function of aggregation number, yielding the micelle size and shape distribution. Nagarajan and Ruckenstein were later able to demonstrate[80] how a phenomenological thought process for the development of the difference in reference state chemical potentials between a surfactant in the micelle and the monomer states could be used to avoid some of the assumptions made in the earlier work. Their theory includes a term related to the loss of rotational degrees of freedom, which was later updated by Puvvada and Blankschtein[81] to include, among other improvements, the use of a single-chain mean-field model[82] for considering packing of the surfactant tail in a more rigorous manner (similar to the theory described in Chapter 4). The MT model developed by Puvvada and Blankschtein has since been applied to a large

variety of surfactant systems with great success, including mixed surfactant systems containing nonionics,[83,84] ionics,[85] anionic and cationic mixtures (leading to vesicle formation),[86] and pH-sensitive surfactants.[87] Addition of consideration of counterion binding to the model has yielded further insight into properties such as the degree of counterion binding in both single-component[88,89] and mixed systems.[90] However, in all this work, the focus has been on linear surfactants with varying head type.

In this Thesis, I will provide a systematic study of branched surfactants using an updated MT approach which I have developed to model these surfactants in a manner that is fully compatible with the modeling of linear surfactants. To demonstrate this, I also apply my MT approach to the study of a wide variety of linear surfactants, examining a much larger set of structures than previously considered. Although I focus on single-component surfactant systems to validate my approach, the multicomponent equations of the MT theory can be readily applied to the study of mixtures of surfactants, which could even include mixtures of linear and branched surfactants (since the theory is compatible with both). The study of multicomponent branched surfactant solutions represents an important future research direction.

1.1.4 Computer Simulations of Surfactant Micelles

Computer simulations offer an exciting tool to investigate the behavior of surfactants and other amphiphiles in bulk aqueous solution. Indeed, they provide a molecular-level insight into the intermolecular interactions giving rise to self-assembly phenomena, including providing a detailed description of the equilibrium and dynamic properties of the aggregates formed. The comprehensive reviews of Shelley and Shelley[91] and Brodskaya[92] are recommended to obtain a full picture of the efforts aimed at understanding surfactant micelles using various simulation techniques to date. Here, I will discuss only a few representative studies. As will be shown in this Thesis, computer simulations can also provide a practical means to identify whether a particular molecule exhibits amphiphilic behavior, when the chemical structure is unconventional and the hydrophobic and hydrophilic regions are therefore difficult to identify. For example, a recent molecular dynamics study of a sulforhodamine dye clarified its

ability to form mixed micelles with sodium dodecyl sulfate – an unexpected observation, due to the incorrect, long-held view that this dye was completely hydrophilic.[93]

As an example that has inspired many of my own advances, including the initialization of micelles for simulation discussed in Appendix C, Bogusz and coworkers used all-atomistic molecular dynamics simulations to study micelles comprising α and β anomeric forms of octyl glucoside (OG), using CHARMM[94] and a TIP3P[95] water model. Specifically, these researchers studied OG micelles of varying shape and size, and identified that aggregation numbers of at least 10 are required to maintain a stable micelle.[96] In addition to quantifying shape through ellipsoid fitting to their spherical micelles, Bogusz and coworkers examined tail length, dihedral angle distributions, and the rates at which dihedrals transitioned between energy minima. They found, through a comparison with octane, that these isomerization rates are reduced, indicating reduction in chain mobility for a surfactant tail in a micelle relative to an oil in a droplet. In terms of head groups, Bogusz and coworkers examined clustering of the glucoside head groups, finding that non-random clustering in fact does occur. Their comparison of the anomeric forms illustrated that the two forms have similar micelle size and surfactant tail properties, but differing solvent accessible surface areas and head orientation. In a follow-up study, these researchers studied the dynamic properties of OG micelles, investigating time-dependent fluctuations in shape and the diffusion of OG surfactants within the micelle. They also were able to demonstrate aggregation from dispersed OG monomers and merger of two separate micelles during 4-6 ns simulations.[97] Chong and coworkers later performed a study comparing the β anomeric form of octyl glucoside with the β anomeric form of octyl galactoside (using AMBER[98] and GLYCAM_2000[99] force fields with a TIP3P[95] water model), with a focus on the hydrogen-bonding network in the vicinity of these sugar-based heads, which differ only at a single stereochemical center.[100] These researchers found similar degrees of hydrogen bonding between headgroups, but found an increase in solvation of the galactoside head group, possibly due to an improved accommodation to the structure of the hydrogen-bonding network of water. Recently, Abel and coworkers have investigated the α and β anomeric forms of dodecyl-β-maltoside (based on an

updated force-field parameter set).[101] These researchers found that the β anomeric form has a tendency to exhibit more ellipsoidal micelle behavior, in agreement with experiments, along with an increased radius of gyration. These researchers also measured the translational diffusion of water entrapped in the rough surface of the micelle, finding about a ten-fold decrease.

Bruce and coworkers used all-atomistic molecular dynamics simulations to probe the distribution of counterions about sodium dodecyl sulfate (SDS) micelles (using the parm98 force field in AMBER 6 and a TIP3P[95] water model, with sulfate head parameters due to Schweighofer and coworkers[102]), finding that the diffusion coefficient of sodium ions decreases markedly from the bulk environment through to the first hydration shell.[103] These researchers additionally investigated the structural properties of these micelles in terms of shape (radius of gyration, eccentricity, solvent accessible surface area) and the distribution of surfactant atomic groups within the micelles (order parameters, distributions of internal dihedral angles).[103] An additional study by these researchers[104] investigated the behavior of water about SDS micelles, including studying orientational effects, distortion of the hydrogen-bonding network, and water penetration. For a simulation of a micelle containing 60 SDS molecules, they confirmed the existence of a small water-free core, surrounded by a transition region of gradually-increasing water content.

Although all-atomistic, explicit-solvent simulations provide a detailed understanding of the static and dynamic behavior of surfactant micelles, it is difficult to study the full aggregation process itself using these techniques (note that the above examples generally made use of pre-formed micelles). Two alternatives are coarse-graining and the use of implicit solvent. As an example of coarse-graining, Nelson and coworkers[105] used lattice Monte Carlo simulations to investigate idealized surfactants represented by two head beads and two tail beads, in a solvent represented by single water beads. These researchers were able to determine the volume fraction of micelles as a function of aggregation number and temperature. The size distribution was then interpreted assuming that the surfactants formed either spheres or cylinders, reproducing a peaked distribution for spheres, and an exponential distribution for cylinders

in agreement with thermodynamic models, indicating that the usual assumption of a well-defined (rather than amorphous) shape is appropriate. As an example of use of implicit solvent, Jusufi and coworkers demonstrated a two-step parametrization technique for creating an implicit solvent model,[106] wherein they first performed full all-atomistic, explicit solvent calculations to determine structural properties, such as head group/counterion interaction. Then, the implicit solvent model was adjusted to replicate these results. Finally, hydrophobic attractions between tail beads were determined using Monte Carlo simulations of the system for parameter fitting to reproduce the cmc of the reference surfactant SDS. These researchers were able to demonstrate a promising agreement with experimental data when they changed head groups, indicating transferability of the approach. The effect of changing counterion was also included in the parametrization. An implicit solvent approach was also used by Lazaridis and coworkers[107] to study dodecylphosphocholine (DPC) micelles. Here, adjustment of head group parameters was performed to replicate average experimental aggregation numbers, upon which an accurate cmc was obtained without specific fitting to that property. A full size distribution of micelles was obtained. However, it was noted that the micelle shapes themselves were more irregular in comparison to those obtained using explicit solvent simulations, indicating that the presence of solvent plays a key role in micelle shape determination.

An approach that enables the most accurate simulation of micelles (using all-atomistic, explicit solvent simulations), but that also allows the determination of cmc's and micelle shape and size distribution is clearly warranted. To this end, Stephenson and coworkers[108–110] introduced an alternative to the need for coarse-graining or use of implicit solvent, through their combination of molecular thermodynamics with hydration results obtained from pre-formed micelles simulated using molecular dynamics. Since this advance, and my generalizations to it, will be discussed in Chapter 3, I will not discuss this topic further here.

1.2 Thesis Overview

In the various Chapters and Appendices of this Thesis, I have provided a detailed description of my theoretical and simulations-based framework for the prediction of the micellization properties of linear and branched surfactants. In Chapter 2, I provide a full derivation of a general molecular-thermodynamic (MT) theory for surfactant micellization. In Chapter 3, I present a combined computer-simulation/molecular-thermodynamic (CSMT) framework, which makes use of molecular dynamics simulations to develop a more detailed quantitative characterization of the hydrophobic driving force for micellization in aqueous media. I also present an iterative dividing surface approach which I formulated to locate the position of the micelle core-water interface of simulated micelles, and to enable the direct determination of surfactant head and tail. In Chapter 4, I derive a single-chain, mean-field approach to model the behavior of surfactant tails within the confinement of a liquid-like micelle core, and to select a suitable chain model for evaluating tail conformations. In Chapter 5, I implement my complete computational framework to make predictions of the micellization properties of a large set of charged (anionic, cationic, zwitterionic) and nonionic (ethoxylated and sugar-based) surfactants, investigating the role of the surfactant head in determining the critical micelle concentration (cmc). In Chapter 6, I select three types of branched surfactants: alkyl sulfonates, alkyl benzene sulfonates, and 1-methyl, 4-alkyl pyridiniums, and investigate the effect that branch location and size has on cmc's and optimal micelle shape, size and degree of counterion binding. In Chapter 7, I summarize the key results of my Thesis, describe directions for future research, and make some concluding remarks.

The various Appendices in my Thesis also contain valuable information regarding the specific details of my computational framework, in addition to providing some additional information regarding the software that I have developed in support of my research. In Appendix A, I discuss vector rotation and atom placement algorithms which I have used in several of my computer programs. In Appendix B, I describe a Monte Carlo approach to chain packing, which enables the results of Chapter 4

to be applied to longer or more highly-branched surfactant tails. In Appendix C, I describe my own implementation of a molecular dynamics engine for use in developing numerically-stable initial configurations of surfactant micelles for simulation. In Appendix D, I document the force expressions used in my software. In Appendix E, I provide valuable information regarding force-field development for novel surfactants. Finally, in Appendix F, I provide a wealth of data in support of the determination of surfactant molecular properties for MT modeling.

Chapter 2

A Molecular-Thermodynamic Approach to Micellization

2.1 Overview

A molecular-thermodynamic (MT) approach to predict surfactant micellization behavior involves combining the equations of bulk-phase thermodynamics with a molecular description of the self-assembly process.[81,111,112] In particular, using classical equilibrium thermodynamics, one can derive an expression for the Gibbs free energy of a solution containing freely-dispersed and aggregated molecules in coexistence, under conditions of constant temperature, pressure, and overall species mole numbers.[35,83] Aggregate mole fractions can then be shown to be a function of the difference between reference-state chemical potentials of the aggregates and their constituents.[35] However, actual calculation of this difference requires a molecular description of the aggregation process.[34] It is this molecular description which enables fully predictive calculations.[81]

In Section 2.2 below, I describe the general thermodynamic modeling framework invoked in the MT approach. In Section 2.3, I provide a full description of the molecular-level thought process which enables actual derivation of the micellization properties of interest to a formulator (i.e., critical micelle concentrations, aggregation numbers, and monomer and micelle compositions, in the case of surfactant mix-

tures[88]). In Section 2.4, I provide some of the practical implementation details of the theory. Finally, in Section 2.5, I provide some concluding remarks which summarize the contents of this Chapter.

2.2 Thermodynamic Modeling of the Micellization Process

2.2.1 Introduction

Consider a solution containing a single phase of water, surfactants, and possibly additional components (such as fully-dissociated salts or organic solubilizates) at equilibrium. The Gibbs free energy of this solution can always be written in functional form as[113]

$$\underline{G} = \underline{G}\left(T, P, \{N_i\}\right), \tag{2.2.1}$$

where $\underline{G}$ is the extensive Gibbs free energy, T is the system temperature, P is the system pressure, and $\{N_i\}$ is the set of mole numbers for all the species present, where the index i is reserved for indexing chemical species throughout the discussion which follows. (Note that I also follow a convention of using an underbar to denote quantities which depend on the extent of the system.[113])

Based on the variables specified in Eq. 2.2.1, one can also write the total differential of $\underline{G}$ as[113]

$$d\underline{G} = \left(\frac{\partial \underline{G}}{\partial T}\right)\Big|_{P,N_i} dT + \left(\frac{\partial \underline{G}}{\partial P}\right)\Big|_{T,N_i} dP + \sum_i \left(\frac{\partial \underline{G}}{\partial N_i}\right)\Big|_{T,P,N_{j\neq i}} dN_i. \tag{2.2.2}$$

Applying the well-known relationships $\left(\frac{\partial \underline{G}}{\partial T}\right)\Big|_{P,N_i} = -\underline{S}$, where $\underline{S}$ is the total system entropy; $\left(\frac{\partial \underline{G}}{\partial P}\right)\Big|_{T,N_i} = \underline{V}$, where $\underline{V}$ is the system volume; and $\left(\frac{\partial \underline{G}}{\partial N_i}\right)\Big|_{T,P,N_{j\neq i}} = \mu_i$, where μ_i is the chemical potential of species i (partial molar Gibbs free energy), Eq. 2.2.2 becomes[113]

$$d\underline{G} = -\underline{S}dT + \underline{V}dP + \sum_i \mu_i dN_i. \tag{2.2.3}$$

As partial molar quantities of $\underline{G}$, the chemical potentials $\{\mu_i\}$ partition $\underline{G}$ into portions ascribed to each molecule, such that $\underline{G}$ may be recovered directly from knowledge of the set $\{\mu_i\}$ via the summability relation[114]

$$\underline{G} = \sum_i N_i \mu_i. \tag{2.2.4}$$

In practice, the chemical potentials $\{\mu_i\}$ are typically related to reference-state chemical potentials by[113]

$$\mu_i = \mu_i^0 + k_B T \ln a_i, \tag{2.2.5}$$

where μ_i^0 is the chemical potential for species i in some specified reference state, k_B is the Boltzmann constant, and a_i is the activity of species i, which in turn may be expressed as

$$a_i = \gamma_i x_i, \tag{2.2.6}$$

where γ_i is the activity coefficient of species i, and x_i is the mole fraction of species i in solution, defined as $x_i \equiv N_i / \sum_i N_i$. Deviation of the activity coefficient from unity is a measure of the non-ideality present in the system, and requires a system-specific model for real systems.[114]

For systems in which aggregation occurs, development of models for γ_i is a highly non-trivial undertaking. In the approach that I will discuss below, I build on the modeling work carried out by members of the Blankschtein Group, whose molecular-thermodynamic (MT) models have been applied to a large variety of single and mixed surfactant systems with great success.[81,83,85,88–90,112,115–117]

One begins with a direct model for the functionality of the Gibbs free energy (as expressed in Eq. 2.2.1) for systems in which one or several of the species in solution (commonly referred to as "monomers"[66] or "unimers"[19] in dissociated form) may assemble into transient aggregates. By "aggregate", I refer to conceptual groupings of more than one molecule, where molecules within the grouping are localized in space such that their individual translational motions are correlated for time scales exceeding those for molecular diffusion within the bulk solvent. That is, there is

a criterion for some temporary "cohesiveness" between the molecules to be satisfied in order to consider them as constituting aggregates, although any one molecule within such a grouping may be found to re-enter the bulk solution rapidly (as Israelachvili discusses,[35] this time scale can be as short as tens of microseconds for small aggregates). When speaking of thermodynamic equilibrium for such transient aggregates, one means equilibration of the overall population distribution of the aggregates,[34,83,105,118] *not* that the aggregates themselves persist indefinitely.

For aggregating systems, it is useful to draw a distinction between the chemical *components* of a solution and its *species*, where the term *component* is used to refer to solvent, ions, free monomers, or aggregates,[111] while the term *species* refers to the unique types of molecules present in the system, but is not concerned with their aggregation state. Accordingly, a given surfactant is represented by a single species index, but may be found in one of many possible components. To be more concrete, some examples of components in a surfactant solution include: (a) a single water molecule, (b) a single surfactant monomer, and (c) a surfactant micelle of a specified aggregation number, shape, and size. For the purposes of the modeling below, water is always assumed to be mobile, such that, while micelles are certainly hydrated, the water itself surrounding the micelle is not considered part of the aggregate – that is, components containing both surfactants and water will not be modeled.

I will need to make use of a variety of indices in the following analysis. As already mentioned, species will be indexed i, with surfactant species indexed by s, ions indexed by k, and water as a species labeled as w. In addition, components will be indexed by c, with surfactant aggregate components (surfactant micelles) indexed by ξ, and single, free molecule components indexed by m (i.e., surfactant monomers, individual ions, or water – other free molecules, including organic solutes, could also be present in solubilization applications).

Each component (indexed by c) will be specified by a number of parameters (or "descriptors") needed to fully specify its free energy. These descriptors will typically include: (i) a set of aggregation numbers for each species, $\{n_{c,i}\}$ (where a particular $n_{c,i}$ may be zero), from which a total aggregation number $n_c = \sum_i n_{c,i}$ may be

calculated (note that a valid solution component must have $n_c \geq 1$), (ii) the aggregate geometry (shape and size), in the case of an aggregate component, and (iii) the distribution of species within the aggregate, if there is any partitioning or compartmentalization. The total number of a particular component found in solution will be denoted N_c, and the number of a particular species in solution can be recovered by summing across components: $N_i = \sum_c N_c n_{c,i}$, a single-species material balance.

For point (iii) above, an example of partitioning can be found in solubilization applications, where organic solutes may preferentially localize to the core, interface, or palisade (head) regions depending on their chemical structure (i.e., demonstrating a locus of solubilization).[12,119,120] An example of compartmentalization involves polymeric micelles with hydrophobic domains composed of fluorocarbon and hydrocarbon units, which exhibit enthalpic antagonism and demixing in the core, leading to compartments which can preferentially solubilize various solutes.[121,122] In the systems that I study in this Thesis, which involve single-component surfactant solutions, with purely hydrocarbon-based surfactants, consideration of (iii) is unnecessary. I mention this aspect of micellization for completeness.

I would like to reiterate at this point that the population distribution of aggregates (with respect to the three types of descriptors discussed above) is what is stable at thermodynamic equilibrium. It is the purpose of an MT theory to determine this population distribution.

2.2.2 Component Chemical Potential Derivation

System Gibbs Free Energy

The system Gibbs free energy, expressed in Eq. 2.2.1, is first decomposed into the following three free-energy contributions:[81]

$$\underline{G} = \underline{G}_{\text{formation}} + \underline{G}_{\text{mixing}} + \underline{G}_{\text{interaction}}, \tag{2.2.7}$$

where $\underline{G}_{\text{formation}}$ is the formation free energy, $\underline{G}_{\text{mixing}}$ is a mixing free energy, and $\underline{G}_{\text{interaction}}$ is an interaction free energy. The term $\underline{G}_{\text{interaction}}$ captures any component-

component, non-ideal interactions not captured in the explicit formulation of $\underline{G}_{\text{mixing}}$, which is usually, in some sense, an ideal mixing term.[123]

Formation Free-Energy Contribution

In terms of components, the formation free energy is a function of the component reference-state chemical potentials:

$$\underline{G}_{\text{formation}} = \sum_c N_c \mu_c^0, \tag{2.2.8}$$

where μ_c^0 is the reference-state chemical potential of component c. Relating μ_c^0 to the individual $\mu_{c,i}^0$ values of a component's constituents will require a separate model in general, which will be presented in the subsection entitled "Relating the Component and Species Chemical Potentials" below. The typical reference state used for solutes is infinite dilution, while the reference state for the solvent (water) is that of a pure liquid phase at the same temperature and pressure conditions as the solution under consideration.[111]

Mixing Free-Energy Contribution

In the language of statistical mechanics, the mixing free energy may be written generally as follows:

$$\underline{G}_{\text{mixing}} = -k_B T \ln Z + P\underline{V}, \tag{2.2.9}$$

where Z is the overall system partition function, and P and $\underline{V}$ are the system pressure and volume, respectively. When dealing with aggregating systems, there are a variety of approaches that have been used to estimate Z.[123] Here, I make use of a translational entropy model, which views the solution as a regular lattice, consisting of N_{sites} total lattice sites of equal volume v_0, such that $\underline{V} = N_{sites} v_0$ and

$$N_{sites} \equiv \sum_c N_c n_c = \sum_i N_i. \tag{2.2.10}$$

Note that the lattice site volume, v_0, can be computed as $\underline{V}/N_{sites}$ once an equation of state has been specified which allows determination of $\underline{V} = \underline{V}(T, P, \{N_i\})$. In fact, actual determination of v_0 is usually not required during the derivation.

The translational entropy model neglects size differences between molecules but takes into account aggregation number differences between components, in contrast to an ideal solution model. The partition function of the translational entropy model is formulated as follows:[123]

$$Z = \frac{\prod_c (N_{sites})^{N_c}}{\prod_c N_c!}, \tag{2.2.11}$$

which can be interpreted as modeling the mixing of components as the mixing of point particles allowed to occupy any one of the N_{sites} lattice sites, with no intercomponent interactions. Note that individual instances of a particular type of component in solution are considered to be indistinguishable. Taking the natural logarithm of Eq. 2.2.11, we find that

$$\begin{aligned} \ln Z &= \sum_c (N_c \ln N_{sites} - \ln N_c!) \\ &= \sum_c (N_c \ln N_{sites} - N_c \ln N_c + N_c) \\ &= \sum_c N_c \left[1 - \ln\left(\frac{N_c}{\sum_{c'} N_{c'} n_{c'}}\right)\right], \end{aligned} \tag{2.2.12}$$

or

$$\ln Z = \sum_c N_c \left[1 - \ln X_c\right], \tag{2.2.13}$$

where we have made use of Stirling's approximation ($\ln N_c! \approx N_c \ln N_c - N_c$, for large N_c), the definition for N_{sites} found in Eq. 2.2.10, and the introduction of a new term, referred to hereafter as the modified mole fraction of component c, defined as follows:

$$X_c \equiv N_c / \sum_{c'} N_{c'} n_{c'}. \tag{2.2.14}$$

Note that the modified mole fraction X_c will only be equivalent to the usual mole fraction if all the components in solution are unimolecular (i.e., if $n_{c'} = 1$ for all c' in Eq. 2.2.14). Generally, $\sum_c X_c \neq 1$.

Inserting Eq. 2.2.13 into Eq. 2.2.9, we find that

$$\underline{G}_{\text{mixing}} = -k_B T \sum_c N_c \left[1 - \ln X_c\right] + P\underline{V}. \tag{2.2.15}$$

We now need an equation of state to evaluate $P\underline{V}$ in Eq. 2.2.15. Similar to the case with solutions containing only unimolecular components,[113] one can generally define a compressibility factor for a solution containing aggregate components as[123]

$$\widetilde{Z} \equiv \frac{P\underline{V}}{k_B T \sum_c N_c}, \tag{2.2.16}$$

which, upon rearrangement, yields

$$P\underline{V} = \widetilde{Z} k_B T \sum_c N_c, \tag{2.2.17}$$

where I have used the symbol $\widetilde{Z}$ for the compressibility factor to avoid confusing it with the partition function, Z. Selecting an equation of state is equivalent to determining the functional form of $\widetilde{Z}$ as a function of the natural variables of the system. Inserting Eq. 2.2.17 into Eq. 2.2.15, we obtain

$$\underline{G}_{\text{mixing}} = -k_B T \sum_c N_c \left[\left(1 - \widetilde{Z}\right) - \ln X_c\right]. \tag{2.2.18}$$

As one possible model for $\widetilde{Z}$, one could propose $\widetilde{Z} = 1$, making an ideal gas-like assumption. This yields

$$\underline{G}_{\text{mixing}} \left(\widetilde{Z} = 1\right) = k_B T \sum_c N_c \ln X_c. \tag{2.2.19}$$

Note that Eq. 2.2.19 is equivalent to an ideal solution result in systems where all the components are unimolecular (i.e., where $n_c = 1$ for all components c, such that X_c in Eq. 2.2.14 becomes the usual mole fraction).

Nagarajan suggests another possible model for $\widetilde{Z}$, recognizing that, for a liquid state, the compressibility factor defined in Eq. 2.2.16 may be closer to 0 (i.e., in an

incompressible state where components are tightly packed).[123] Assuming $\widetilde{Z} \approx 0$ (i.e., $\widetilde{Z} \ll 1$), Eq. 2.2.18 becomes

$$\underline{G}_{\text{mixing}}\left(\widetilde{Z} \approx 0\right) = -k_B T \sum_c N_c \left[1 - \ln X_c\right] \tag{2.2.20}$$

In this latter case, $\underline{G}_{\text{mixing}} = \underline{F}_{\text{mixing}}$, where $\underline{F}_{\text{mixing}}$ is the Helmholtz free energy of mixing of the components,

$$\underline{F}_{\text{mixing}} = -k_B T \ln Z. \tag{2.2.21}$$

The drawback of Eq. 2.2.20 is that it does not reduce to the ideal solution result when all the components are unimolecular. Moreover, Blankschtein, Thurston, and Benedek demonstrated that Eq. 2.2.19 is an effective model when used to make practical predictions in systems exhibiting micellization and phase separation behavior.[124] As a result, in the analysis which follows, I will make preferential use of Eq. 2.2.19, since $\underline{G}_{\text{mixing}}$ is envisioned to represent the ideal mixing of components, and this expression should hold in the limit where no aggregate components are present (e.g., below the critical micelle concentration in a surfactant solution[35]).

Interaction Free-Energy Contribution

If Eq. 2.2.19 is analogous to ideal mixing in a fully-dispersed system (i.e., containing no aggregates), the interaction free-energy contribution is similarly analogous to a non-ideal mixing correction term (an excess function). Specifically, the interaction free-energy contribution is of importance in assessing phase separation in surfactant systems (e.g., for predicting cloud points in ethoxylated surfactant systems). Since my thesis research focuses on single-phase, dilute systems, I will proceed with setting $\underline{G}_{\text{interaction}} = 0$ (an assumption appropriate for dilute micellar solutions). However, for completeness, an example of a functional form for $\underline{G}_{\text{interaction}}$, used by Puvvada and Blankschtein in studying ethoxylated surfactants, is as follows:[81]

$$\underline{G}_{\text{interaction}} = -\frac{1}{2} C\left(T, P\right) N_{surf} \phi, \tag{2.2.22}$$

where $C(T,P)$ captures the strength of interactions between micelles, N_{surf} is the total number of surfactant molecules in solution ($N_{surf} = \sum_s N_s$), and ϕ is the total volume fraction of these surfactants (assuming some characteristic molar volume).

Favorable intermicellar interactions in ethoxylated systems, leading to phase separation, are believed to be due to hydration shells that form around the ethoxylate units in the surfactant heads,[125] which extend into water similar to an end-tethered polymer brush.[126] When ethoxylate chains overlap as two micelles approach one another, water is released, leading to a driving force for association similar to the hydrophobic effect which drives the micellization process.[125] This driving force for intermicellar association is partially offset by other terms, such as a difference in chemical potential of solvent in the overlapping, brush-like region relative to the bulk state – an "osmotic" effect which imparts some degree of steric stabilization.[127]

Note that an alternative approach to studying intermicellar and intramicellar interactions is available which does not involve the decomposition of $\underline{G}$ into $\underline{G}_{\text{mixing}}$ and $\underline{G}_{\text{interaction}}$ terms. This approach is one developed by Zoeller and coworkers using the McMillan-Mayer theory,[128,129] but this theory is beyond the scope of this Thesis.

Component Chemical Potential Calculation

Having defined the various contributions to $\underline{G}$, one can now insert Eqs. 2.2.8 and 2.2.19 into Eq. 2.2.7. Assuming that the micellar solution is dilute, such that $\underline{G}_{\text{interaction}} = 0$, the following expression for $\underline{G}$ is obtained:

$$\underline{G} = \sum_c N_c \left(\mu_c^0 + k_B T \ln X_c\right). \tag{2.2.23}$$

We now want to compute the chemical potential of any one component in the solution. This is accomplished in a manner similar to the calculation of a species chemical potential. Specifically,[83]

$$\mu_c = \left(\frac{\partial \underline{G}}{\partial N_c}\right)\Bigg|_{T,P,N_{c'\neq c}}, \tag{2.2.24}$$

which leads to

$$\begin{aligned}
\mu_c &= \frac{\partial}{\partial N_c}\left(\sum_{c'} N_{c'}\left(\mu_{c'}^0 + k_B T \ln X_{c'}\right)\right) \\
&= \sum_{c'}\left[\frac{\partial N_{c'}}{\partial N_c}\left(\mu_{c'}^0 + k_B T \ln X_{c'}\right) + N_{c'}\left(k_B T \frac{\partial}{\partial N_c}\ln X_{c'}\right)\right] \\
&= \sum_{c'}\left[\delta_{cc'}\left(\mu_{c'}^0 + k_B T \ln X_{c'}\right) + N_{c'}\left(\frac{k_B T}{X_{c'}}\frac{\partial X_{c'}}{\partial N_c}\right)\right], \qquad (2.2.25)
\end{aligned}$$

where $\delta_{cc'}$ is a Kronecker delta (equal to 1 if $c = c'$, and zero otherwise). Proceeding, we find

$$\begin{aligned}
\frac{\partial X_{c'}}{\partial N_c} &= \frac{\partial}{\partial N_c}\left(\frac{N_{c'}}{\sum_{c''} N_{c''} n_{c''}}\right) \qquad (2.2.26) \\
&= \frac{\delta_{cc'} - X_{c'} n_c}{\sum_{c''} N_{c''} n_{c''}}.
\end{aligned}$$

Inserting Eq. 2.2.26 into Eq. 2.2.25, one obtains

$$\begin{aligned}
\mu_c &= \sum_{c'}\left[\delta_{cc'}\left(\mu_{c'}^0 + k_B T \ln X_{c'}\right) + N_{c'}\left(\frac{k_B T}{X_{c'}}\frac{\delta_{cc'} - X_{c'} n_c}{\sum_{c''} N_{c''} n_{c''}}\right)\right] \\
&= \sum_{c'}\left[\delta_{cc'}\left(\mu_{c'}^0 + k_B T \ln X_{c'}\right) + k_B T\left(\delta_{cc'} - X_{c'} n_c\right)\right]. \qquad (2.2.27)
\end{aligned}$$

After evaluation of the summations and Kronecker deltas, Equation 2.2.27 simplifies to

$$\mu_c = \mu_c^0 + k_B T\left(\ln X_c + 1 - n_c \sum_{c'} X_{c'}\right). \qquad (2.2.28)$$

To cast the chemical potential of a component in terms of its activity, one can write

$$\mu_c = \mu_c^0 + k_B T \ln a_c, \qquad (2.2.29)$$

where

$$a_c = X_c \exp\left(1 - n_c \sum_{c'} X_{c'}\right). \qquad (2.2.30)$$

We can also define an activity coefficient, γ_c, as $\gamma_c \equiv a_c/X_c$, which yields

$$\gamma_c = \exp\left(1 - n_c \sum_{c'} X_{c'}\right). \tag{2.2.31}$$

If all the components are unimolecular, $\gamma_c = 1$ (since $\sum_{c'} X_{c'} = 1$ and $n_c = 1$), reflecting the fact that the expression for $\underline{G}_{\text{mixing}}$ in Eq. 2.2.19 becomes identical to the expression for an ideal solution in this limit.

As a check of the chemical potential result in Eq. 2.2.28, one can demonstrate that the Gibbs free energy summability relation holds. Specifically,

$$\begin{aligned}
\sum_c N_c \mu_c &= \sum_c N_c \left[\mu_c^0 + k_B T \left(\ln X_c + 1 - n_c \sum_{c'} X_{c'}\right)\right] \\
&= \underline{G}_f + \underline{G}_m + \sum_c N_c k_B T \left(1 - n_c \sum_{c'} X_{c'}\right) \\
&= \underline{G}_f + \underline{G}_m + k_B T \left(\sum_c N_c - \left(\sum_c N_c n_c\right) \sum_{c'} \frac{N_{c'}}{\sum_{c''} N_{c''} n_{c''}}\right) \\
&= \underline{G}_f + \underline{G}_m.
\end{aligned} \tag{2.2.32}$$

That is,

$$\underline{G} = \sum_c N_c \mu_c, \tag{2.2.33}$$

since $\underline{G}_{\text{interaction}} = 0$, an assumption made early in the derivation of Eq. 2.2.28.

Relating the Component and Species Chemical Potentials

Since the individual species within multimolecular components (i.e., aggregates) rapidly exchange with the monomer population, and all the components are fully dispersed at equilibrium, the entire solution is considered to be a single phase. We have already shown that Eq. 2.2.33 holds. However, in a single phase, Eq. 2.2.4 must also hold, yielding

$$\sum_c N_c \mu_c = \sum_i N_i \mu_i. \tag{2.2.34}$$

Recall that the species material balance for each species i is given by

$$N_i = \sum_c N_c n_{c,i}. \tag{2.2.35}$$

Inserting Eq. 2.2.35 into Eq. 2.2.34 yields

$$\begin{aligned} \sum_c N_c \mu_c &= \sum_i \left(\sum_c N_c n_{c,i} \right) \mu_i \\ &= \sum_c N_c \left(\sum_i n_{c,i} \mu_i \right). \end{aligned} \tag{2.2.36}$$

Rearranging Eq. 2.2.36 yields

$$\sum_c N_c \left(\mu_c - \sum_i n_{c,i} \mu_i \right) = 0. \tag{2.2.37}$$

Equation 2.2.37 would clearly be satisfied if

$$\mu_c = \sum_i n_{c,i} \mu_i, \tag{2.2.38}$$

but the various N_c's are not independent quantities, such that this result may not be the only possible solution. To arrive at Eq. 2.2.38 conclusively, then, one must take an alternative approach. This involves another relationship that must hold at equilibrium for the single phase:

$$d\underline{G}|_{T,P,\{N_i\}} = 0. \tag{2.2.39}$$

Inserting Eq. 2.2.33 into Eq. 2.2.39 yields

$$\sum_c (\mu_c dN_c + N_c d\mu_c) = 0. \tag{2.2.40}$$

A constraint equation arises from holding the species mole numbers constant in Eq. 2.2.39, which is equivalent to requiring $dN_i = 0$ for all species i. Taking the

differential of Eq. 2.2.35, and applying this fact yields

$$\sum_c n_{c,i} dN_c = 0. \tag{2.2.41}$$

One can always apply the Gibbs-Duhem equation $\sum_i N_i d\mu_i = 0$ at equilibrium at constant T and P. If the components in solution are treated as distinct chemical species, a similar Gibbs-Duhem equation $\sum_c N_c d\mu_c = 0$ would also hold. Making this assumption, Eq. 2.2.40 becomes

$$\sum_c \mu_c dN_c = 0, \tag{2.2.42}$$

which can be divided into monomeric and aggregate contributions as follows:

$$\sum_m \mu_m dN_m + \sum_\xi \mu_\xi dN_\xi = 0. \tag{2.2.43}$$

The constraint equation, Eq. 2.2.41, can likewise be decomposed into monomer and aggregate contributions:

$$\sum_m n_{m,i} dN_m + \sum_\xi n_{\xi,i} dN_\xi = 0. \tag{2.2.44}$$

Note that there is always a one-to-one mapping between the identity of monomers and species in solution (i.e., there is one unique monomer component in solution for each species, and each monomer component corresponds to one species). If we denote the monomer m corresponding to species i as $m(i)$, then Eq. 2.2.44 simplifies to

$$dN_{m(i)} = -\sum_\xi n_{\xi,i} dN_\xi. \tag{2.2.45}$$

This use of the notation $m(i)$ also allows one to convert the monomer summation

in Eq. 2.2.43 into a species summation as follows:

$$\sum_i \mu_{m(i)} dN_{m(i)} + \sum_\xi \mu_\xi dN_\xi = 0. \tag{2.2.46}$$

We can now substitute Eq. 2.2.45 into Eq. 2.2.46, which yields

$$\sum_i \mu_{m(i)} \left(- \sum_\xi n_{\xi,i} dN_\xi \right) + \sum_\xi \mu_\xi dN_\xi = 0, \tag{2.2.47}$$

which can be rearranged to obtain

$$\sum_\xi dN_\xi \left(\mu_\xi - \sum_i n_{\xi,i} \mu_{m(i)} \right) = 0. \tag{2.2.48}$$

Since the dN_ξ's may be considered to vary independently (the constraint of Eq. 2.2.45 has already been applied to eliminate the dependent mole numbers), this relationship will hold in general only if

$$\mu_\xi = \sum_i n_{\xi,i} \mu_{m(i)}. \tag{2.2.49}$$

This is referred to as the principle of multiple chemical equilibrium[81,130,131] (derived assuming a Gibbs-Duhem equation for the components), which states that an aggregate chemical potential is equal to the sum of the monomer chemical potentials of its constituents.

If one returns to the assumed Gibbs-Duhem equation for components, $\sum_c N_c d\mu_c = 0$, and expands this expression in terms of monomers and aggregates,

$$\sum_m N_m d\mu_m + \sum_\xi N_\xi d\mu_\xi = 0, \tag{2.2.50}$$

then makes use of the $m(i)$ notation, and substitutes in the differential of Eq. 2.2.49, one obtains

$$\sum_i N_{m(i)} d\mu_{m(i)} + \sum_\xi N_\xi \sum_i n_{\xi,i} d\mu_{m(i)} = 0, \tag{2.2.51}$$

which simplifies to

$$\sum_i \left(N_{m(i)} + \sum_\xi N_\xi n_{\xi,i} \right) d\mu_{m(i)} = 0. \tag{2.2.52}$$

The term in parentheses is simply N_i (i.e., $\sum_c N_c n_{c,i}$), the mole number of species i. Furthermore, thermodynamic equilibrium requires that a particular molecule have the same chemical potential at any location within the solution[35] (diffusional equilibrium). This means that we can use the monomer chemical potential as representative of the corresponding species chemical potential, such that $\mu_{m(i)} = \mu_i$ and $d\mu_{m(i)} = d\mu_i$. Therefore, Eq. 2.2.52 is equivalent to $\sum_i N_i d\mu_i = 0$. That is, one can see that the principle of multiple chemical equilibrium is consistent with a species-level treatment of the solution thermodynamics. This consistency is also evident in applying $\mu_{m(i)} = \mu_i$ to Eq. 2.2.49, which yields the relationship expressed in Eq. 2.2.38, thereby satisfying Eq. 2.2.37, which was presented at the beginning of this subsection.

Gibbs-Duhem Equation

As an aside, I would like to briefly return to the Gibbs-Duhem equation to illustrate the nature of the activity coefficients in Eq. 2.2.31. At a species level, at constant temperature and pressure, the Gibbs-Duhem equation is[113]

$$\sum_i N_i d\mu_i = 0. \tag{2.2.53}$$

If the species chemical potentials are defined as in Eq. 2.2.5, one can show that

$$\sum_i N_i d\ln\gamma_i = 0. \tag{2.2.54}$$

If we start instead with the component Gibbs-Duhem equation, expressed in the previous section as

$$\sum_c N_c d\mu_c = 0, \tag{2.2.55}$$

we can now make use of Eq. 2.2.29 and our subsequent definition of the component

activity coefficient. This yields

$$\begin{aligned}\sum_c N_c d\mu_c &= \sum_c N_c d\left(\mu_c^0 + k_B T \ln a_c\right) \\ &= k_B T \sum_c N_c d\ln\left(\gamma_c X_c\right) \\ &= k_B T \sum_c N_c \left[d\ln\gamma_c + \frac{1}{X_c} dX_c\right] \\ &= k_B T \sum_c N_c d\ln\gamma_c + k_B T \left(\sum_{c'} N_c' n_{c'}\right) \sum_c dX_c. \end{aligned} \tag{2.2.56}$$

Differentiating Eq. 2.2.14 yields

$$\begin{aligned} dX_c &= X_c \left[\frac{dN_c}{N_c} - \frac{\sum_{c'} n_{c'} dN_{c'}}{\sum_{c'} N_{c'} n_{c'}}\right] \\ &= \frac{1}{\sum_{c'} N_{c'} n_{c'}} \left[dN_c - X_c \sum_{c'} n_{c'} dN_{c'}\right]. \end{aligned} \tag{2.2.57}$$

Substituting Eq. 2.2.57 into Eq. 2.2.56 then leads to

$$\begin{aligned}\sum_c N_c d\mu_c &= k_B T \sum_c N_c d\ln\gamma_c + k_B T \sum_c \left[dN_c - X_c \sum_{c'} n_{c'} dN_{c'}\right] \\ &= k_B T \sum_c N_c d\ln\gamma_c + k_B T \sum_c dN_c \left[1 - n_c \sum_{c'} X_{c'}\right]. \end{aligned} \tag{2.2.58}$$

Utilizing Eq. 2.2.55, this implies that

$$\sum_c N_c d\ln\gamma_c + \sum_c dN_c \left[1 - n_c \sum_{c'} X_{c'}\right] = 0. \tag{2.2.59}$$

For a system containing only dispersed, unimolecular components, it is clear that $n_c = 1$ for all components c and $\sum_{c'} X_{c'} = 1$ (the $X_{c'}$ becoming normal mole fractions in this case), such that Eq. 2.2.54 is recovered.

2.2.3 Aggregate Modified Mole Fraction Derivation

Revisiting Eq. 2.2.49, both μ_ξ and $\mu_{m(i)}$ must obey Eq. 2.2.28. This leads to

$$\begin{aligned}\mu_\xi &= \mu_\xi^0 + k_B T \left(\ln X_\xi + 1 - n_\xi \sum_{c'} X_{c'} \right) \\ &= \sum_i n_{\xi,i} \mu_{m(i)} \\ &= \sum_i n_{\xi,i} \left[\mu_{m(i)}^0 + k_B T \left(\ln X_{m(i)} + 1 - \sum_{c'} X_{c'} \right) \right]. \end{aligned} \tag{2.2.60}$$

The right-hand side expressions in the first and third lines of Eq. 2.2.60 can be directly equated to each other, then further rearranged and simplified, to obtain

$$\ln X_\xi = -\beta \left(\mu_\xi^0 - \sum_i n_{\xi,i} \mu_{m(i)}^0 \right) + \left(\sum_i n_{\xi,i} \ln X_{m(i)} \right) + \left(\sum_i n_{\xi,i} \right) - 1, \tag{2.2.61}$$

where $\beta = 1/k_B T$ is the thermodynamic beta, and $n_\xi = \sum_i n_{\xi,i}$ has been used to cancel the terms involving $\sum_{c'} X_{c'}$.

Taking the exponential of Eq. 2.2.61 yields

$$X_\xi = \frac{1}{e} \exp \left(-\beta \left(\mu_\xi^0 - \sum_i n_{\xi,i} \mu_{m(i)}^0 \right) \right) \prod_i \left(e X_{m(i)} \right)^{n_{\xi,i}}. \tag{2.2.62}$$

At this stage, we can define a micellization free energy, $\underline{G}_{mic,\xi}$, which is a function of the extent of the micelle. Specifically,

$$\underline{G}_{mic,\xi} \equiv \mu_\xi^0 - \sum_i n_{\xi,i} \mu_{m(i)}^0. \tag{2.2.63}$$

Calculation of Eq. 2.2.63 requires a phenomenological, molecular-level thought process, of the type presented in Section 2.3.

Equation 2.2.63 can also be expressed on a per-surfactant molecule basis as follows:

$$g_{mic,\xi} \equiv \underline{G}_{mic,\xi} / n_{surf,\xi}, \tag{2.2.64}$$

where $n_{surf,\xi}$ is the surfactant aggregation number in the micelle (i.e., $n_{surf,\xi} = \sum_s n_{\xi,s}$). Substituting Eq. 2.2.63 into Eq. 2.2.64, and splitting the summation over species i into a summation over surfactants s and ions k (recall that water, w, is not considered to be part of any aggregate component), we arrive at the following useful expression for $g_{mic,\xi}$:

$$g_{mic,\xi} = \frac{\mu^0_\xi}{n_{surf,\xi}} - \sum_s \alpha_{\xi,s}\mu^0_{m(s)} - \sum_k \widehat{\beta}_{\xi,k}\mu^0_{m(k)} \tag{2.2.65}$$

where $\alpha_{\xi,s} = n_{\xi,s}/n_{surf,\xi}$ is the mole fraction of surfactant s in the micelle (on a per-surfactant molecule basis), and $\widehat{\beta}_{\xi,k} = n_{\xi,k}/n_{surf,\xi}$ is the extent (or degree) of binding of ion k to the micelle (non-zero only for ions of charge opposite to that of the surfactants, such that non-zero $\widehat{\beta}_{\xi,k}$'s are referred to as degrees of counterion binding).[90]

Substituting Eq. 2.2.65 in Eq. 2.2.62, one obtains

$$X_\xi = \frac{1}{e}\exp\left(-\beta n_{surf,\xi} g_{mic,\xi}\right)\prod_i \left(eX_{m(i)}\right)^{n_{\xi,i}}. \tag{2.2.66}$$

One can manipulate this expression further by multiplying and dividing by $X_1^{n_{surf,\xi}}$, where $X_1 \equiv \sum_s X_{m(s)}$ is the total modified mole fraction for monomers in solution, and breaking the product term in Eq. 2.2.66 into surfactant and counterion terms. This yields

$$X_\xi = \frac{1}{e}X_1^{n_{surf,\xi}}\exp\left(-\beta n_{surf,\xi} g_{mic,\xi}\right)\prod_s \left(\frac{eX_{m(s)}}{\sum_{s'} X_{m(s')}}\right)^{n_{\xi,s}} \prod_k \left(eX_{m(k)}\right)^{n_{\xi,k}}. \tag{2.2.67}$$

Note that

$$\frac{X_{m(s)}}{\sum_{s'} X_{m(s')}} = \frac{N_{m(s)}}{\sum_{s'} N_{m(s')}} \equiv \alpha_{m(s)}, \tag{2.2.68}$$

where $\alpha_{m(s)}$ is defined as the mole fraction of surfactant species s in monomeric form (i.e., as a unimolecular component), on a total surfactant monomer basis.

Continuing with manipulation of the two product terms in Eq. 2.2.67:

$$\begin{aligned}
&\prod_s \left(\frac{eX_{m(s)}}{\sum_{s'} X_{m(s')}}\right)^{n_{\xi,s}} \\
&= \exp\left(-\beta n_{surf,\xi}\left[-\frac{1}{\beta n_{surf,\xi}}\ln\left(\prod_s \left(e\alpha_{m(s)}\right)^{n_{\xi,s}}\right)\right]\right) \\
&= \exp\left(-\beta n_{surf,\xi}\left[-\frac{1}{\beta n_{surf,\xi}}\sum_s n_{\xi,s}\left(1+\ln\alpha_{m(s)}\right)\right]\right) \\
&= \exp\left(-\beta n_{surf,\xi}\left[-k_BT\left(1+\sum_s \alpha_{\xi,s}\ln\alpha_{m(s)}\right)\right]\right), \qquad (2.2.69)
\end{aligned}$$

and

$$\begin{aligned}
&\prod_k \left(eX_{m(k)}\right)^{n_{\xi,k}} \\
&= \exp\left(-\beta n_{surf,\xi}\left[-\frac{1}{\beta n_{surf,\xi}}\ln\left(\prod_k \left(eX_{m(k)}\right)^{n_{\xi,k}}\right)\right]\right) \\
&= \exp\left(-\beta n_{surf,\xi}\left[-\frac{1}{\beta n_{surf,\xi}}\sum_k n_{\xi,k}\left(1+\ln X_{m(k)}\right)\right]\right) \\
&= \exp\left(-\beta n_{surf,\xi}\left[-k_BT\left(\sum_k \widehat{\beta}_{\xi,k}\left(1+\ln X_{m(k)}\right)\right)\right]\right). \qquad (2.2.70)
\end{aligned}$$

Substituting Eqs. 2.2.69 and 2.2.70 back into Eq. 2.2.67, one finds

$$X_\xi = \frac{1}{e}X_1^{n_{surf,\xi}}\exp\left(-\beta n_{surf,\xi}\left[\begin{array}{c} g_{mic,\xi} - k_BT\left(1+\sum_s \alpha_{\xi,s}\ln\alpha_{m(s)}\right) + \\ -k_BT\left(\sum_k \widehat{\beta}_{\xi,k}\left(1+\ln X_{m(k)}\right)\right)\end{array}\right]\right). \qquad (2.2.71)$$

One can now define a modified free energy of micellization as follows:[90]

$$g_{m,\xi} \equiv g_{mic,\xi} - k_BT\left(1+\sum_s \alpha_{\xi,s}\ln\alpha_{m(s)}\right) - k_BT\left(\sum_k \widehat{\beta}_{\xi,k}\left(1+\ln X_{m(k)}\right)\right). \qquad (2.2.72)$$

Note that, in Ref. [90], the authors have chosen to insert the term $-k_BT\left(1+\sum_k \widehat{\beta}_{\xi,k}\right)$ contained in Eq. 2.2.72 into their expression for $g_{mic,\xi}$ rather than present it as a sep-

arate term in $g_{m,\xi}$. However, there is no compelling reason to do this (and it is not done in Ref. [84], for example), since the magnitude of $g_{mic,\xi}$ is most clearly understood when it represents a difference in reference state chemical potentials, rather than some offset value obtained by inclusion of $-k_BT\left(1+\sum_k \widehat{\beta}_{\xi,k}\right)$. As a result, I proceed according to the definition of $g_{m,\xi}$ presented in Eq. 2.2.72.

Using the definition in Eq. 2.2.72, one can now simplify Eq. 2.2.71, which leads to

$$X_\xi = \frac{1}{e} X_1^{n_{surf,\xi}} \exp\left(-\beta n_{surf,\xi} g_{m,\xi}\right). \tag{2.2.73}$$

This is the final working expression for the modified mole fraction of an aggregate ξ in solution. The simplest form for $g_{m,\xi}$ is encountered in the case of a single-component, nonionic surfactant solution, where one finds that Eq. 2.2.72 reduces to $g_{m,\xi} = g_{mic,\xi} - k_BT$. For a binary nonionic surfactant solution, one finds that $g_{m,\xi} = g_{mic,\xi} - k_BT\left(1 + \alpha \ln \alpha_1 + (1-\alpha)\ln(1-\alpha_1)\right)$, where I use α to indicate $\alpha_{\xi,s}$ for one of the two surfactants s present in solution, and α_1 to indicate $\alpha_{m(s)}$ for that same surfactant (defined in Eq. 2.2.68). This is the same result as Eq. [4] of Ref. [84].

The monomer modified mole fraction for a given species i can be determined by rearranging the species material balance in Eq. 2.2.35 and dividing through by $\sum_i N_i$. This yields

$$X_{m(i)} = \frac{N_i}{\sum_i N_i} - \sum_\xi n_{\xi,i} X_\xi, \tag{2.2.74}$$

where the various N_i's in Eq. 2.2.74 are considered fixed by the formulator. The term X_1 in Eq. 2.2.73 can then be calculated from Eq. 2.2.74 as $X_1 = \sum_s X_{m(s)}$. Since the $g_{m,\xi}$'s will generally depend upon the number of monomers in solution (see Eq. 2.2.72), closing the material balances for all the species requires use of a non-linear solver.

For a single-component, nonionic surfactant solution, Eq. 2.2.72 reduces to

$$g_{m,\xi} = g_{mic,\xi} - k_BT \tag{2.2.75}$$

2.2.4 Property Prediction with Gibbs Free Energy Minimization

Making predictions for the equilibrium behavior of surfactants in solution using the MT theory involves either: (i) prediction of the aggregation behavior of a surfactant system at a specified set of overall solution conditions (i.e., specification of temperature, pressure, and mole numbers of all surfactants, salts, and other additives), or (ii) prediction of the solution conditions which give rise to a desired aggregation behavior. Both cases require a model for $g_{m,\xi}$ for any possible aggregate that can form in solution, such that calculation of Eq. 2.2.73 is possible across all feasible aggregate shapes and sizes. Development of such a model is the subject of Section 2.3.

An important example of the second class of predictions involves determination of the critical micelle concentration (cmc) for a single or mixed surfactant system. Here, the formulator is typically interested in determining the minimum total surfactant concentration at which micelles in solution can be practically found, when keeping the relative compositions of the surfactants fixed. These compositions could be freely chosen or dictated by the nature of the industrial surfactants used in the formulation. For example, certain surfactant manufacturing processes naturally lead to a polydispersity in surfactant chain lengths.[132]

Determination of the cmc is complicated by the fact that Eq. 2.2.73 is a smooth function of $\{N_i\}$.[35] That is, as one steadily increases the total surfactant concentration from zero, there is no single total surfactant concentration at which micelles suddenly and discontinuously appear in solution (as would be the case with a first-order phase transition[133]). Instead, one observes a (usually narrow) transition region in which micelles begin to appear, rapidly increase in number, and then ultimately dominate the component population distribution relative to the monomers.[34] As a result, several definitions for the cmc can be proposed which, although similar, are not strictly identical in value.[134] Experimentally, the onset of aggregation impacts many measurable properties (such as surface tension, or electrical conductivity, in the case of ionic surfactants) in a manner which leads to a transition in slope in the plot of

these properties against total surfactant concentration. The cmc is often determined as the point of intersection between lines fit to the property curve on either side of the property transition.[134] Theoretically, one common strategy is to define the cmc as that concentration at which the total fraction of surfactant molecules present in micellar form achieves a small, fixed value (e.g., 2%).[133] Another strategy is to define the cmc as that concentration at which the number of surfactant molecules in micellar form achieves some fixed, absolute value.[135,136] The cmc can also be defined based on consideration of the free energy of micellization itself.[35,88] This last strategy to define the cmc is the one that I will pursue here, due to the convenience of its calculation within an MT framework.

Writing Eq. 2.2.73 in a slightly rearranged manner illustrates the origin of the transition region and the appropriateness of this last definition of the cmc:

$$X_\xi = \frac{1}{e}\left(\frac{X_1}{\exp\left(\beta g_{m,\xi}\right)}\right)^{n_{surf,\xi}}. \tag{2.2.76}$$

In Eq. 2.2.76, the term $X_1^{n_{surf,\xi}}$ can be viewed as the probability of finding $n_{surf,\xi}$ surfactant molecules at a particular point in space to form the micelle ξ. First, for $g_{m,\xi} \geq 0$, micellization is clearly not favored, since $\exp\left(\beta g_{m,\xi}\right) > 1$, and $X_\xi \ll X_1$ under all conditions, due to the $n_{surf,\xi}$ exponent, where $n_{surf,\xi} \gg 1$, typically. For $g_{m,\xi} < 0$, as X_1 increases (through the addition of surfactant to the system), there may come a point prior to the solubility limit of the surfactant where the term $\exp\left(\beta g_{m,\xi}\right)$ is of the same order as X_1. In this case, micelles will clearly have non-negligible values for X_ξ. Recall that, by definition, X_ξ must be less than 1. As a result, the term $\exp\left(\beta g_{m,\xi}\right)$ provides a clear limit to X_1: the exponent $n_{surf,\xi}$ in Eq. 2.2.76 will cause a violation of $X_\xi > 1$ for $X_1/\exp\left(\beta g_{m,\xi}\right)$ slightly greater than 1. Of course, this situation is avoided by the specific non-linear functional form of the material balances – further addition of surfactant to the system leads to addition of surfactant to the micelle population, rather than to the monomer population.[35]

The functional form of Eq. 2.2.76 has other ramifications. Specifically, the population distribution of micelles is quite peaked around an optimal micelle shape, size, and

composition, due to the exponential term containing $g_{m,\xi}$.[112] Consider, for example, two micelles, labeled ξ and ξ' and possessing modified free energies of micellization $g_{m,\xi}$ and $g_{m,\xi'}$ and aggregation numbers $n_{surf,\xi}$ and $n_{surf,\xi'}$. Then, the ratio of modified mole fractions is given by

$$\frac{X_\xi}{X_{\xi'}} = X_1^{n_\xi - n_{\xi'}} \exp\left(-\beta\left(n_{surf,\xi} g_{m,\xi} - n_{surf,\xi'} g_{m,\xi'}\right)\right). \tag{2.2.77}$$

For the special case where the aggregation numbers are equal (say, to some value n_{surf}, but differing in actual composition), Eq. 2.2.77 yields

$$\frac{X_\xi}{X_{\xi'}} = \exp\left(-\beta n_{surf}\left(g_{m,\xi} - g_{m,\xi'}\right)\right). \tag{2.2.78}$$

Clearly, Eq. 2.2.78 shows that small differences in g_m between the two micelles are magnified through multiplication by $n_{surf,\xi}$ and exponentiation, such that the micelle with lower (more negative) g_m will greatly dominate.

As a result, the arguments above – regarding the point at which X_1 approximately balances $\exp(\beta g_{m,\xi})$ as being representative of the transition region – can be more precisely quantified by determining the micelle shape, size, and composition which leads to a minimum value for $g_{m,\xi}$. This optimal micelle (denoted ξ^*) can then be considered as representative of the population as a whole, due to the peakedness of the typical distribution.[112] The cmc can then be defined as that total surfactant concentration equal to $\exp(\beta g_{m,\xi^*})$, where g_{m,ξ^*} is the modified free energy of micellization for the optimal micelle ξ^*.[81,88]

For a deeper discussion of the implementation of this approach, including a means for interpolating geometries when the optimal micelle is found to be a cylinder or bilayer (modeled in Section 2.3 as infinite in extent), see Section 2.4.

2.3 Molecular Model for the Micellization Free Energy

2.3.1 Phenomenological Thought Process

In Section 2.2, a definition for the cmc for a surfactant system was provided which relies on the minimization of the modified free energy of micellization, $g_{m,\xi}$, relative to the properties which describe the micelle ξ (i.e., micelle shape, size, and composition, where composition indicates the number and type of surfactants and bound counterions present). This quantity was defined in Eq. 2.2.72, which I repeat here for clarity:

$$g_{m,\xi} \equiv g_{mic,\xi} - k_B T \left(1 + \sum_s \alpha_{\xi,s} \ln \alpha_{m(s)}\right) - k_B T \left(\sum_k \widehat{\beta}_{\xi,k} \left(1 + \ln X_{m(k)}\right)\right), \quad (2.3.1)$$

where the free energy of micellization $g_{mic,\xi}$ represents the net free-energy change (on a per-surfactant molecule basis) that is obtained by the formation of a micelle from its constituents under infinite-dilution conditions. The functionality of $g_{mic,\xi}$ was specified in Eq. 2.2.65 as

$$g_{mic,\xi} = \frac{\mu^0_\xi}{n_{surf,\xi}} - \sum_s \alpha_{\xi,s} \mu^0_{m(s)} - \sum_k \widehat{\beta}_{\xi,k} \mu^0_{m(k)}. \quad (2.3.2)$$

It is the development of an expression for $g_{mic,\xi}$ that requires a phenomenological thought process, because the reference-state chemical potentials are not otherwise known. Because the Gibbs free energy is a state function, any path which describes the micellization process and affords convenient calculation of the change in free energy along each step of the path can be utilized (even one which is not physically realizable in practice). Here, I will present a thought process which meets these criteria and has been developed and refined in the Blankschtein group at MIT.[81] For the remainder of this section, I will assume that a particular micelle ξ is being discussed (with its relevant descriptors), and I will drop the index ξ itself, for convenience. I will use

the symbol g to represent Gibbs free energies on a per-surfactant molecule basis, and the symbol f to represent Helmholtz free energies on a per-surfactant molecule basis. Subscripts of g or f will denote a particular step in the thought process, and the choice of g or f itself depends on whether a particular step is associated with a constant pressure or constant volume process, respectively. All steps involve constant temperature and aggregation number.

(1) Starting with the surfactant monomers and counterions in their reference states of infinite dilution, the system is completely discharged, such that all point charges are moved to infinity (a *discharge* step, with free-energy change g_{disch}).

(2) The bonds between surfactant head and tail are chemically broken.

(3) Surfactant tails are removed from solution into an oil-like phase of pure tails (a *transfer* step, with free-energy change g_{tr}).

(4) For a multicomponent system, the tail phases are mixed in the specified composition (a *mixing* step, with free-energy change g_{mix}).

(5) A cavity is created in solution with a volume of the appropriate shape and size for the micelle, through work against the system pressure (a *pressure-volume work* step, with free-energy change $P_{ext}v_{avg}$, where P_{ext} is the pressure external to the micelle: the solution pressure). [Note that I have introduced this term for the first time in the Blankschtein group, recognizing its value to the thought process.]

(6) The micelle core is formed by transferring $n_{surf,\xi}$ surfactant tails from the oil-like phase into the cavity, introducing a core-water interface (an *interface formation* step, with free-energy change f_{int}).

(7) The surfactant tails are partially ordered such that one end lies at the interface (for re-attachment of the head groups in the next step), and the core is pressurized non-uniformly to ensure constant density (a *packing* step, with free-energy change f_{pack}).

(8) Head groups are chemically re-attached to the tails, giving rise to a steric layer on the core exterior (a *steric* step, with free-energy change f_{st}).

(9) Charges of head groups and counterions are restored in an electrostatic charging process, bringing the point charges back from infinity (a *charging* step, with

free-energy change f_{ch}).

Note that the micelle core is typically assumed to be incompressible (i.e., has a constant volume). Therefore, free energies related to the micelle core or to the micelle core-water interface (which correspondingly exhibits a constant surface area), including Steps (6) to (9), are Helmholtz free energies. My inclusion of the pressure-volume work term in Step (5) reconciles the inclusion of these Helmholtz free energies with the Gibbs free energies of the other steps, Steps (1), (3), and (4), and their inclusion in g_{mic} as a whole. Note that the free energies related to bond breaking and forming encountered in Steps (2) and (8) are assumed to be equal and opposite in sign. As a result, they are not included in subsequent analysis. (The term f_{st} will be seen to include only the free energy due to head group interactions and not a free-energy contribution due to bond formation.) Application of the thought process to a single-surfactant system (for which Step (4) is not needed) is illustrated in Figure 2-1.

Keeping the comments above in mind, we can now write an expression for g_{mic} as a linear sum of the free-energy changes associated with the various steps in the thought process (thereby traversing the fictitious path linking the two real end states of dispersed and aggregated constituents). Specifically,

$$g_{mic} = g_{disch} + g_{tr} + g_{mix} + (P_{ext}v_{avg} + f_{int} + f_{pack} + f_{st} + f_{ch}), \qquad (2.3.3)$$

where the terms grouped in parentheses on the right-hand side represent a Gibbs free energy term related to the construction of the fixed-volume micelle in solution. (Recall that g_{mic} and its various free-energy contributions are cast on a per-surfactant molecule basis. Multiplication of any of the terms in Eq. 2.3.3 by n_{surf} yields the total free-energetic contribution of that term to the overall free energy of micellization, $\underline{G}_{mic}$. For example, the total packing free energy can be computed as $\underline{F}_{pack} = n_{surf}f_{pack}$.)

Note that the discharge, transfer, and mixing terms are functions of aggregate composition, but not typically of the shape or size of an aggregate. In fact, this is why these terms can be directly cast as Gibbs free energies (i.e., Steps (1), (3), and

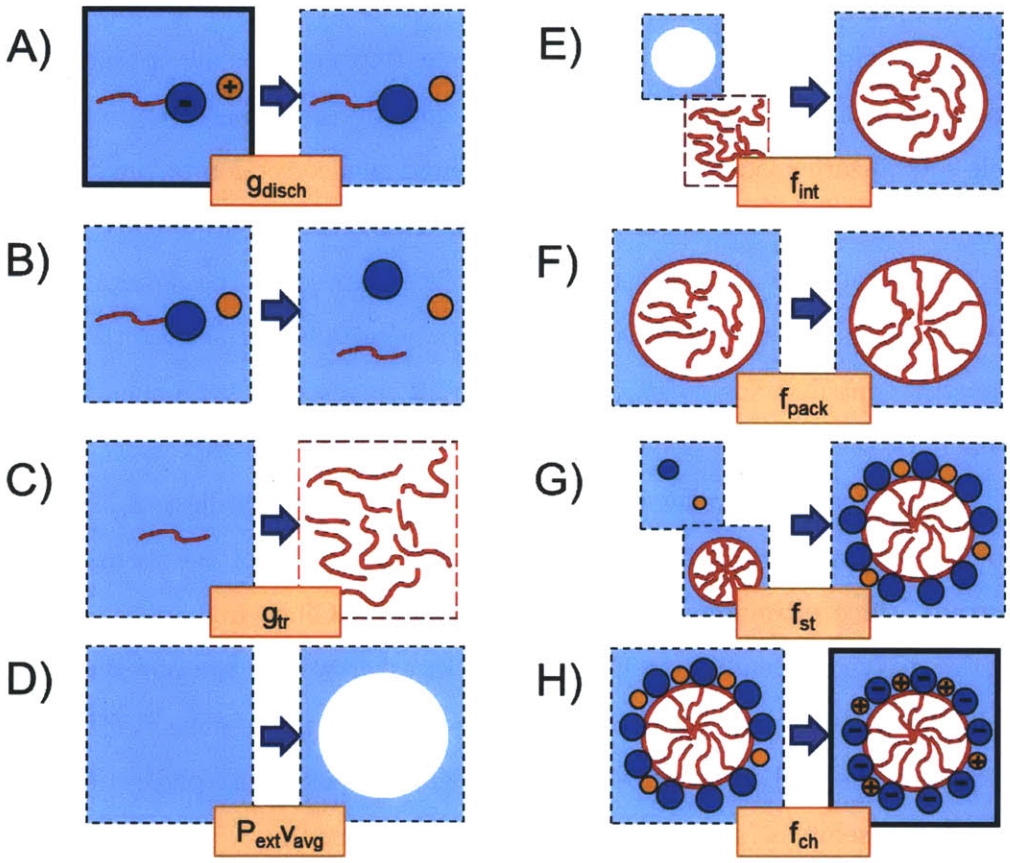


Figure 2-1: Illustration of the phenomenological thought process used in the molecular-thermodynamic theory for predicting surfactant micellization behavior. In this example, the various steps A) through H) link the initial state of an anionic surfactant at infinite dilution to the fully-formed micelle at infinite dilution. The anionic surfactant is depicted as possessing a charged head (blue circle) and linear alkyl tail (red line), with a positively-charged counterion (orange). Step A) involves discharging of the system (Step (1) in the text). Step B) involves separation of the head and tail of the surfactant (Step (2) in the text). Step C) involves transfer of the surfactant tail into a bulk oil-like phase (Step (3) in the text). Step D) involves formation of a cavity in solution (Step (5) in the text). Step E) involves filling the cavity with an oil droplet obtained from the bulk oil-like phase (Step (6) in the text). Step F) involves packing of the surfactant chains under various constraints (Step (7) in the text). Step G) involves reattachment of the head groups and binding of counterions (Step (8) in the text). Step H) involves re-charging the species (Step (9) in the text).

(4) are carried out in solution or in a bulk oil-like phase, under constant pressure). However, when solubilizates are present (not considered in this Thesis), the mixing term can vary throughout the core, and the thought process needs to be modified accordingly. Specifically, an additional mixing step needs to be coupled with the packing step, since there is a translational entropy of mixing attributable to the mobile solubilizate species.[119] This can be viewed as a correction to the bulk mixing, arising from the inhomogeneous nature of the micelle core environment.

In the literature, one finds various notations employed for the terms in Eq. 2.3.3. Since the pressure-volume term is very small in magnitude (see below), the Helmholtz free-energy contributions to g_{mic} are sometimes represented symbolically as Gibbs free energies (i.e., all the free-energy contributions are expressed using the symbol g).[81] Some researchers prefer using the symbol $\Delta\mu$.[111] Regardless of the notation, of course, the precise functional form used to model each of the terms is what is important, since ultimately g_{mic} itself is certainly a Gibbs free energy, as is clear from the derivation leading to Eq. 2.2.65. In other words, the observables of the overall system (the surfactant solution) correspond to those of an NPT ensemble, even though the individual micelles are modeled as possessing properties derivable from NVT ensembles (since incompressibility is assumed).

I will now proceed to describe and motivate the functional forms of each of the free-energy contributions to g_{mic} found in Eq. 2.3.3.

2.3.2 Discharge Free-Energy Contribution

The discharge free energy is derived assuming that the initially dispersed, charged species in solution (surfactants and ions) possess charges that are localized such that they can be modeled as charged spheres. For example, with a surfactant such as sodium dodecyl sulfate, the -1 charge of this surfactant (when dissociated from its sodium counterion) is assumed to be localized onto the sulfate head group. I have found that my own ab initio results typically indicate that some charge does extend into the alkyl tail, but typically less than 5-10%. Therefore, for modeling purposes, a charged sphere representing the sulfate head is a reasonable model. For monoatomic

ions such as sodium, calcium, etc., the charged-sphere approximation is of course a natural one.

The total discharge free energy is a summation of contributions from each micelle constituent in solution:

$$\underline{G}_{disch} = \sum\nolimits_s n_s g_{disch,s} + \sum\nolimits_k n_k g_{disch,k}, \tag{2.3.4}$$

where, as in earlier sections, s indicates a surfactant index, and k indicates an ion index. On a per-surfactant molecule basis (i.e., dividing through by n_{surf}, the micelle surfactant aggregation number), Eq. 2.3.4 becomes

$$g_{disch} = \sum\nolimits_s \alpha_s g_{disch,s} + \sum\nolimits_k \widehat{\beta}_k g_{disch,k}, \tag{2.3.5}$$

where the definitions of α_s and $\widehat{\beta}_k$ introduced in Section 2.2.3 have been used.

It now remains to develop a functional form for the discharge free energy of individual ions. Note that, if a species is nonionic, there is no charge to release, and, therefore, $g_{disch,s} = 0$.

For anions and cations, one can determine $g_{disch,x}$ (where x represents s or k) through the use of a charging process calculation, which involves integration of a differential equation

$$dw = \psi_r dq, \tag{2.3.6}$$

where dw is the infinitesimal work related to charging a sphere of radius r and having an existing surface electric potential ψ_r with an amount of charge equal to dq.[137] The discharge free energy is equal to the negative of the integral of this equation with bounds on q ranging from 0 to q_x, the final charge of the ion x. The Debye-Hückel result is given by:[88]

$$g_{disch,x} = -\frac{e_0^2}{2\epsilon_b} \frac{z_x^2}{r_{h,x}\left(1 + \kappa r_{h,x}\right)}, \tag{2.3.7}$$

where e_0 is the elementary charge, $z_x = q_x/e_0$, $\epsilon_b = 4\pi\epsilon_0\eta_w$ (ϵ_0 is the vacuum dielectric permittivity, and η_w is the relative dielectric permittivity (i.e., the dielectric constant) of the aqueous solvent), $r_{h,x}$ is the hydrated radius of ion x, and κ is the inverse Debye-

Hückel screening length, which can be computed as[90]

$$\kappa = \left(\frac{8\pi e_0^2 I}{\epsilon_b k_B T} \right)^{1/2}, \tag{2.3.8}$$

where I is the ionic strength of the bulk solution,

$$I = \frac{1}{2} \sum_x z_x^2 C_x, \tag{2.3.9}$$

which is a function of $\{C_x\}$, the set of all ion concentrations in the bulk (monomer form). Note that this includes ions k that may have $\beta_k = 0$ (e.g., co-ions). It does not include micelles, since these are assumed to be neutralized by a diffuse cloud of ions surrounding each individual micelle (see Sections 2.3.9 and 2.4.3).[90] Concentrations can be estimated quite accurately from modified mole fractions for monomer species (surfactant or ions) through multiplication by the molarity of water (see Appendix F for values), since water is the primary component of these dilute solutions. (Note that the conceptual introduction of diffuse layers surrounding micelles[88,90] will lead to enhancement or depletion of ions from the bulk, affecting the value of I – see Section 2.4.3 for more details.)

For zwitterions (dipolar ions), the Onsager model[138] is used to obtain[90]

$$g_{disch,x} = -\frac{e_0^2}{4\pi\epsilon_0 (2\eta_w + 1)} \frac{d_{sep,x}^2}{r_{h,x}^3}, \tag{2.3.10}$$

where $d_{sep,x}$ is the estimated distance between charges within the dipole. A common example of a dipolar ion is a betaine surfactant head group.

In summary,

$$g_{disch,x} = \begin{cases} 0, \text{ if nonionic} \\ -\frac{e_0^2}{2\epsilon_b} \frac{z_x^2}{r_{h,x}\left(1+\kappa r_{h,x}\right)}, \text{ if ionic} \\ -\frac{e_0^2}{4\pi\epsilon_0 (2\eta_w+1)} \frac{d_{sep,x}^2}{r_{h,x}^3}, \text{ if zwitterionic} \end{cases}. \tag{2.3.11}$$

2.3.3 Transfer Free-Energy Contribution

For the transfer free-energy contribution, it is desired to link the infinite-dilution reference state of each surfactant tail in aqueous solution with the state of that surfactant tail in its pure, oil-like phase. The relation of these two reference states is most effectively calculated at the solubility limit of the surfactant tail, which is the condition at which the oil-like phase is in thermodynamic equilibrium with the aqueous phase. Assuming solution ideality (i.e., the solubility limit is still quite dilute), the chemical potential of the surfactant tail at saturation is

$$\mu_{tail(s),aq} = \mu^0_{tail(s),aq} + k_B T \ln X^{sat}_{tail(s),aq}, \tag{2.3.12}$$

where $\mu^0_{tail(s),aq}$ is the reference-state chemical potential for the *tail* of surfactant s at infinite-dilution conditions, and $X^{sat}_{tail(s),aq}$ is the solubility limit (ordinary mole fraction, at saturation) of this tail. In the pure oil-like phase, the chemical potential of the surfactant tail is

$$\mu_{tail(s),pure} = \mu^0_{tail(s),pure}, \tag{2.3.13}$$

where $\mu^0_{tail(s),pure}$ is the reference-state chemical potential of the *tail* of surfactant s in a pure state.

Equating the chemical potentials in Eqs. 2.3.12 and 2.3.13 (a procedure possible at the solubility limit, where the two phases are in thermodynamic equilibrium), one obtains

$$\mu^0_{tail(s),aq} + k_B T \ln X^{sat}_{tail(s),aq} = \mu^0_{tail(s),pure}. \tag{2.3.14}$$

Through rearrangement of Eq. 2.3.14, the transfer free-energy contribution associated with surfactant s (which is, specifically, that associated with the transfer of its tail) is calculated as the difference between the two reference-state chemical potentials associated with the tail. That is,

$$g_{tr,s} = \mu^0_{tail(s),pure}(T,P) - \mu^0_{tail(s),aq}(T,P) = k_B T \ln X^{sat}_{tail(s),aq}. \tag{2.3.15}$$

The mole fraction of the surfactant tail in aqueous solution at saturation, $X^{sat}_{tail(s),aq}$, can be estimated from solubility models derived for analogous oils. It is computed as

$$X^{sat}_{tail(s),aq} = \frac{S_{tail(s),aq}}{M_{aq}}, \tag{2.3.16}$$

where $S_{tail(s),aq}$ is the solubility of the tail (in molar units) of surfactant s, and M_{aq} is the molarity of water (the dominant component in dilute solution). The solubility $S_{tail(s),aq}$ can be obtained from one of the group-contribution models described in Section F.5 of Appendix F.

The total transfer free-energy contribution is then given by

$$\underline{G}_{tr} = \sum\nolimits_s n_s g_{tr,s}, \tag{2.3.17}$$

where the counterions are not considered to have a transfer contribution. (For counterions possessing hydrophobic portions, see the work of Srinivasan and Blankschtein[88] for an appropriate treatment.)

The corresponding transfer free-energy contribution on a per-surfactant molecule basis is obtained by dividing Eq. 2.3.17 through by n_{surf}, as follows:

$$g_{tr} = \sum\nolimits_s \alpha_s g_{tr,s}. \tag{2.3.18}$$

2.3.4 Mixing Free-Energy Contribution

The mixing free-energy contribution involves a purely entropic, ideal combinatorial approach to mixing that does not discriminate between subdomains of the micelle core.[88,90] That is, mixing is assumed to be homogeneous throughout the small volume in solution occupied by a micelle at infinite dilution, and the molecules being mixed are assumed to exhibit neither enthalpic synergism nor antagonism in their discharged states. For systems containing hydrocarbon and fluorocarbon-based surfactants, this contribution could be modified according to a regular solution theory approach (i.e., an approach involving enthalpic interaction parameters),[139] to reflect

their generally antagonistic behavior.[121] The total mixing free-energy contribution, under these assumptions, is given by

$$\underline{G}_{mix} = k_B T \left(\sum\nolimits_s n_s \ln\left(\frac{n_s}{n}\right) + \sum\nolimits_k n_k \ln\left(\frac{n_k}{n}\right) \right), \tag{2.3.19}$$

where n is the total aggregation number of the micelle, n_s is the number of surfactants s in the micelle, and n_k is the number of ions k in the micelle, such that $n = \sum_s n_s + \sum_k n_k$. The arguments of the logarithms in Eq. 2.3.19 can also be cast in terms of compositional variables, as follows:

$$\underline{G}_{mix} = k_B T \left[\sum\nolimits_s n_s \ln\left(\frac{\alpha_s}{1 + \sum_{k'} \widehat{\beta}_{k'}}\right) + \sum\nolimits_k n_k \ln\left(\frac{\widehat{\beta}_k}{1 + \sum_{k'} \widehat{\beta}_{k'}}\right) \right]. \tag{2.3.20}$$

On a per-surfactant molecule basis, the mixing free-energy contribution is given by[90]

$$g_{mix} = k_B T \left[\sum\nolimits_s \alpha_s \ln\left(\frac{\alpha_s}{1 + \sum_{k'} \widehat{\beta}_{k'}}\right) + \sum\nolimits_k \widehat{\beta}_k \ln\left(\frac{\widehat{\beta}_k}{1 + \sum_{k'} \widehat{\beta}_{k'}}\right) \right]. \tag{2.3.21}$$

2.3.5 Pressure-Volume Free-Energy Contribution

The pressure-volume work term, $P_{ext} v_{avg}$, represents the work required to create a cavity in the solvent to accommodate the micelle core,[140,141] on a per-surfactant molecule basis. This contribution is typically negligible in magnitude compared to the other free-energy contributions to g_m. To see this, note that typical tail volumes for surfactants possessing tails of 6 to 16 carbons in length range from roughly 200 to 500 $\AA^3$ (about 30 $\AA^3$ per CH_2 group and 50 $\AA^3$ per CH_3 group – see Appendix F for a more accurate analysis). The system pressure, P_{ext}, is usually taken to be atmospheric pressure. Through unit conversion, 1 atm is equal to 101325 Pa (i.e., J/m^3) or 1.01325×10^{-25} $J/\AA^3$. At a system temperature of $T = 298.15K$, the thermal energy $k_B T$ is approximately $4.1 \times 10^{-21} J$, since $k_B = 1.381 \times 10^{-23}$ J/K. Therefore, 1 atm equals $2.5 \times 10^{-5} k_B T/\AA^3$ at this temperature. If we take the upper value for the typical molecular tail volume encountered in surfactant systems, 500 $\AA^3$,

the pressure-volume term $P_{ext}v_{avg}$ is found to be 0.013 k_BT. This is much smaller than the other free-energy contributions, which are typically between +1 to +5 k_BT for the interfacial, packing, steric, and electrostatic contributions, and between −10 to −30 k_BT for the transfer contributions (all on a per-surfactant molecule basis). In fact, typical errors in these other terms are usually in the range of 0.1 to 0.5 k_BT in practice. Therefore, the pressure-volume work term can be safely neglected.[141] I include it in the formalism here for completeness, where it provides a clear bridge between the Helmholtz free-energy terms related to the fixed volume micelle and the Gibbs free-energy terms related to the fixed pressure solution.

Note that the total pressure-volume work term is obtained by multiplying $P_{ext}v_{avg}$ by n_{surf}. This yields $P_{ext}\underline{V}$, where $\underline{V}$ is the total volume of the aggregate under consideration (i.e., $\underline{V}_\xi$).

2.3.6 Interfacial Free-Energy Contribution

The interfacial-free energy contribution captures the free-energy penalty due to restoring contacts between the surfactant tails and the bulk solvent at the micelle core-water interface when the cavity created by the pressure-volume work term is filled with surfactant tail material. This can most easily be modeled using a macroscopic treatment of the interface, where the Helmholtz free energy due to creation of the interface is found by integrating σdA (under conditions of constant temperature and core volume) from zero contact area up to the final contact area.[142] Here, σ is the interfacial tension of the micelle core-water interface, which is a function of the curvature of the interface, but presumed to be independent of the amount of area of that interface in contact with water. That is, during integration of dA, σ is presumed constant. If the final contact area is taken to be the area of the micelle core, subtracting out a shielded area due to the presence of head-tail bonds (which have a non-zero width of approximately 21$\mathring{A}^2$ for a carbon-carbon bond), then the total interfacial free energy is given by[81]

$$\underline{F}_{int} = \sigma\left(\underline{A} - \underline{A}_0\right), \tag{2.3.22}$$

where $\underline{A}_0$ is the shielded area, determined from the individual surfactant head-tail bond cross-sectional areas as $\underline{A}_0 = \sum_s n_s a_{0,s}$ (again, typically $a_{0,s} = 21\mathring{A}^2$).[81]

The per-surfactant molecule form of Eq. 2.3.22 is given by[90]

$$f_{int} = \sigma \left(a - \sum_s \alpha_s a_{0,s} \right), \tag{2.3.23}$$

where a is the micelle core area per surfactant molecule ($\underline{A}/n_{surf}$).

It now remains to determine the micelle core-water interfacial tension, σ. This is an interfacial tension that can be presumed to be similar to the interfacial tension of an oil droplet composed of oils analogous to the tails within a micelle core. One typically assumes that σ can be composition-weighted, similar to composition-weighted models for surface tension.[143] Specifically,[90]

$$\sigma = \sum_s \alpha_s \sigma_s. \tag{2.3.24}$$

Finally, the curvature-dependent interfacial tensions are related to planar interfacial tensions via the Gibbs-Tolman-Koenig-Buff equation.[81,144–146] Specifically,

$$\sigma_s = \frac{\sigma_{0,oil(s)}}{1 + (S-1)\,\delta/l_c}, \tag{2.3.25}$$

where $\sigma_{0,oil(s)}$ is the planar interfacial tension between an oil analogous to the tail of surfactant s, S is a shape factor (equal to 1 for bilayers, 2 for cylinders, and 3 for spheres), δ is the Tolman length, and l_c is the core-minor radius of the micelle under consideration (recall that the integration leading to Eq. 2.3.22 does not change l_c). The Tolman length is the distance from the surface of tension to the equimolar dividing surface along the radial coordinate.[144] A correlation for the Tolman length for linear surfactant tails has been developed by Puvvada and Blankschtein,[81] using a Tolman length of $\delta_{C_{11}} = 2.25\mathring{A}$ for a C_{11} tail, and the following scaling relationship:

$$\delta_{C_N} = \frac{\delta_{C_{11}}}{l_{\max,C_{11}}} l_{\max,C_N}, \tag{2.3.26}$$

where C_N represents a linear chain containing N carbon groups, and $l_{\max}$ is the maximum extended chain length, which can be determined from geometric arguments as follows:[35]

$$l_{\max,C_N} = 1.54 + 1.265N \left[\mathring{A}\right]. \tag{2.3.27}$$

2.3.7 Packing Free-Energy Contribution

A thorough discussion of calculation of the packing free energy can be found in Chapter 4. Briefly, a single-chain mean field (SCMF) approach can be used to model single or mixed surfactant micelles. A rotational isomeric state (RIS) model is used as a basis for modeling the internal conformations of surfactant chains, and a maximum entropy formalism is used to derive a probability distribution function (pdf), which allows calculation of chain conformation-dependent properties, the single chain partition function, and expressions for the packing free energy.

Since packing calculations are usually time-intensive, libraries of packing results are prepared in advance to distribute to users of the software implementation of the molecular-thermodynamic theory. These results are stored in the form of a polynomial for each shape type[90,147] (bilayer, cylinder, and sphere) and surfactant tail system, from which data for any particular micelle size can be interpolated. For single surfactant systems, the polynomial is a function of size alone, expressed as powers of $l_c/l_{c,\max}$, where l_c is the core-minor radius of the micelle of interest, and $l_{c,\max}$ is the maximum micelle size attainable (equal to the maximum fully-extended chain length found in the micelle). For mixed systems, the polynomial is a function of size and compositional variables ($N-1$ mole fractions for N total types of surfactant tails), where cross-terms are included in the fit.[148]

The total packing free-energy contribution is expressed in Eqs. 4.2.54 and 4.2.55 of Chapter 4 as a difference in free energy of the surfactant tails in the bulk oil-like state (a "free", unconstrained state, with free energy $\underline{F}_{free}$) and the packed, constrained

micelle core state (with free energy $\underline{F}_{core}$), as follows:

$$\begin{aligned} \underline{F}_{pack} &= \underline{F}_{core} - \underline{F}_{free} \\ &= n_{surf} \sum_s \alpha_s \left[-k_B T \ln \frac{Z_s}{Z_{s,free}} - \sum_l \pi_l \langle v_{\omega_s,l} \rangle \right], \end{aligned} \tag{2.3.28}$$

where the various terms in Eq. 2.3.28 are defined in detail in Chapter 4 (in brief, the first term in the square brackets is a ratio of partition functions, and the second term is an energetic contribution arising from the non-uniform pressurization of the micelle core to achieve constant density throughout). Determination of f_{pack} is readily accomplished by dividing Eq. 2.3.28 by n_{surf}. That is,

$$f_{pack} = \frac{\underline{F}_{pack}}{n_{surf}} = \sum_s \alpha_s \left[-k_B T \ln \frac{Z_s}{Z_{s,free}} - \sum_l \pi_l \langle v_{\omega_s,l} \rangle \right]. \tag{2.3.29}$$

2.3.8 Steric Free-Energy Contribution

The steric free-energy contribution to g_{mic} arises due to excluded-volume interactions between all the surfactant heads and bound counterions in the micelle of interest. These interactions are typically assumed to take place in a narrow shell, or steric layer, concentric to the micelle core-water interface. That is, head groups are assumed to be neither chain-like nor exhibiting a diffuse arrangement. Instead, the steric layer is assumed to be sufficiently thin, such that all steric interactions can be adequately modeled by free-energy calculations for excluded-area disks on a surface representing the steric layer. Then, the theory of gaseous insoluble films can be used to derive the free energy, as I will show below. Note that the steric free-energy contribution should be viewed as a departure function, in that it relates the real interactions of the head groups and bound counterions in this gaseous film to the ideal-gas case. (See Ref. [113] for a discussion of departure functions.)

To derive $\underline{F}_{st}$, one first represents the micelle core in aqueous solution as a two-phase system of oil and water, possessing an interfacial tension σ_0 in the absence of surfactant head groups and bound counterions. The extensive Helmholtz free energy

for this system is a function of the following independent variables:

$$\underline{F} = \underline{F}\left(T, \underline{V}, \underline{A}, \{N_i\}\right), \tag{2.3.30}$$

where T is the system temperature, $\underline{V}$ is the system volume, $\underline{A}$ is the extensive area of the interface, and $\{N_i\}$ represents the set of mole numbers for all the species present. The change in this free energy upon (reversible) extension of the interface is given by[142]

$$\left(\frac{\partial \underline{F}}{\partial \underline{A}}\right)\Bigg|_{T,\underline{V},\{N_i\}} = \sigma \tag{2.3.31}$$

where σ is the interfacial tension (equal to σ_0 when the interface is clean – i.e., devoid of head groups or counterions). In total differential form,

$$d\underline{F} = \sigma d\underline{A}, \quad \text{for fixed } T, \underline{V}, \{N_i\}. \tag{2.3.32}$$

The free energy of a gaseous film representing the steric layer is calculated as[142]

$$\underline{F}^{film} = \underline{F} - \underline{F}_0, \tag{2.3.33}$$

where $\underline{F}_0$ is the Helmholtz free energy for the system lacking surfactant heads or bound counterions (for which $\sigma = \sigma_0$). The total differential of $\underline{F}^{film}$, for fixed $T, \underline{V}, \{N_i\}$, is then given by

$$d\underline{F}^{film} = d\underline{F} - d\underline{F}_0 = (\sigma - \sigma_0)\, d\underline{A}, \tag{2.3.34}$$

where Eq. 2.3.32 has been used to obtain the right-hand side relationship. Substituting the definition of surface pressure,[142]

$$\Pi \equiv \sigma_0 - \sigma, \tag{2.3.35}$$

into Eq. 2.3.34, one obtains

$$d\underline{F}^{film} = -\Pi d\underline{A}. \tag{2.3.36}$$

For an ideal gas film, Eq. 2.3.36 reduces to

$$d\underline{F}^{film}_{ideal} = -\Pi^{ideal} d\underline{A}. \tag{2.3.37}$$

Writing the total differential of the steric free energy as a difference between the total differential of the Helmholtz free energies of the real and ideal films, one obtains

$$\begin{aligned} d\underline{F}_{st} &= d\underline{F}^{film} - d\underline{F}^{film}_{ideal} \\ &= -\left(\Pi - \Pi^{ideal}\right) d\underline{A}. \end{aligned} \tag{2.3.38}$$

This departure function[113] is then integrated from conditions of ideality to the real conditions. Here, ideality is obtained in the limit of infinite area, where head groups and bound counterions do not interact. Therefore, the bounds of integration are from $+\infty$ ($\underline{F}_{st} = 0$) to $\underline{A}$, the area of the micelle core under consideration. Specifically,

$$\underline{F}_{st} = -\int_{\infty}^{\underline{A}} \left(\Pi - \Pi^{ideal}\right) d\underline{A}'. \tag{2.3.39}$$

where the prime is used to distinguish between the area as a variable of integration and in the limits of integration.

Under ideal conditions, the surface pressure is given by:[142]

$$\Pi^{ideal} = \frac{N k_B T}{\underline{A}}, \tag{2.3.40}$$

where N is the total number of molecules participating in the steric layer.

However, under real conditions, a more appropriate model (analogous to the steric part of a van der Waals equation of state for a three-dimensional gas) is[142]

$$\Pi = \frac{N k_B T}{\underline{A} - \underline{A}_H}, \tag{2.3.41}$$

where $\underline{A}_H$ is an extensive co-area, taken to be equal to the sum of surfactant and bound counterion head areas. Note that $\underline{A}_H$ is fixed when $\{N_i\}$ is fixed, since these

head areas are considered independent of the number of molecules present. Note also that, if the heads are instead flexible, this may no longer be the case[86,126]! In this case, it would be preferable to implement a packing theory for the surfactant head, similar to the packing theory for the surfactant tail described in Chapter 4, but incorporating the effect of solvent (consider, for example, the approach applied to polymer brushes described in Ref. [149]). Finally, another alternative model for compact heads is that provided by Scaled-Particle Theory.[150-152] The interested reader is referred to a paper by Yuet and Blankschtein for further details of applying this approach in the context of the $\underline{F}_{st}$ calculations.[86]

Inserting Eqs. 2.3.40 and 2.3.41 into Eq. 2.3.39 and integrating yields

$$\begin{aligned} \underline{F}_{st} &= -Nk_BT\int_{\infty}^{\underline{A}}\left(\frac{1}{\underline{A}'-\underline{A}_H}-\frac{1}{\underline{A}'}\right)d\underline{A}' \\ &= -Nk_BT\ln\left(\frac{\underline{A}-\underline{A}_H}{\underline{A}}\right), \end{aligned} \tag{2.3.42}$$

where $\underline{A}$ is again the area of the micelle core, and the prime is used in the integrand for clarity (as in Eq. 2.3.39).

Assuming that all the species in the micelle have a head group (note that, for certain solubilization systems, encapsulated organic solubilizates may not), N is equal to n, the total aggregation number of the micelle (the sum of surfactant and counterion counts). Then, division through by n_{surf} yields[111]

$$f_{st} = -k_BT\left(1+\sum_k \widehat{\beta}_k\right)\ln\left(1-\frac{\underline{A}_H/n_{surf}}{a}\right) \tag{2.3.43}$$

where $a = \underline{A}/n_{surf}$ is the micelle core area per surfactant molecule. Equation 2.3.43 can be applied equally well to single-component or mixed surfactant systems.[83]

Recall that $\underline{A}_H$ corresponds to the sum of all the head areas in the micelle. If this is written as $\underline{A}_H = \sum_s n_s a_{H,s} + \sum_k n_k a_{H,k}$, where $a_{H,s}$ is the head area of a single surfactant s, and $a_{H,k}$ is the cross-sectional area of a single counterion k, then Eq.

2.3.43 becomes[90]

$$f_{st} = -k_B T \left(1 + \sum_k \widehat{\beta}_k\right) \ln\left(1 - \frac{\sum_s \alpha_s a_{H,s} + \sum_k \widehat{\beta}_k a_{H,k}}{a}\right). \tag{2.3.44}$$

Finally, the determination of the head area of a particular surfactant or bound counterion is typically made using geometric arguments, where one uses bond lengths, bond angles, and van der Waals radii to determine an appropriate disk area for modeling purposes. Ethoxylated surfactants with short head chain lengths can also be reasonably modeled using this approach.[90]

2.3.9 Charging Free-Energy Contribution

Once the surfactant micelle has been completely assembled from its constituent surfactant molecules and bound counterions, the final step involves re-charging the ionic and zwitterionic species. Similar to the discharging free-energy step discussed in Section 2.3.2, a charging process is used to determine the free-energy contribution due to this step, involving integration of an expression for the differential work of charging, $dw = \psi_r dq$, where ψ_r is again a surface potential (at radial position r, where radial has the meaning of a direction normal to the half-plane of a bilayer, normal to the central axis of a cylinder, or coincident with the usual radius of a sphere), and dq is an infinitesimal unit of charge. The process of charging a body from zero initial charge to a final charge against a background of all other ions in solution, which are themselves fully charged, is referred to as the Guntelberg charging process.[137]

A key difference between charging a micelle and charging an ion is that the micelle may comprise multiple individual charged species, of differing molecular geometry. Assuming that each surfactant species has a characteristic "distance-to-charge" for each charged group in the head (a geometric parameter representing the distance between the micelle core-water interface and the average location of the charge, along the radial direction normal to the interface; zwitterions have two such parameters), each unique distance-to-charge represents a surface with its own potential ψ_r which must be charged through the Guntelberg charging process. Figure 2-2 illustrates this

situation for a micelle containing two anionic surfactants with differing distances-to-charge, giving rise to two charged surfaces. Binding of counterions to the micelle is considered to be limited to binding to the charged surfaces, thereby simplifying the modeling of the microstructure of the micelle interfacial region.[88] Since all the species in the micelle core are considered to be laterally mobile, the charged surfaces are considered to be well-represented by a smeared, uniform charge density (in contrast to the use of discrete charges). The total free energy of charging is a sum of the energies associated with the Guntelberg charging process for each charged surface, and is given by

$$\underline{F}_{ch} = \sum_{cs} \int_0^{Q_{cs}} \psi_{cs} dq_{cs}, \tag{2.3.45}$$

where cs indicates a particular charged surface, ψ_{cs} is the surface potential of that charged surface (a function of q_{cs}, the charge of the surface), and Q_{cs} is the final charge of the surface, which can be calculated as

$$\begin{aligned} Q_{cs} &= n_{surf} e_0 z_{cs} \\ &= n_{surf} e_0 \left(\sum\nolimits_s \alpha_s z_{s,cs} + \sum\nolimits_k \widehat{\beta}_{k,cs} z_{k,cs} \right), \end{aligned} \tag{2.3.46}$$

where e_0 is the elementary charge, z_{cs} is the average valence of charge on the surface on a per-surfactant molecule basis, s indexes surfactants, k indexes ions, and $z_{s,cs}$ and $z_{k,cs}$ are the valences of surfactant s and ion k, respectively, found on charged surface cs. Note that $\widehat{\beta}_{k,cs}$ is the degree of binding of ion k to charged surface cs. The overall degree of binding for ion k is simply $\widehat{\beta}_k = \sum_{cs} \widehat{\beta}_{k,cs}$, a sum across charged surfaces, since bound ions are limited to these surfaces, by assumption. If a particular surfactant s does not contribute to the charge of a particular charged surface cs, then $z_{s,cl} = 0$ for that surface (this will always be the case for nonionic surfactants in a charged micelle environment, for example, since they do not themselves possess charge). Finally, if a particular ion k does not bind to charged surface cs, then $\widehat{\beta}_{k,cs} = 0$.

Another concept introduced in Figure 2-2 is the Stern surface. This surface marks

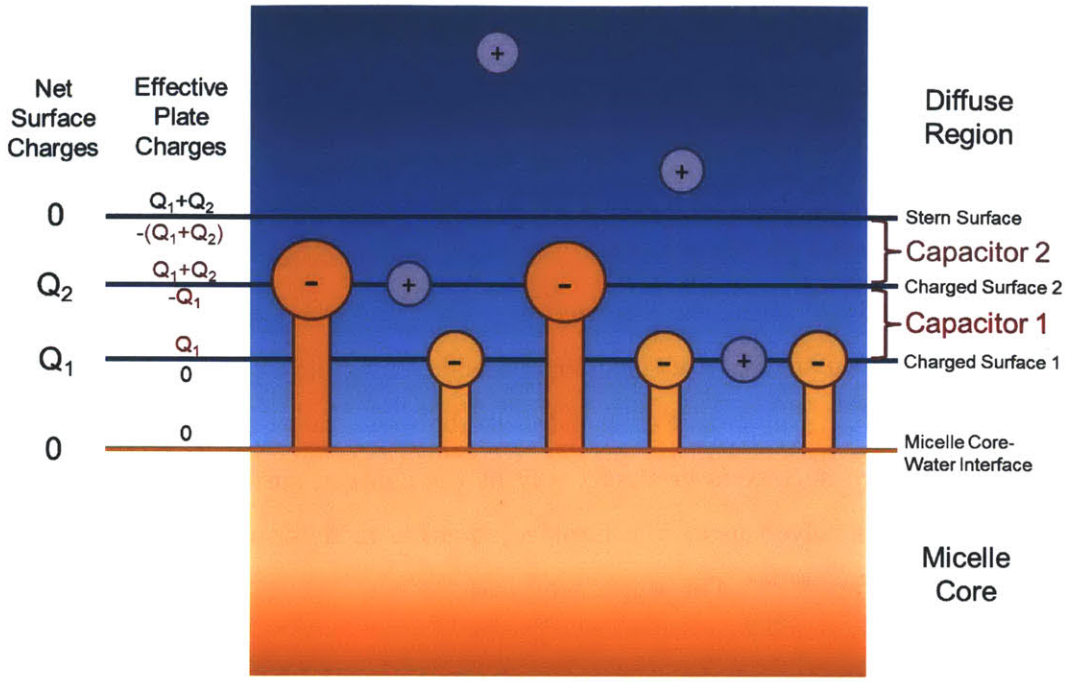


Figure 2-2: Schematic of an electrostatic model for micelles containing charged surfactants, as presented in the text. A small section of a micelle core-water interface is shown, with the micelle core represented as an orange background, and the aqueous solution represented as a blue background. Two types of anionic surfactant heads and their common counterion are shown. (Water and surfactant tails are not shown, for clarity.) The first surfactant head type (in light orange) possesses a negative charge whose location gives rise to charged surface 1 (with a net charge Q_1). The second surfactant head type (in dark orange) possesses a negative charge whose location gives rise to charged surface 2 (with a net charge Q_2). The positively-charged counterions (in purple) are assumed to be either: (i) bound to one of the charged layers, (ii) present in the diffuse region, or (iii) present in the bulk solution (not shown).[88] The counterions located in the diffuse region serve to neutralize the net charge of the micelle, such that the micelle is not considered to contribute to solution ionic strength. The diffuse region begins at the Stern surface, which is positioned at a distance from the micelle core-water interface equal to the length of the longest charged surfactant head (here, the second head type) plus the hydrated radius of the counterion.[90] In the interfacial region between the micelle core and the Stern surface, bound counterions are assumed to be restricted to the charged surfaces, such that the Laplace equation can be applied between charged surfaces. This permits the use of a capacitor-in-series model,[112] with each surface acting as a capacitor plate (effective plate charges for capacitor charging calculations are shown). In the diffuse region, counterions are assumed to distribute according to the Boltzmann statistics, and the Poisson-Boltzmann equation can be applied, with the effective charge at the Stern surface equal to that necessary to cancel out the outermost capacitor plate charge (since the Stern surface itself contains no charge). This yields an effective charge equal to the net charge of the micelle.

the start of a diffuse layer surrounding the charged micelle, which contains ions that are assumed to distribute according to the Boltzmann statistics and the Poisson equation, based on the surface potential of the micelle at the Stern surface.[153] The position of the Stern surface is located beyond the outermost charged surface at a distance equal to the radius of the surfactant head giving rise to that charged surface plus the radius of the hydrated counterion. Counterions within the Stern surface are considered bound (and must be localized to a charged surface), while counterions at or beyond the Stern surface are considered mobile (unbound).

Since the radially-dependent charge density between the micelle core-water interface and the Stern surface is non-zero only at the charged surfaces, by assumption, Eq. 2.3.45 can be solved using the Laplace equation to determine the potentials at each charged surface.[90,154] This approach yields[85]

$$\underline{F}_{ch} = n_{surf} \sum_{cs} \left[\frac{2\pi}{4\pi\epsilon_0 \eta_{int} k_B T} \frac{1}{a_{cs}} \left(\sum_{cs'=1}^{cs} \frac{Q_{cs'}}{n_{surf}} \right)^2 F\left[S, d_{cs}, R_{cs}\right] \right] + \int_0^{Q_f} \psi_{ss} dq_{ss}, \tag{2.3.47}$$

where the first term represents the work of charging capacitors in series (where each charged surface and the Stern surface represent the charged plates of a capacitor), and the second term is the work of charging the micelle to a total final charge, based on a potential situated at the Stern surface (ss). In Eq. 2.3.47, η_{int} is the dielectric constant in the micelle interfacial region (which extends from the micelle core-water interface to the Stern surface) – chosen by Goldsipe and Blankschtein[90] to be equal to $\eta_w/2$, a_{cs} is the area per surfactant molecule ($\underline{A}_{cs}/n_{surf}$) at the charged surface, $\sum_{cs'=1}^{cs} Q_{cs'}/n_{surf}$ is the cumulative charge per surfactant molecule up to the current charge surface under consideration, and F is a term which captures the geometry dependence of each capacitor – a function of: (i) S, the micelle shape, (ii) d_{cs}, the distance between the current charged surface and the next charged surface (i.e., the capacitor-plate separation), and (iii) R_{cs}, the radial position of the charged surface, from the center of mass of the micelle. The shape-dependence of the function F is as

follows:[90]

$$F\left[S, d_m, R_m\right] = \begin{cases} d_m, \text{ for bilayers } (S=1) \\ R_m \ln\left(1 + d_m/R_m\right), \text{ for cylinders } (S=2) \\ d_m/\left(1 + d_m/R_m\right), \text{ for spheres } (S=3) \end{cases} . \tag{2.3.48}$$

The final term in Eq. 2.3.47, a charging process which effectively charges the micelle to its final charge Q_f as if that charge were uniformly located at the Stern surface, is based on the electrostatic potential calculated at the Stern surface, ψ_{ss}. (Note that $Q_f = \sum_{cs} Q_{cs}$, and that the capacitor terms can be viewed as corrections arising from the actual distribution of charge into charged layers.) Solution for this potential involves consideration of the ions in the diffuse layer, and, hence, solution of a Poisson-Boltzmann (PB) equation. Ohshima, Healy, and White have provided an analytical approximation to the full non-linear PB equation for monovalent counterions,[155] which simplifies calculation of the charging process integral. Specifically,[90]

$$\int_0^{Q_f} \psi_{ss} dq_{ss} = \frac{k_B T}{e_0} Q_f \left[y_f - \frac{1}{s_f} \left(4\cosh\left(\frac{y_f}{2}\right) - 4 + \frac{8\left(S-1\right)}{x_0} \ln\left(\cosh\left(\frac{y_f}{4}\right)\right)\right)\right], \tag{2.3.49}$$

where the following dimensionless groups apply:

$$s_f = \frac{4\pi e_0}{\epsilon_w k_B T \kappa} \frac{q_f}{a_s}, \tag{2.3.50}$$

where e_0 is the elementary charge, κ is the inverse Debye-Hückel screening length (defined in Eq. 2.3.8), $\epsilon_w = 4\pi\epsilon_0\eta_w$ (where ϵ_0 is the dielectric permittivity of vacuum, and η_w is the relative dielectric permittivity of the aqueous solvent), $q_f = Q_f/n_{surf}$ is the final micelle charge on a per-surfactant molecule basis, and a_{ss} is the area per surfactant molecule at the Stern surface, given by

$$y_f = \frac{e_0 \psi_{ss,f}}{k_B T}, \tag{2.3.51}$$

where $\psi_{ss,f}$ is the final value of the surface potential at the Stern surface upon com-

pletion of charging, and

$$x_0 = \kappa R_{ss}, \tag{2.3.52}$$

where R_{ss} is the radial position of the Stern surface from the center of mass of the micelle.

The key result of Ohshima, Healy, and White is a non-linear, shape-dependent relationship between s_f and y_f, given by[155]

$$s_f = 2\sinh\left(\frac{y_f}{2}\right) + \frac{2(S-1)}{x_0}\tanh\left(\frac{y_f}{4}\right). \tag{2.3.53}$$

Given s_f and x_0, both of which can be calculated during a molecular-thermodynamic solution process (see Section 2.4) for any trial micelle (fixing geometry, size, and charge) and bulk monomer composition (fixing ionic strength), y_f can be determined from Eq. 2.3.53 using a non-linear solver such as MINPACK.[156] Equation 2.3.49 can then be evaluated, followed by Eq. 2.3.47 to determine $\underline{F}_{ch}$. Finally, the charging free-energy contribution per surfactant molecule can be computed as follows:

$$f_{ch} = \underline{F}_{ch}/n_{surf}. \tag{2.3.54}$$

This completes the description of the various free-energy contributions to g_{mic}. I will now proceed with a description of the computational process used to predict the micellization properties for a surfactant solution using the molecular-thermodynamic theory.

2.4 Description of the Computational Process

2.4.1 Program CS-PREDICT Overview

I have implemented the molecular-thermodynamic (MT) theory presented in this Chapter into a computer program called CS-PREDICT, which includes capabilities for making predictions using the traditional MT theory (including my new contributions regarding the individual terms: (i) updated densities, solubilities, and interfacial

tensions, (ii) revised packing equations, and (iii) the addition of new capabilities for branched surfactant property prediction). Program CS-PREDICT is also capable of making predictions using a combined computer-simulation/molecular-thermodynamic (CS-MT) theory based on the work of Stephenson and coworkers[108–110] and my own contributions, as discussed in Chapter 3. Since the CS-MT theory ultimately involves replacement of the transfer free-energy contribution (g_{tr}) to the free energy of micellization (g_{mic}) with an updated value, $g_{tr,CSMT}$, the solution process for the two approaches (MT and CS-MT) once this replacement has been made is identical.

Once details of the surfactants and ions in the solution have been entered by the user, the program proceeds as follows.

Overall, three environments are defined: "solution" (total solution conditions), "bulk" (monomer components), and "micelle", for which compositional variables for all species are tracked. The micelle environment contains additional parameters related to the geometry of the micelle and its size (i.e., its core-minor radius). There is also the concept of a diffuse layer which surrounds charged micelles and ensures electroneutrality.[90]

The following six steps are then undertaken by the user:

(1) A vector is initialized for the bulk environment containing modified mole fractions for surfactant monomers, counterions, and coions (if salt is present). Initial guesses for these values are based, when possible, on user-selected solution environment conditions (total salt added, total surfactant added, surfactant composition). If a cmc prediction is requested, the total surfactant modified mole fraction in the bulk environment is set equal to a guess (based on the relationship between the cmc and the modified free energy of micellization, g_m – see Section 2.2.4) of $\exp\left(\beta g^*_{m,guess}\right)$, where $g^*_{m,guess} = g_{tr,avg} + \Delta g$, $g_{tr,avg}$ is the total solution environment composition-weighted g_{tr} (note that the user *must* specify the relative compositions of surfactants added), and Δg is a guess for the sum of free-energy penalties arising from the other terms. A number on the order of $\Delta g = 5.5\ k_BT$ has been found to work quite well for this purpose for most systems (i.e., this guess typically leads to convergence).

(2) *Outer Loop*

A non-linear solver, MINPACK,[156] is used with a numerical Jacobian (HYBRD routine) to solve a set of non-linear material balances of the form

$$X_i = \widehat{X}_{m(i)} + \sum_{\xi} n_{\xi,i} X_\xi + E_{dl,i}, \tag{2.4.1}$$

where X_i is the total mole fraction of species i in solution, $\widehat{X}_{m(i)}$ is an effective mole fraction for the monomeric form of species i associated with the bulk environment, $n_{\xi,i}$ is the aggregation number of species i in aggregate ξ, X_ξ is the modified mole fraction of that aggregate in solution ($X_\xi = N_\xi / \sum_i N_i$, where N is a mole number), and $E_{dl,i}$ is an enhancement (or depletion) term arising from the diffuse layer surrounding electrostatically-charged micelles.[90] A hat has been used with $\widehat{X}_{m(i)}$ to distinguish it from $X_{m(i)}$, the actual mole fraction of species i in monomeric form. The role of the diffuse layer is that it enhances ions possessing charges with sign opposite to the sign of the micelle net charge, and depletes ions possessing charges with the same sign. The $\left\{\widehat{X}_{m(i)}\right\}$ are used to calculate the effective ionic strength of the bulk solution.[90] Note that, by writing out the species material balances in terms of components, along the lines of Eq. 2.2.35, dividing the sum over components into monomer and aggregate contributions, and dividing this expression by the total number of molecules in the solution, one obtains $X_i = X_{m(i)} + \sum_\xi n_{\xi,i} X_\xi$, which, when compared to Eq. 2.4.1, shows that $X_{m(i)} = \widehat{X}_{m(i)} + E_{dl,i}$. The functional form for $E_{dl,i}$ is discussed in Section 2.4.3.

As discussed in Section 2.2, water is not assumed to be a constituent of any aggregate (although water, of course, does hydrate aggregates). Water is also assumed to be neither enhanced or depleted by diffuse layer considerations, since it is neutral in charge (i.e., $E_{dl,w} = 0$). As a result, the water material balance is always satisfied by setting the bulk environment mole fraction for water equal to the solution mole fraction.

The non-linear solver seeks the set $\left\{\widehat{X}_{m(i)}\right\}$ which leads to satisfaction of Eq. 2.4.1. The X_i's are either determined through user input, or, for a cmc calculation,

$X_{surf} = \sum_s X_s = \exp(\beta g_m^*)$ is the target (such that the $\{X_i\}$ may change for each iterative step). Note that $\{X_\xi\}$ and $\{E_{dl,i}\}$ are functions of $\left\{\widehat{X}_{m(i)}\right\}$ through the ionic strength. Determination of g_m^* once $\left\{\widehat{X}_{m(i)}\right\}$ has been set represents an inner loop of the solution process.

(3) A vector for the micelle environment containing estimates for the core-minor radius, the surfactant mole fractions, and the degrees of counterion binding per charge layer is initialized. Reasonable estimates include radii close to the maximum chain length of surfactants present in the system (a consideration arising from packing), non-zero mole fractions equal to the relative surfactant mole fractions in the bulk solution, and non-zero values for the degrees of counterion binding in each charge layer (say, $\widehat{\beta}_{k,cl,\xi} = 0.01$, if k is a counterion). Note that the use of non-zero values for the mole fractions and the degrees of counterion binding is necessary, because entropy of mixing terms containing natural logarithms do not allow arguments of zero.

(4) *Inner Loop (Across Shapes)*

Once the bulk environment is set, optimal modified free energies of micellization can be computed for the principal geometries of interest. An optimizer capable of extremizing a nonlinear objective function with constraints, such as CONMAX,[157] is suitable. The optimizer minimizes g_m^{shape} by varying the micelle descriptors of size and composition to obtain the optimized values g_m^{shape*} for the three principal shapes of sphere, cylinder, and bilayer. The constraints include: (i) $R_{\min} \leq l_{c,\xi} \leq R_{\max}$, where $R_{\min}$ and $R_{\max}$ are limits on $l_{c,\xi}$ set by the domain of the packing polynomials available for the surfactants present (technically functions of the surfactant tail and the shape of the micelle, with $R_{\max}^{bilayer} < R_{\max}^{sphere}$, due to packing considerations), (ii) $0 \leq \alpha_{s,\xi} \leq 1$, and (iii) $\widehat{\beta}_{k,cl,\xi} \geq 0$. Note that, for the evaluation of the constraints related to $\{\alpha_{s,\xi}\}$, all but one of the $\alpha_{s,\xi}$ values are obtained from the parameter vector of CONMAX, with the remaining value computed from $\sum_s \alpha_{s,\xi} = 1$.

(5) Once the inner loop is completed for the three principal geometries, the geometry yielding the minimal (most negative) free energy of micellization is determined. If this is a cylinder or bilayer, an interpolation scheme of the type described in Section 2.4.2 is necessary in order to evaluate the term $\sum_\xi n_{\xi,i} X_\xi$ in Eq. 2.4.1. Otherwise, for

spheres, this summation becomes a single term centered on the optimal parameters $l_{c,\xi}^{sph*}$, $\left\{\alpha_{s,\xi}^{sph*}\right\}$, and $\left\{\widehat{\beta}_{k,cl,\xi}^{sph*}\right\}$. Diffuse layer calculations are also performed to obtain $\{E_{dl,i}\}$.

(6) Once the outer loop is completed, the mass balance constraints are satisfied, and, therefore, the optimal micelle properties can be determined from the final CONMAX results. In particular, monomer and micelle compositions and concentrations can be determined, and, for cmc predictions, the total surfactant concentration X_{surf} in the solution is reported. (This value can be converted into molar units by multiplying by the molarity of water at the temperature of interest.)

2.4.2 Geometry Interpolation

The functional forms of the various free-energy contributions to the free energy of micellization ($g_{mic,\xi}$) that depend on the size and shape of aggregate ξ are based on constant-curvature micelle geometries. (These are the terms found within the parentheses of Eq. 2.3.3 – that is, the interfacial, packing, steric, and electrostatic contributions.) Specifically, the three principal shapes considered include spheres, infinite cylinders, and infinite bilayers, where only the spheres are finite in extent. Infinite cylinders and infinite bilayers are idealized geometries, but it is known experimentally that surfactants can undertake finite rodlike[158] or discoidal[159,160] geometries in solution, which are finite counterparts to infinite cylinders and infinite bilayers, respectively. To enable prediction of the properties of these finite shapes from the infinite ones, one can introduce interpolating functions which describe the free energy of micellization based on calculations for the idealized geometries. For example, for discoidal micelles, one may consider the disk as a finite bilayer wrapped around its waist with a toroidal rim.[161] For rodlike micelles, a typical approach is to use a spherocylindrical model, in which the finite cylindrical body is capped at both ends with a hemispherical endcap.[81,162,163] Each of the domains may have different attributes. For example, there is experimental evidence to suggest that the hemispherical endcaps of a finite rodlike micelle possess a radius which is larger than that of the cylindrical body.[164]

In the remainder of this section, I focus on the interpolation scheme for spherocylindrical micelles. The reader interested in a discoidal model is directed to the work of Srinivasan and Blankschtein in this area.[161] (At the cmc, most surfactants exhibit either spherical or spherocylindrical aggregate formation. The formation of disks is comparatively rare.[165]) As in Section 2.3, I will drop use of the index ξ for convenience, except where necessary for clarity.

Under the prevailing bulk solution conditions (temperature, pressure, monomer concentrations), one can always determine the minimum free energy of micellization for each of the three principal shapes based on the MT model developed in Section 2.3. Each of these shapes will have an optimal parameter set (denoted $*$): (i) a core-minor radius, l_c^{shape*}, (ii) a set of surfactant mole fractions $\left\{\alpha_s^{shape*}\right\}$, where s indicates the species of surfactant, and (iii) a set of degrees of counterion binding $\left\{\widehat{\beta}_{k,cl}^{shape*}\right\}$, where k indicates the species of counterion and cl indicates the charge layer for binding ($\widehat{\beta}_k^{shape*} = \sum_{cl} \widehat{\beta}_{k,cl}^{shape*}$). When the geometry which attains the global minimum free energy of micellization is an infinite cylinder, the parameters for the optimal sphere are taken to provide the details for the composition and size (l_c^{sph*}) of the hemispherical endcaps. That is, the energetics of the endcaps and the cylindrical body are assumed to be decoupled. The fact that the hemispherical endcaps are observed to be larger than the cylinder experimentally[164] lends some credence to this approach: spheres are usually found to have optimal core-minor radii larger than their cylindrical counterparts due to packing considerations.[166] The optimal cylindrical parameters are then taken to provide the details for the cylindrical body. Note that this fixes the core-minor radius of the body as l_c^{cyl*}, but does not fix the cylinder length – that is, one finds monodispersity in the core minor radius but polydispersity in the length! This will be handled accordingly below in developing the population distribution for spherocylinders.

Assuming incompressibility, the total aggregation number corresponding to the surfactant molecules located in the endcaps can be calculated from the optimal sphere

parameters as[90]

$$n_{surf}^{sph*} = \frac{\frac{4}{3}\pi \left(l_c^{sph*}\right)^3}{v_{avg}^*}, \tag{2.4.2}$$

where v_{avg}^* is the optimal composition-average tail volume, given by

$$v_{avg}^* = \sum_s \alpha_s^* v_{tail,s}. \tag{2.4.3}$$

For a given spherocylinder of total surfactant aggregation number n_{surf}^{tot} (cylinder plus endcaps), the interpolated free energy of micellization is taken to be[81,90]

$$g_{mic}\left(n_{surf}^{tot}\right) = g_{mic}^{cyl*} + \frac{n_{surf}^{sph*}}{n_{surf}^{tot}}\left(g_{mic}^{sph*} - g_{mic}^{cyl*}\right) \tag{2.4.4}$$

which is a simple composition-weighted average according to the number of surfactant molecules present in each domain. Note that $g_{mic}^{sph*} > g_{mic}^{cyl*}$, since the cylinder is presumed to be the global optimum shape. This implies that $g_{mic}\left(n_{surf}^{tot}\right) > g_{mic}^{cyl*}$ for all finite n_{surf}^{tot} values, which would appear to indicate preferable formation of infinite cylinders in solution. Recall Eq. 2.2.76, which provides the following expression for the modified mole fraction of any aggregate:

$$X_\xi = \frac{1}{e}\left(\frac{X_1}{\exp\left(\beta g_m\right)}\right)^{n_{surf}}. \tag{2.4.5}$$

Since the material balances guarantee that X_1, the total monomer modified mole fraction, will be less than $\exp\left(\beta g_m\right)$ (otherwise the constraint $X_\xi < 1$ would be violated for almost all values of n_{surf}), a large increase in n_{surf}^{tot} (i.e., n_{surf}) will lead to X_ξ, the modified mole fraction of the spherocylinder, approaching zero. Accordingly, the dominant shape in solution will, in fact, be finite spherocylinders, although they possess a slightly less negative g_{mic} (and g_m).

In order to evaluate Eq. 2.4.5, one requires evaluation of g_m, the modified free energy of micellization. Recall that this is related to g_{mic} via Eq. 2.2.72. Specifically,

$$g_m \equiv g_{mic} - k_B T\left(1 + \sum_s \alpha_s \ln \alpha_{m(s)}\right) - k_B T\left(\sum_k \widehat{\beta}_k \left(1 + \ln X_{m(k)}\right)\right), \tag{2.4.6}$$

and that $\alpha_{m(s)}$ and $X_{m(k)}$ are surfactant and counterion monomer properties, respectively. Calculation of the surfactant mole fractions $\{\alpha_s\}$ and the degrees of counterion binding $\left\{\widehat{\beta}_k\right\}$ in Eq. 2.4.6 follow an identical interpolation scheme to the one used in Eq. 2.4.4:[90]

$$\alpha_s\left(n_{surf}^{tot}\right) = \alpha_s^{cyl*} + \frac{n_{surf}^{sph*}}{n_{surf}^{tot}}\left(\alpha_s^{sph*} - \alpha_s^{cyl*}\right), \tag{2.4.7}$$

and

$$\widehat{\beta}_k\left(n_{surf}^{tot}\right) = \widehat{\beta}_k^{cyl*} + \frac{n_{surf}^{sph*}}{n_{surf}^{tot}}\left(\widehat{\beta}_k^{sph*} - \widehat{\beta}_k^{cyl*}\right). \tag{2.4.8}$$

Inserting Eqs. 2.4.4, 2.4.7, and 2.4.8 into Eq. 2.4.6 yields

$$g_m\left(n_{surf}^{tot}\right) = \left(\begin{array}{l} \left[g_{mic}^{cyl*} + \frac{n_{surf}^{sph*}}{n_{surf}^{tot}}\left(g_{mic}^{sph*} - g_{mic}^{cyl*}\right)\right] + \\ -k_BT\left(1 + \sum_s\left[\alpha_s^{cyl*} + \frac{n_{surf}^{sph*}}{n_{surf}^{tot}}\left(\alpha_s^{sph*} - \alpha_s^{cyl*}\right)\right]\ln\alpha_{m(s)}\right) + \\ -k_BT\left(\sum_k\left[\widehat{\beta}_k^{cyl*} + \frac{n_{surf}^{sph*}}{n_{surf}^{tot}}\left(\widehat{\beta}_k^{sph*} - \widehat{\beta}_k^{cyl*}\right)\right]\left(1 + \ln X_{m(k)}\right)\right) \end{array}\right), \tag{2.4.9}$$

or

$$\begin{aligned} g_m\left(n_{surf}^{tot}\right) = & \left(\begin{array}{l} g_{mic}^{cyl*} - k_BT\left(1 + \sum_s \alpha_s^{cyl*}\ln\alpha_{m(s)}\right) + \\ -k_BT\left(\sum_k \widehat{\beta}_k^{cyl*}\left(1 + \ln X_{m(k)}\right)\right) \end{array}\right) + \\ & + \frac{n_{surf}^{sph*}}{n_{surf}^{tot}}\left(\begin{array}{l} \left(g_{mic}^{sph*} - g_{mic}^{cyl*}\right) - k_BT\left(\sum_s\left(\alpha_s^{sph*} - \alpha_s^{cyl*}\right)\ln\alpha_{m(s)}\right) + \\ -k_BT\left(\sum_k\left(\widehat{\beta}_k^{sph*} - \widehat{\beta}_k^{cyl*}\right)\left(1 + \ln X_{m(k)}\right)\right) \end{array}\right) \end{aligned} \tag{2.4.10}$$

which, through comparison with Eq. 2.4.6 simplifies to

$$g_m\left(n_{surf}^{tot}\right) = g_m^{cyl*} + \frac{n_{surf}^{sph*}}{n_{surf}^{tot}}\left(g_m^{sph*} - g_m^{cyl*}\right). \tag{2.4.11}$$

In other words, the interpolating scheme applied in Eqs. 2.4.4, 2.4.7, and 2.4.8 leads to an identical interpolation on g_m as a consequence.

At this stage, it is convenient to introduce the following definition:

$$\Delta\mu \equiv n_{surf}^{sph*}\left(g_m^{sph*} - g_m^{cyl*}\right) + k_BT, \tag{2.4.12}$$

where $\Delta\mu$ is a measure of the extent of the free energy penalty paid by the n_{surf}^{sph*} surfactant molecules present in the endcaps for being in the suboptimal hemispherical environment relative to the optimal cylindrical environment. Note that the addition of the thermal energy k_BT was done for convenience to allow for a simplification described below.

Rearranging Eq. 2.4.12, one obtains

$$g_m^{sph*} - g_m^{cyl*} = \frac{\Delta\mu - k_BT}{n_{surf}^{sph*}}, \tag{2.4.13}$$

which can be substituted into Eq. 2.4.11 to yield

$$g_m\left(n_{surf}^{tot}\right) = g_m^{cyl*} + \frac{\Delta\mu - k_BT}{n_{surf}^{tot}}. \tag{2.4.14}$$

Inserting this expression into Eq. 2.4.5 yields the following result:

$$\begin{aligned} X_\xi &= \frac{1}{e}\left(\frac{X_1}{\exp\left(\beta\left(g_m^{cyl*} + \frac{\Delta\mu - k_BT}{n_{surf}}\right)\right)}\right)^{n_{surf}} \\ &= \frac{1}{\exp\left(\beta\Delta\mu\right)}\left(\frac{X_1}{\exp\left(\beta g_m^{cyl*}\right)}\right)^{n_{surf}}. \end{aligned} \tag{2.4.15}$$

At this stage, one can next write explicit expressions for the total amount of surfactant molecules in spherocylindrical form in solution, and for the total number of spherocylindrical micelles (in modified mole fraction form). As discussed above, the optimal cylinder fixes the core-minor radius of the spherocylinder bodies, but not their length. Both of the expressions which follow are therefore infinite sums over all possible spherocylinder lengths, which range from the suboptimal sphere limit (no cylindrical growth) to the infinite cylinder limit.

Defining $q \equiv X_1 \exp\left(-\beta g_m^{cyl*}\right)$ and $K = \exp\left(\beta\Delta\mu\right)$, where q and K are both quantities independent of the length of the spherocylinder, the total amount of surfactant in spherocylindrical form (the first moment, M_1, of the micelle population

distribution[124]) is given by

$$M_1 \equiv \sum_{n_{surf}=n_{surf}^{sph*}}^{\infty} n_{surf} X_\xi = \frac{1}{K} \sum_{n_{surf}=n_{surf}^{sph*}}^{\infty} n_{surf} q^{n_{surf}}. \tag{2.4.16}$$

The total number of spherocylindrical micelles in solution (the zeroth moment, M_0, of the micelle population distribution[124]) is given by

$$M_0 \equiv X_{mic} = \sum_{n_{surf}=n_{surf}^{sph*}}^{\infty} X_\xi = \frac{1}{K} \sum_{n_{surf}=n_{surf}^{sph*}}^{\infty} q^{n_{surf}}. \tag{2.4.17}$$

If $q < 1$ holds, the following convenient mathematical relationships for power series involving q hold:[167]

$$\begin{aligned} \sum_{n=a}^{\infty} n q^n &= \sum_{n'=0}^{\infty} (n' + a)\, q^{(n'+a)} \\ &= -q^a \frac{(-q - a + qa)}{(1-q)^2}, \end{aligned} \tag{2.4.18}$$

and

$$\begin{aligned} \sum_{n=a}^{\infty} q^n &= \sum_{n'=0}^{\infty} q^{(n'+a)} \\ &= \frac{q^a}{1-q}. \end{aligned} \tag{2.4.19}$$

Inserting Eq. 2.4.18 into Eq. 2.4.16 yields

$$M_1 = \frac{1}{K} q^{n_{surf}^{sph*}} \left[\frac{n_{surf}^{sph*}}{(1-q)} + \frac{q}{(1-q)^2} \right]. \tag{2.4.20}$$

Inserting Eq. 2.4.19 into Eq. 2.4.17 yields

$$M_0 = \frac{1}{K} \frac{q^{n_{surf}^{sph*}}}{1-q}. \tag{2.4.21}$$

In order to satisfy the individual surfactant and counterion material balances, the following expressions (similar in formulation to that of M_1) are also useful. For

surfactants, one finds

$$
\begin{aligned}
& \sum_{n_{surf}=n_{surf}^{sph*}}^{\infty} \alpha_s n_{surf} X_\xi \\
= & \sum_{n_{surf}=n_{surf}^{sph*}}^{\infty} \left(\alpha_s^{cyl*} + \frac{n_{surf}^{sph*}}{n_{surf}} \left(\alpha_s^{sph*} - \alpha_s^{cyl*} \right) \right) n_{surf} X_\xi \\
= & \frac{1}{K} \alpha_s^{cyl*} \sum_{n=n^{sph*}}^{\infty} n_{surf} q^{n_{surf}} + \frac{1}{K} n_{surf}^{sph*} \left(\alpha_s^{sph*} - \alpha_s^{cyl*} \right) \sum_{n=n^{sph*}}^{\infty} q^{n_{surf}} \\
= & \alpha_s^{cyl*} M_1 + n_{surf}^{sph*} \left(\alpha_s^{sph*} - \alpha_s^{cyl*} \right) M_0 \\
= & \alpha_s^{cyl*} \left(\frac{1}{K} q^{n_{surf}^{sph*}} \left[\frac{n_{surf}^{sph*}}{(1-q)} + \frac{q}{(1-q)^2} \right] \right) + n_{surf}^{sph*} \left(\alpha_s^{sph*} - \alpha_s^{cyl*} \right) \left(\frac{1}{K} \frac{q^{n_{surf}^{sph*}}}{1-q} \right) \\
= & \frac{1}{K} q^{n_{surf}^{sph*}} \left(\alpha_s^{sph*} \frac{n_{surf}^{sph*}}{(1-q)} + \alpha_s^{cyl*} \frac{q}{(1-q)^2} \right).
\end{aligned} \tag{2.4.22}
$$

For counterions, it is straightforward to show that the following similar expression holds:

$$
\sum_{n_{surf}=n_{surf}^{sph*}}^{\infty} \widehat{\beta}_k n_{surf} X_\xi = \frac{1}{K} q^{n_{surf}^{sph*}} \left(\widehat{\beta}_k^{sph*} \frac{n_{surf}^{sph*}}{(1-q)} + \widehat{\beta}_k^{cyl*} \frac{q}{(1-q)^2} \right). \tag{2.4.23}
$$

In MT modeling, cylinders and bilayers are viewed as being rigid and possessing constant curvature. This assumption is valid only in dilute regimes. Once rodlike micelles grow to length scales much greater than their species-dependent persistence length, l_p (due to added salt, surfactant, or temperature changes), they become quite flexible.[158] Other concerns may also arise when rodlike micelles grow, including the formation of branches in the cylindrical bodies of the micelles,[168] or interpenetration and entanglement between rods, which greatly impacts solution viscosity and phase behavior.[48,158]

Many studies involving the flexibility of rodlike micelles have been reported. Imae and coworkers investigated a solution of cetyl trimethylammonium bromide (CTAB) surfactant with added NaBr salt at $T = 35°C$,[169] using static light scattering. At zero salt concentration, CTAB forms spherical micelles (with aggregation number 91). At $0.1M$ NaBr, short rods are found to form (with aggregation number 401, and in coexistence with spheres), and these rods lengthen with increasing salt concentration

(with aggregation numbers 1820 for $0.2M$, 5410 for $0.3M$, and 9530 for $0.5M$ NaBr). Based on the reported calculated persistence length of $l_p = 44nm$ in Ref. [169] for rodlike CTAB micelles, one can estimate, based on: (i) a $C15$ model for the alkyl tail (assuming that the methylene group adjacent to the CTAB head remains hydrated[34]), (ii) a fully-extended tail length (and maximum core-minor radius) $l_{c,\max} = 21\mathring{A}$, and (iii) a tail volume $v_{tail} = 432\mathring{A}^3$, that the volume in a rigid subsection of the micelle is approximately $\underline{V}_{rigid} \tilde{} \pi l_{c,\max}^2 l_{persist} = 609600\mathring{A}^3$, with an associated aggegration number $n_{surf,rigid} = \underline{V}_{rigid}/v_{tail} = 1400$. Another study, by Magid and coworkers,[170] focused on using small-angle neutron scattering (SANS) to determine the flexibility and persistence lengths of rodlike micelles formed by sodium dodecyl sulfate (SDS) under conditions of 1 to $2M$ added NaCl at $T = 45°C$. These authors determined persistence lengths having l_p values between 15 and $20nm$, with l_p decreasing with increasing salt concentration. Using $l_p = 15nm$, $l_{c,\max} = 15\mathring{A}$, and $v_{tail} = 324\mathring{A}^3$ (C_{11} tail), one finds $n_{surf,rigid} = 330$. Several other surfactants have been studied in the C_{12} to C_{16} range; these all exhibit l_p values greater than, or equal to, the values discussed here.[158] Therefore, these calculations for $n_{surf,rigid}$ can be viewed as providing a first-pass check for the applicability of the MT model. For cylinders exhibiting substantial growth, adjustments to the MT model may be required, including possible introduction of a suitable $\underline{G}_{\text{interaction}}$ term in Eq. 2.2.7.

In practice, most common linear surfactants at the cmc form micelles with low aggregation numbers at room temperature,[171] indicative of spherical or globular micelles.[72] One common surfactant that does form cylinders at room temperature is hexaethylene glycol monododecyl ether ($C_{12}E_6$), which has an aggregation number of around 300 at $T = 20°C$.[171] $C_{12}E_6$ has a sphere-to-rod shape transition temperature of $T^* = 15°C$.[172] As the number of ethylene oxide groups is increased, it is known that T^* increases, such that dodecyl ethoxylated surfactants with 7 or more ethoxylate units are spherical at room temperature. (T^* also increases with decreasing alkyl chain length: $C_{10}E_6$ forms rods at $T = 30°C$,[173] but not at $T = 25°C$,[174] although $C_{10}E_5$ does.[173]) Note that, based on the approximate calculation for SDS above (another C_{12} surfactant), modeling of $C_{12}E_6$ rods as rigid would appear appropriate.

(In fact, Ravey suggests a value of $25nm$ for the l_p value of $C_{12}E_6$, based on fitting to neutron scattering results,[175] which is even more accommodating in terms of an acceptable aggregation number).

2.4.3 Electrostatic Diffuse Layer

As discussed in Section 2.4.1, the presence of charged micelles in solution gives rise to a diffuse layer which enhances, or depletes, ions on the basis of the sign of their charge relative to the sign of the charge of the micelles. Goldsipe and Blankschtein have provided a thorough discussion of the derivation of these contributions.[90] Below, I summarize their discussion and results for completeness.

First, one assumes that the presence of charged micelles in solution will affect the distribution of the monomeric forms of all charged species (dissociated surfactants and salts). If, on average, micelles are separated in dilute solution at a distance which is greater than the Debye-Hückel screening length, κ^{-1} (a measure of the range of electrostatic interactions), where κ was defined in Eq. 2.3.8, then the total volume of the solution containing monomers (i.e., the volume resulting from subtracting the total volume occupied by micelles from the solution volume) can be partitioned conceptually into diffuse layer and bulk regions. Assuming that these regions have volumes $\underline{V}_{diff}$ and $\underline{V}_{bulk}$, respectively, one can write a material balance for species i (whether a surfactant, a counterion, or a coion) as follows:

$$N_i = C_{i,bulk}\left(\underline{V}_{bulk} + \int_{\underline{V}_{diff}} \exp\left(-\frac{z_i e_0 \psi(r)}{k_B T}\right) dV\right) + \sum_{\xi} n_{\xi,i} N_{\xi}, \qquad (2.4.24)$$

where $C_{i,bulk}$ is the concentration of species i in the bulk region, the integrand represents a Boltzmann distribution for the ions in the diffuse layer, z_i is the charge number of species i, e_0 is the elementary charge, $\psi(r)$ is the electrostatic potential at radial coordinate r for the geometry of interest (where radial represents the direction normal to the half-plane of a bilayer, normal to the central axis of a cylinder, or coincident to the usual radius of a sphere), and the final term is the total number of species i found in aggregate form. If $z_i e_0 \psi(r) \ll k_B T$, then one can simplify the

integrand in Eq. 2.4.24 using a Taylor series expansion, truncated to leading order, to obtain

$$N_i = C_{i,bulk}\left(\underline{V}_{bulk} + \underline{V}_{diff} - \frac{z_i e_0}{k_B T}\int_{\underline{V}_{diff}} \psi\left(r\right) dV\right) + \sum_{\xi} n_{\xi,i} N_{\xi}. \tag{2.4.25}$$

Both the overall solution and the bulk solution are considered electroneutral, which gives rise to the following two constraints:

$$\sum_i z_i N_i = 0, \tag{2.4.26}$$

and

$$\sum_i z_i C_{i,bulk} = 0. \tag{2.4.27}$$

Inserting Eq. 2.4.25 into Eq. 2.4.26, one obtains

$$\sum_i z_i \left[C_{i,bulk}\left(\underline{V}_{bulk} + \underline{V}_{diff} - \frac{z_i e_0}{k_B T}\int_{\underline{V}_{diff}} \psi\left(r\right) dV\right) + \sum_{\xi} n_{\xi,i} N_{\xi}\right] = 0. \tag{2.4.28}$$

The use of Eq. 2.4.27 in Eq. 2.4.28 yields a convenient cancellation, eliminating the bulk and diffuse layer volume terms, such that

$$-\sum_i z_i^2 C_{i,bulk}\left(\frac{e_0}{k_B T}\int_{\underline{V}_{diff}} \psi\left(r\right) dV\right) + \sum_{\xi}\sum_{i} z_i n_{\xi,i} N_{\xi} = 0. \tag{2.4.29}$$

Through further rearrangement of Eq. 2.4.29, the following expression for the unknown integral is obtained:

$$\left(\frac{e_0}{k_B T}\int_{\underline{V}_{diff}} \psi\left(r\right) dV\right) = \frac{\sum_{\xi}\sum_i z_i n_{\xi,i} N_{\xi}}{\sum_i z_i^2 C_{i,bulk}}. \tag{2.4.30}$$

Inserting Eq. 2.4.30 back into Eq. 2.4.25, one obtains

$$N_i = C_{i,bulk}\left(\underline{V}_{bulk} + \underline{V}_{diff}\right) - z_i C_{i,bulk}\frac{\sum_{\xi}\sum_i z_i n_{\xi,i} N_{\xi}}{\sum_i z_i^2 C_{i,bulk}} + \sum_{\xi} n_{\xi,i} N_{\xi}. \tag{2.4.31}$$

Dividing Eq. 2.4.31 through by $\sum_i N_i$ and rearranging yields

$$X_i - \sum_\xi n_{\xi,i} X_\xi = \frac{C_{i,bulk}\left(\underline{V}_{bulk} + \underline{V}_{diff}\right)}{\sum_i N_i} - z_i C_{i,bulk} \frac{\sum_\xi \sum_i z_i n_{\xi,i} X_\xi}{\sum_i z_i^2 C_{i,bulk}}, \tag{2.4.32}$$

where the right-hand side can be interpreted as the mole fraction of species i in solution which is not found in micellar form (i.e., $X_{m(i)} = \widehat{X}_{m(i)} + E_{dl,i}$). One can identify the first term on the right-hand side of Eq. 2.4.32 as the effective mole fraction of species i in monomeric form, based on the bulk concentration $C_{i,bulk}$,

$$\widehat{X}_{m(i)} = \frac{C_{i,bulk}\left(\underline{V}_{bulk} + \underline{V}_{diff}\right)}{\sum_i N_i}. \tag{2.4.33}$$

It follows that the second term on the right-hand side of Eq. 2.4.32 is $E_{dl,i}$, where:

$$E_{dl,i} = -z_i C_{i,bulk} \frac{\sum_\xi \sum_i z_i n_{\xi,i} X_\xi}{\sum_i z_i^2 C_{i,bulk}}. \tag{2.4.34}$$

Next, assume that the micelle size distribution is tightly peaked, such that $\sum_i z_i n_{\xi,i}$ is always of one sign. Then, one can see from Eq. 2.4.34 that species with z_i of the same sign charge as that of the micelles will have $E_{dl,i} < 0$, such that $\widehat{X}_{m(i)} > X_{m(i)}$, an effective enhancement in the bulk concentration of species i (depletion in the diffuse layer). When z_i is of opposite sign (e.g., with counterions), the reverse occurs: $E_{dl,i} > 0$, leading to effective depletion in the bulk and enhancement in the diffuse layer. Again, these depletion and enhancement effects are important in assessing the ionic strength, which is computed as follows:

$$I = \frac{1}{2} \sum_i z_i^2 C_{i,bulk}. \tag{2.4.35}$$

Equation 2.4.33 provides a direct relation between $C_{i,bulk}$ and $\widehat{X}_{m(i)}$.

2.5 Conclusions

In this Chapter, I presented the full derivation of the molecular-thermodynamic (MT) theory for surfactant micellization which I will use in Chapters 5 and 6 to make predictions of the micellization properties of a variety of linear and branched surfactants, respectively. This theory is based on the original MT approach developed by Puvvada and Blankschtein[81] and refined by numerous members of the Blankschtein group,[85–90,112,148,161] including me.

In Section 2.2, I provided a thermodynamic treatment of a bulk phase containing species capable of forming aggregates, and I defined the notion of a component, which can consist of any number of associated molecules, for which a surfactant monomer and surfactant micelle are two examples. Consideration of the free energies of formation and mixing of these components in solution ultimately led to an expression for the mole fraction of a micelle component as a function of the difference in reference state chemical potentials between the micelle and its constituents. I identified this difference as the free energy of micellization, g_{mic}, and described how a phenomenological thought process was necessary to estimate its value. Given a model for g_{mic}, the complete population distribution of micelles in solution as a function of total surfactant concentration can be determined, and useful micellization properties such as the critical micelle concentration can be estimated.

In Section 2.3, I described a suitable thought process for calculating g_{mic} in detail. This thought process provides a convenient, reversible thermodynamic path from an initial state of individual, dispersed surfactants to a final state of a fully-formed micelle at infinite dilution. Each step of the thought process yields a free-energy contribution to g_{mic} which can be calculated using the models provided in Sections 2.3.2 to 2.3.9. I resolved a long-standing issue regarding the precise nature of the various free-energy contributions to the free energy of micellization, through the introduction of a new pressure-volume work term related to the cavity formation in solution necessary to accommodate the micelle core.

Finally, in Section 2.4, I provided the details of my specific computer program

implementation of the MT theory, along with a discussion of two other implementation details: (i) a method for interpolating the values for g_{mic} obtained for spheres and infinite cylinders to estimate g_{mic} for spherocylinders, and (ii) a discussion of the effect of the electrostatic diffuse layer surrounding micelles on the population balances of ions and surfactant species.

In the next Chapter, I will describe how computer simulations can be used to augment the MT theory, providing a platform for the detailed study of the hydration of surfactant micelles. This then allows the determination of several properties of importance to surfactant modeling, including an identification of the hydrated (head) and dehydrated (tail) regions of a surfactant, and a quantification of the change in hydration experienced by a surfactant group upon incorporation of that group into a micelle core.

Chapter 3

Combining Computer Simulations with Molecular Thermodynamics

3.1 Overview

As discussed in Chapter 1, simulations have long been used to gain insight into the equilibrium micellization behavior of surfactants, including studies involving: (i) determination of the overall population distribution of spherical and spherocylindrical micelles,[105,118] (ii) the phase behavior at higher concentrations,[176] (iii) the shape and structure of individual micelles,[96,101,103,177] (iv) the stability of micelles as a function of size,[178] (v) specific ion effects related to the size and charge of counterions,[179,180] (vi) the orientation and ordering of solvent molecules about the micelle,[104,181] and (vii) the orientation of individual surfactant molecules within the micelle.[182] Two excellent reviews of the many contributions in this field are due to Shelley and Shelley[91] for work prior to the year 2000, and to Brodskaya[92] for work since (up to 2011).

Many of the studies referenced above involved a considerable effort in analyzing and interpreting simulated micelle data. However, the work is often limited to the specific simulation performed and the specific surfactant system studied, and, quite frequently, is independent of any theoretical framework for broader predictions. In response, Stephenson and coworkers showed how molecular dynamics simulations could be used within the context of a molecular-thermodynamic (MT) theory in a series

of papers concerned with more accurately quantifying the hydrophobic effect.[108–110] They proposed a framework which blends computer simulations (CS) with an MT theory to create a combined computer-simulation/molecular-thermodynamic (CSMT) framework. In addition to providing a deeper insight into the micellization behavior of common linear surfactants that have already been heavily studied (including the sulfates and ethoxylates), the CSMT framework provides a useful means for investigating increasingly-complex surfactants, including the branched surfactants discussed in Chapter 6.

In this Chapter, I will discuss the CSMT framework in detail and demonstrate how it builds upon the traditional MT approach described in Chapter 2. In Section 3.2, I present the key equations of the CSMT approach. In Section 3.3, I discuss my contribution to the CSMT framework: a novel computational strategy for head/tail identification based on a dividing surface approach. In Section 3.4, I discuss the simulation protocol that I have used in initializing molecular dynamics simulations of surfactants in monomer and micelle form for the linear surfactants discussed in Chapter 5 and for the branched surfactants discussed in Chapter 6. In Section 3.5, I describe the analysis routines that I have applied to my equilibrated simulations in order to extract the information needed to make micellization predictions. Finally, in Section 3.6, I provide some concluding remarks.

3.2 The Combined Computer-Simulation/Molecular-Thermodynamic (CSMT) Framework

3.2.1 Dehydration and Hydration Free-Energy Contributions to the Micellization Free Energy

The CSMT framework is focused on improving those free-energy contributions to the micellization free energy, g_{mic}, which involve characterization of the magnitude of the hydrophobic effect. These are the transfer and interfacial free-energy contributions described in Sections 2.3.3 and 2.3.6 of Chapter 2. In particular, Stephenson and

coworkers suggest that, for a particular simulated aggregate, one can compute an improved estimate of g_{mic} by making the following substitution:[108]

$$g_{tr} + g_{int} \mapsto g_{dehydr} + g_{hydr}, \tag{3.2.1}$$

where g_{tr} is the MT transfer free energy, g_{int} is the Gibbs interfacial free energy, related to the Helmholtz interfacial free energy via

$$g_{int} = f_{int} + P_{ext} v_{avg}, \tag{3.2.2}$$

where the pressure volume term on the right-hand side is described in Section 2.3.5 of Chapter 2 (but is usually negligible), g_{dehydr} is a dehydration free energy associated with partially removing hydrophobic atomic groups from direct contact with water through incorporation of these groups into the micelle core, and g_{hydr} is a hydration free energy associated with the difference in the structuring of water about the relatively large micelle core relative to structuring about each micelle constituent in monomeric form (note that both the micelle and monomer states are considered at infinite dilution). All the terms in Eq. 3.2.1 are on a per-surfactant molecule basis.

The dehydration and hydration free energies are functions of the size of the simulated aggregate, and, consequently, I will use a hat symbol hereafter to denote a simulation-specific free energy quantity (such that g_{dehydr} and g_{hydr} will be represented as $\widehat{g}_{dehydr}$ and $\widehat{g}_{hydr}$). In the discussion which follows I focus on a single-component surfactant system. (Extensions to multicomponent systems have been proposed based on a composition-weighting scheme, but these are outside the scope of this Thesis.)

Let a index the atomic groups within a surfactant (e.g., CH_2, CH_3, O, OH). Such groups typically involve a heavy atom and any directly attached hydrogens. These groups are selected based on the specific types available in the solubility model of interest (see Section F.5 of Appendix F for a thorough discussion of several such models).

Stephenson and coworkers sought to quantify the change in hydration state of

individual groups a with a so-called fractional degree of hydration[108]

$$f_a \equiv \frac{N_{hydr,mic}(a)}{N_{hydr,mon}(a)}, \tag{3.2.3}$$

where $N_{hydr,mic}(a)$ is a function that quantifies the number of hydrating contacts in the micelle environment based on the selected group a, and $N_{hydr,mon}(a)$ is a function that quantifies the number of hydrating contacts in the monomer environment (also based on the selected group a). The term hydrating contact refers to the close proximity (to atomic group a) of a group capable of participating in the hydrogen-bond network of water (including water itself, ethers, ketones, carboxylic acids, amines, and similar groups). The notion of close proximity is made concrete by the selection of a cutoff radius for determining contacts as a function of time (both $N_{hydr,mic}(a)$ and $N_{hydr,mon}(a)$ are time-averaged quantities) relative to the center of mass of group a. A suitable cutoff radius can be determined systematically by examining the location of the first hydration shell about each group of interest. In practice, an average distance, in the range of 0.45 to 0.55 nm is suitable for most groups (see Section 3.5 below for an example). This procedure is very closely related to that used to determine hydration numbers for ions.[183]

By design, the quantity f_a in Eq. 3.2.3 is zero if a group becomes completely dehydrated in the micellar state, and unity if the hydration environment remains unchanged (in terms of total number of contacts, but perhaps not the energetics of these contacts, as will be seen shortly). In practice, values of f_a will always be found between these two extremes, because bringing surfactants together to form a micelle will eliminate some contacts from even the head groups. Sample values of f_a are provided in Table 3.1 for a sodium dodecyl sulfate (SDS) micelle, where the sulfate group exhibits f_a values ranging from 0.88 for the three outer oxygens (carrying most of the charge of the surfactant) to 0.74 for the "ether" oxygen connecting the sulfur atom to the alkyl tail. Values for each methylene and methyl group in the tail are reported, proceeding down the tail in order (see Figure 3-5 inset for the chemical structure of SDS). There is a rapid fall-off of f_a values within the first few CH_2 groups

Table 3.1: Values of f_a for Individual Groups in Sodium Dodecyl Sulfate (SDS), Calculated Based on Simulation Results for a Spherical Micelle of Aggregation Number 48 at $25°C$ with a Cutoff Radius of 0.55 nm

Group	f_a	Group	f_a	Group	f_a
3O	0.88	CH_2 (C3)	0.40	CH_2 (C8)	0.15
S	0.87	CH_2 (C4)	0.29	CH_2 (C9)	0.16
O	0.74	CH_2 (C5)	0.25	CH_2 (C10)	0.15
CH_2 (C1)	0.60	CH_2 (C6)	0.19	CH_2 (C11)	0.17
CH_2 (C2)	0.47	CH_2 (C7)	0.18	CH_3 (C12)	0.18

of the alkyl region, indicating indirectly that the micelle core is quite dry. Note that preparation of Table 3.1 also involved conducting a simulation of the monomer state, such that f_a could be evaluated according to Eq. 3.2.3.

If f_a quantifies the extent to which a group remains hydrated when transitioning from the monomer to the micelle environment, then the quantity $(1 - f_a)$ can be viewed as the fractional degree of dehydration. This quantity is used in developing an expression for $\widehat{g}_{dehydr}$, as follows:[108]

$$\widehat{g}_{dehydr} = g_{tr,0} + \sum_a (1 - f_a)\, g_{tr,a}, \tag{3.2.4}$$

where $g_{tr,0}$ is any constant term used in the solubility model of interest (according to Section F.5 of Appendix F), and $g_{tr,a}$ is the group contribution of group a to the transfer free energy, according to the solubility model. Note that the sum in Eq. 3.2.4 is over all groups in the surfactant. For hydrophilic groups, it is even possible that $g_{tr,a}$ is a positive quantity, as is the case for ether oxygens, for example, in the solubility model of Klopman, Wang, and Balthasar.[5] However, barring a complete set of values for hydrophilic groups, a lower bound for $\widehat{g}_{dehydr}$ can be determined by setting $g_{tr,a} = 0$ for hydrophilic groups. This is indeed what Stephenson and coworkers recommend.[108]

Now, assume that the surfactant has been partitioned into groups associated with the head (hydrated region) and tail (dehydrated region) of the surfactant using the strategy discussed later in Section 3.3. I define a quantity ζ_a for convenience that takes a value of 0 when a is a head group and a value of 1 when a is a tail group.

The hydration free energy is then expressed as follows:[108]

$$\widehat{g}_{hydr} = P_{ext} v_{avg} + \sum_a \zeta_a f_a \left(SASA_a \Delta\widehat{\mu}_{wc,a} \right), \tag{3.2.5}$$

where $SASA_a$ is the solvent accessible surface area, originally defined by Lee and Richards,[184] of group a (calculated in a simulation of the monomer at infinite dilution), $\Delta\mu_{wc,a}$ is a change in free energy, on a per-$SASA$ basis, due to the size-dependent nature of the hydrogen-bonding network,[185] and the change in size encountered when transitioning from the monomer (small) to the micelle (large) state. (I use the generic symbol μ here, since this term is a combination of Gibbs and Helmholtz terms, with the pressure-volume term, $P_{ext} v_{avg}$, included to convert the Helmholtz term into a Gibbs term for consistency.) Note that Stephenson and coworkers present Eq. 3.2.5 as a sum over groups found in the core.[108] The use of ζ_a makes the two sums equivalent, while a full sum over a is more convenient for the discussion in Section 3.2.3. The term $\Delta\mu_{wc,a}$ can be related to an effective interfacial free energy per unit area for group a in the micelle and monomer environments. Stephenson and coworkers propose the following model:[108]

$$\Delta\widehat{\mu}_{wc,a} = \frac{\widehat{\underline{F}}_{int}}{SASA_{core}} - \frac{-g_{tr,a}}{SASA_a}, \tag{3.2.6}$$

where $\widehat{\underline{F}}_{int}$ is the total interfacial free energy of an idealized micelle possessing the same aggregation number as the simulated aggregate, and $-g_{tr,a}$ represents a conceptual interfacial free energy (per surfactant molecule) for group a in the monomer state. The determination of $\widehat{\underline{F}}_{int}$ can be made according to the MT theory:[81]

$$\widehat{\underline{F}}_{int} = \sigma \left(\underline{A} - \underline{A}_0 \right) \tag{3.2.7}$$

where σ is the curvature-corrected interfacial tension of the micelle core-water interface, $\underline{A}$ is the total area of the interface, and $\underline{A}_0$ is the amount of area shielded by the head-tail bond. (See Section 2.3.6 of Chapter 2.)

By substituting Eq. 3.2.7 into Eq. 3.2.6, and the result into Eq. 3.2.5, one obtains

$$\widehat{g}_{hydr} = \left(\sum_a \zeta_a f_a g_{tr,a} \right) + \left(P_{ext} \upsilon_{avg} + \sum_a \zeta_a f_a \sigma \left(\underline{A} - \underline{A}_0 \right) \frac{SASA_a}{SASA_{core}} \right), \quad (3.2.8)$$

where I have bracketed the two terms which can be considered Gibbs free-energy contributions per surfactant molecule.

3.2.2 Calculation of a Modified Transfer Free Energy

The calculation of $\widehat{g}_{dehydr}$ according to Eq. 3.2.4 and $\widehat{g}_{hydr}$ according to Eq. 3.2.8 is readily carried out through micelle and monomer simulations for a surfactant of interest. However, as the hats indicate, these are simulation-dependent quantities. That is, the size and shape of the micelles considered must be taken into account. Technically, the substitution in Eq. 3.2.1 within the MT framework would lead to a requirement of a different simulated micelle for each possible micelle in solution, to enable proper evaluation of g_{mic} during the optimization procedure described in Section 2.4 of Chapter 2. Although there is a finite number of spheres possible for any particular surfactant (with size limited by the dimensions of the surfactant chain), there is no limit on the number of finite spherocylinders or discs that would need to be modeled.

To circumvent this untenable situation, consider the following definition:

$$\widehat{g}_{tr,CSMT} = \widehat{g}_{dehydr} + \widehat{g}_{hydr} - \widehat{g}_{int}, \quad (3.2.9)$$

which is based on a rearrangement of the replacement relationship in Eq. 3.2.1, where $\widehat{g}_{tr,CSMT}$ can be considered analogous to g_{tr} in the original MT approach. Now, in the thought process of the MT model, g_{tr} represents the free-energy benefit associated with transfering a surfactant tail from solution into a bulk phase of tails. On a per-surfactant molecule basis, the transfer free energy does not depend on micelle size or shape, unlike the interfacial term and the dehydration and hydration terms, which depend on f_a, a size and shape-dependent quantity. If $\widehat{g}_{tr,CSMT}$ is also independent

of size and shape, due to convenient cancellation of these dependencies on the right-hand side of Eq. 3.2.9, then one can write $\widehat{g}_{tr,CSMT} = g_{tr,CSMT}$. As a result, a single micelle and monomer simulation could be used to calculate $\widehat{g}_{tr,CSMT}$, and consequently $g_{tr,CSMT}$, a term which can be used to calculate g_{mic} for any possible micelle. By making a replacement of g_{tr} with $g_{tr,CSMT}$, no further simulations would be required to proceed with optimization of g_{mic} for the prediction of optimal micelle properties and micellization behavior (e.g., critical micelle concentrations).

It is certainly not clear a priori that $\widehat{g}_{tr,CSMT} = g_{tr,CSMT}$ should hold, other than as an analogy to g_{tr}. However, in my own results (discussed in Chapters 5 and 6), the size-dependence of $\widehat{g}_{tr,CSMT}$ has been found to be minimal. Furthermore, Stephenson and coworkers found, for a variety of oil droplet sizes and shapes that $\widehat{g}_{tr,CSMT}$ is also reasonably shape-independent.[108]

3.2.3 Comparing the CSMT and MT Transfer Free Energies

In Section 3.2.2, an analogy between g_{tr} and $g_{tr,CSMT}$ was made. It is instructive to examine the difference between these two quantities. Starting with Eq. 3.2.9 and substituting in Eqs. 3.2.2, 3.2.4, and 3.2.8, one obtains

$$g_{tr,CSMT} = g_{tr,0} + \sum_a (1 - f_a + \zeta_a f_a)\, g_{tr,a} + \widehat{f}_{int} \left[-1 + \sum_a \zeta_a f_a n \frac{SASA_a}{SASA_{core}} \right], \tag{3.2.10}$$

where the pressure volume terms have cancelled out from the $\widehat{g}_{hydr}$ and $\widehat{g}_{int}$ terms, and the relationship $\widehat{F}_{int} = n\widehat{f}_{int}$ has been used, with n the simulated micelle aggregation number.

In Eq. 3.2.10, the summation term involving $g_{tr,a}$ can be split into two summations involving head ($\zeta_a = 0$) and tail ($\zeta_a = 1$) groups separately:

$$\sum_a (1 - f_a + \zeta_a f_a)\, g_{tr,a} = \sum_a (1 - \zeta_a)\,(1 - f_a)\, g_{tr,a} + \sum_a \zeta_a g_{tr,a}. \tag{3.2.11}$$

Noting that g_{tr} in the traditional MT approach (indicated hereafter as $g_{tr,MT}$) includes only contributions from atomic groups in the surfactant tail (i.e., $g_{tr,MT} =$

$g_{tr,0} + \sum_a \zeta_a g_{tr,a}$), inserting Eq. 3.2.11 into Eq. 3.2.10 and rearranging yields

$$g_{tr,CSMT} - g_{tr,MT} = \sum_a (1 - \zeta_a)(1 - f_a) g_{tr,a} + \widehat{f}_{int} \left[-1 + \sum_a \zeta_a f_a \frac{SASA_a}{SASA_{core}/n} \right]. \tag{3.2.12}$$

The first term on the right-hand side of Eq. 3.2.12 is the contribution to $g_{tr,CSMT}$ due to groups in the head of the surfactant, for which $\zeta_a = 0$. As already mentioned, $g_{tr,a}$ is typically considered non-zero only for hydrophobic groups (i.e., groups having $g_{tr,a} < 0$). In the case of oils, there are no head groups, indicating that the only difference between $g_{tr,CSMT}$ and $g_{tr,MT}$ for oils would arise from the second term on the right-hand side, and, specifically, from how $\sum_a \zeta_a f_a \frac{SASA_a}{SASA_{core}/n}$ differs from unity.

Recall that f_a, as defined in Eq. 3.2.3, is a ratio of hydrating contacts of group a in the micelle and monomer states. It is therefore reasonable to write

$$f_a = \frac{K_{a,micelle} SASA_{a,micelle}}{K_{a,monomer} SASA_{a,monomer}}, \tag{3.2.13}$$

where the K_a's are proportionality constants in the micelle and monomer environments, respectively, which relate the number of contacts in each environment to the corresponding $SASA$ values for group a. Note that $SASA_{a,monomer}$ is simply a synonym for $SASA_a$ in earlier equations. Accordingly,

$$\sum_a \zeta_a f_a \frac{SASA_a}{SASA_{core}/n} = \sum_a \zeta_a \frac{K_{a,micelle}}{K_{a,monomer}} \frac{SASA_{a,micelle}}{SASA_{core}/n}. \tag{3.2.14}$$

Since the tail groups are the only groups that contribute to $SASA_{core}$, it follows that

$$SASA_{core} = n \sum_a \zeta_a SASA_{a,micelle}. \tag{3.2.15}$$

Equation 3.2.15 shows that, if $K_{a,micelle} = K_{a,monomer}$ for all groups a in Eq. 3.2.14, it would follow that $\sum_a \zeta_a f_a \frac{SASA_a}{SASA_{core}/n} = 1$, and the second term on the right-hand side of Eq. 3.2.12 vanishes. Stephenson and coworkers were able to demonstrate effective (but not complete) agreement between $g_{tr,CSMT}$ and $g_{tr,MT}$ for oil droplets,[108] which seems to suggest that the difference between $K_{a,micelle}$ and $K_{a,monomer}$ is non-

zero but quite small in the absence of head groups. That is, for oil droplets, the number of hydrating contacts is proportional to the exposed area with approximately the same proportionality constant in both the oil droplet and single oil molecule states, despite the restructuring of the hydrogen-bonding network that occurs about larger hydrophobic entities. In contrast, for surfactant systems, which do involve head groups, an increased deviation between $K_{a,micelle}$ and $K_{a,monomer}$ can be expected to occur, due to the physical presence of the surfactant head groups in the simulation. These head groups do not affect calculations of $SASA_{core}$ by definition (see Section 3.5.4 – $SASA_{a,micelle}$ is also unaffected for each tail group a), but do generally reduce the number of hydrating contacts that occur, reducing $K_{a,micelle}$ relative to $K_{a,monomer}$, and leading to deviation of $\sum_a \zeta_a f_a \frac{SASA_a}{SASA_{core}/n}$ from unity.

3.3 Surfactant Head/Tail Identification Using a Dividing Surface Approach

The identification of surfactant atomic groups as belonging to the head or tail region of the surfactant is essential to carry out MT predictions, since all aspects of the thought process presented in Section 2.3.1 of Chapter 2 are predicated on a division of the surfactant into these two regions for the purposes of calculating the free-energy contributions corresponding to the assembly of a micelle from its dispersed constituents. Head/tail identification is also necessary for the evaluation of $g_{tr,CSMT}$ to make predictions using the CSMT framework, since $\widehat{g}_{hydr}$ includes contributions which arise solely from tail groups (see Eq. 3.2.5).

In the second paper in the three-part series on quantifying the hydrophobic effect by Stephenson and coworkers,[109] a method is proposed to carry out the required head/tail identification. It is suggested to examine the fractional degrees of hydration of individual groups, f_a, relative to a cutoff which determines the identification. Specifically, group a is considered to be a tail group if f_a is less than the cutoff. For this purpose, a cutoff of 0.6 is proposed, motivated by the study of nonionic[109]

and ionic and zwitterionic[110] surfactants, considerations of the curvature of the core-water interface,[109] and recognition that the fractional degrees of hydration of terminal methyl groups in the oils considered[108] often had values of $f_a = 0.53$ (a method which identified oils as possessing head groups would clearly be in error).

However, I have identified a problem with this approach: calculation of fractional degrees of hydration involves consideration of both micellar and monomer states. However, intuitively, a head/tail identification for a surfactant molecule in a micelle environment is not expected to depend on the monomer state. Indeed, the head/tail identification is principally meant to define the micelle core for the purpose of calculating the various contributions to the free energy of micellization.[81] While this identification is certainly a function of the absolute hydration state of each group (the micelle core is expected to be quite dry, the head region is expected to be well-hydrated, and intermediate groups are expected to exhibit some intermediate extent of hydration), f_a values are measures of relative hydration states (between micelle and monomer environments). In fact, a group in a surfactant molecule could be quite shielded from contact with water in the monomer environment (e.g., a tetra-functionalized carbon surrounded by bulky groups which prevent contact of the carbon with water), such that transfer of such a group to the micelle core should not greatly change its hydration state. In that case, although this group is embedded in the micelle core, f_a could still be greater than 0.6, leading to the incorrect identification that the group in question belongs to the head region. This could then result in a discontinuous head/tail identification, where one group is identified as head based on its f_a value, but surrounding groups are identified as tail.

In examining the various plots of f_a for nonionic surfactants in the results of Stephenson and coworkers,[109] it becomes evident that these situations do occur, even for common, linear surfactants. For example, in Figure 8 of Ref. [109], which depicts the f_a values for an ethoxylated surfactant, the first ether oxygen attached to the alkyl tail is the first group encountered (starting from the terminal methyl) which has an f_a value greater than 0.6. However, the two CH_2 groups between the first ether and the second ether oxygen clearly have f_a values less than 0.6 (all other remaining

groups have $f_a > 0.6$ and would be considered head by this criterion). This is clearly a discontinuous identification, and presents difficulties in subsequent modeling. It appears that Stephenson and coworkers mainly focused on the traditional head/tail identifications used in MT modeling for the purposes of making predictions using the CSMT framework and comparing to the MT results. The head/tail identifications described elsewhere in Ref. [109] are the traditional ones, and do not match in general with the 0.6 cutoff approach. As a result, attention was not brought to the important issues that I am raising here.

A key contribution that I have made in this Thesis is to address the issue of proper head/tail identification. To accomplish this, I conceived of a dividing surface approach, which involves consideration of surfactants in the micelle state only, and does not depend on the counting of water contacts or on the determination of f_a values. This work was partially developed with the assistance of my colleague Shangchao Lin in the early stages of development. I also designed and implemented a self-consistent iterative scheme for determining the location of the dividing surface, which I will describe shortly. With respect to the median location of atomic groups relative to the micelle core center of mass, the head/tail identification is necessarily continuous in this approach.

The notion of using a dividing surface was motivated by the thermodynamic models within the MT framework itself. In Chapter 2, I described how the micelle core-water interface is conceptualized as being a sharp boundary between the aggregated surfactant tail groups and the bulk water environment (in which the surfactant heads and ions also reside). For example, in the evaluation of the packing free-energy contribution within the MT theory, there is a strict assumption that surfactant tails cannot leave the core, and that surfactant head groups and water cannot enter the core. This idealized portrayal of the micelle core has a counterpart in bulk-phase thermodynamics, where two immiscible phases are considered to contact each other at a planar interface of zero width. The location of this interface for modeling purposes is typically chosen based on the determination of a Gibbs dividing surface between the two phases.[142] The Gibbs dividing surface is an equimolar dividing surface selected

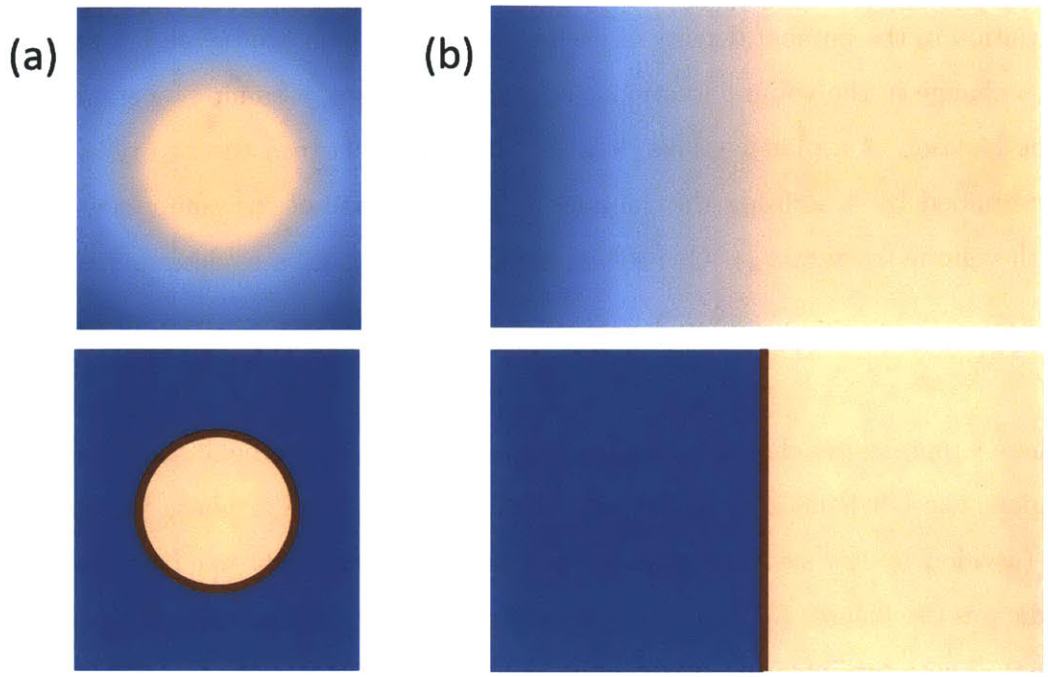


Figure 3-1: Analogy between the micelle core-water interface and a Gibbs dividing surface between two immiscible macroscopic phases. In (a), at top, a real micelle core (light orange) is pictured, which exhibits a diffuse interfacial region, with some water (blue) penetration and surface roughness. At bottom, an idealized micelle core is pictured, with a sharp boundary (dark brown) separating the core and water region (surfactant heads, not shown, are assumed to remain in the water region). In (b), at top, the interfacial region separating two bulk immiscible phases (e.g., oil (light orange) and water (blue)) is pictured. At bottom, an idealized division of the system into pure phases separated by a Gibbs dividing surface (dark brown) is portrayed.

such that the surface excess of one of the components of the system (comprising one of the pure phases) is zero. Figure 3-1 illustrates the analogy between the micelle core-water system and the interface resulting from contact between two immiscible phases. In both cases, the inhomogeneities take place over a molecular length scale in an interphase region.[142] Note that the dividing surface in the case of the micelle core will adapt the geometry of the micelle itself, whether spherical, cylindrical, or planar (bilayer).

For the planar case in Fig. 3-1, the location of the dividing surface is determined as follows. Assume that species A and B constitute the two immiscible macroscopic

phases in equilibrium, respectively. It is assumed that there is a region of rapid variation in the number density of each component, which can be readily represented as a change in the volume fraction ϕ of each component, assuming incompressibility. The location of a planar Gibbs dividing surface, d, between the two phases can be determined by examining the volume fraction ϕ of one of the components (say B, with volume fraction ϕ_B). One solves the following equation to determine d:

$$\int_{-\infty}^{d} \phi_B(r)\, dr = \int_{d}^{\infty} [1 - \phi_B(r)]\, dr, \tag{3.3.1}$$

where r represents a distance coordinate along an axis perpendicular to the interfacial region, the left-hand side as a whole indicates the volume in phase A occupied by B (divided by the system cross-sectional area), and the right-hand side as a whole indicates the volume in phase B occupied by A (divided by the system cross-sectional area). These two integrals are depicted as light red, shaded regions in Fig. 3-2, which is a depiction of a generic volume fraction profile for ϕ_B in the vicinity of the interface. When d is selected such that Eq. 3.3.1 is satisfied, the region corresponding to $r < d$ can be modeled as pure A and the region corresponding to $r > d$ can be modeled as pure B. The total volume of B in solution (equal to the value of the integral $\int_{-\infty}^{\infty} \phi_B(r)\, dr$ multiplied by the system cross-sectional area) will then be found to be equal to the volume of the region corresponding to $r > d$ (again, assuming fixed molecular volumes and incompressibility).

The dividing surface for a micelle can be found similiarly to what was done using Eq. 3.3.1. As discussed in Section 3.5, from molecular dynamics simulations of micelles, one has access to information on the average location of a collection of atomic groups relative to others through the concept of a radial distribution function (rdf). The following equation can be used to find d in this case:

$$\int_{V(r<d)} g(r)\, dV = \int_{V(r>d)} [\widetilde{g} - g(r)]\, dV \tag{3.3.2}$$

where $g(r)$ is the radial distribution function for the bulk phase relative to the center of mass of the micelle core (the composition of both of these regions will be discussed

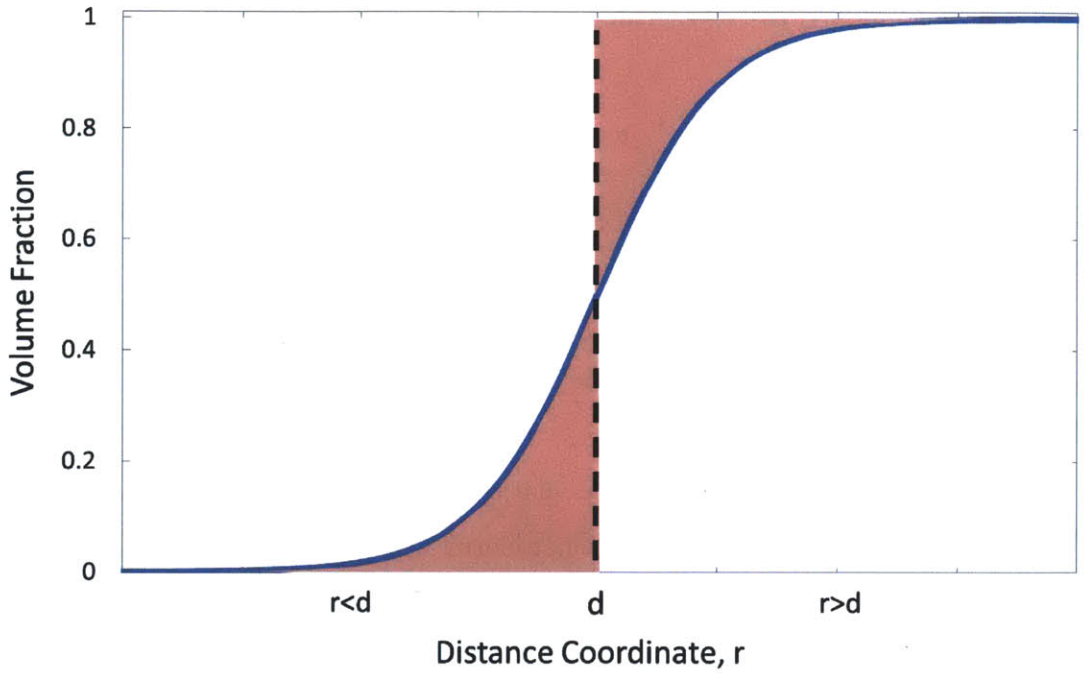


Figure 3-2: Illustration of the Gibbs dividing surface, as applied to two immiscible phases in equilibrium. The location of the dividing surface (at position d) is determined according to Eq. 3.3.1 in the text. A representative window for the distance coordinate, r, in the vicinity of the interface is shown, but the full domain is $(-\infty, +\infty)$. The two areas shaded in light red are equal.

shortly), $V(r < d)$ indicates the volume of the region characterized by the radial coordinate (spherical radius, cylindrical cross-sectional radius, or bilayer half-width) being less than the location of the dividing surface (i.e., the core volume), $V(r > d)$ indicates the volume of the bulk region, and $\widetilde{g}$ is the bulk value obtained for $g(r)$ in the simulation cell when evaluated far from the micelle core. Note that $\widetilde{g}$ approaches unity from above as the simulation cell size is increased for a fixed micelle core. By evaluating dV for the three geometries considered (sphere, cylinder, bilayer), Eq. 3.3.2 can be simplified to

$$\int_0^d g(r)\, r^{S-1} dr = \int_d^D [\widetilde{g} - g(r)]\, r^{S-1} dr, \tag{3.3.3}$$

where S is a shape factor ($S = 1$ for bilayers, 2 for cylinders, and 3 for spheres), the left-hand side lower integral bound corresponds to the center of mass of the micelle (centered at the origin), and D is a practical upper limit for integration, imposed by the periodic boundary conditions of the simulation cell (D is equal to the half-width of the simulation cell box).

In order to carry out practical calculations involving Eq. 3.3.3, one needs to identify the groups found in the bulk (i.e., identify which groups in the simulation cell are not part of the micelle core). For this purpose, I have proposed the following self-consistent iterative process. One begins by examining $g(r)$ as determined for water and ions relative to the center of mass of the micelle (specifically, the center of mass of the aggregated surfactants, excluding bound counterions, which are themselves included in the groups contributing to $g(r)$). Once d has been determined using Eq. 3.3.3 (the actual implementation involving histograms and discrete sums rather than integrals, according to the output of GROMACS), head groups are identified based on their median positions relative to the dividing surface location (median positions greater than d lead to a head identification). The next step of the iteration involves including these head groups with the water and ions, generating a new $g(r)$ profile, and repeating until no new head groups are identified. The final head/tail identification is that obtained at convergence.

Figures 3-3 and 3-4 below illustrate the effect of the iterative scheme on the determination of the location of the dividing surface, as applied to a sodium dodecyl sulfate (SDS) and octaethylene glycol monododecyl ether ($C_{12}E_8$) micelle, respectively. (Note that both figures show only the initial and final iterations, for clarity.) It is clear that the inclusion of head groups in the rdf determination will always shift the dividing surface location to the left, corresponding to a smaller core identification. However, the magnitude of the shift is a strong function of the type of surfactant under consideration.

In the case of SDS (see Figure 3-3), only three iterations are required to obtain a stable solution (the first identifies the sulfate as a head group, the second identifies an additional methylene group as belonging to the head, and the third confirms the head/tail identification). The dividing surface location shifts from 1.686 nm (blue dashed line) to 1.569 nm (orange dashed line), a difference of 0.117 nm, and a change in core volume of approximately 3900 Å^3. With SDS identified as possessing a C_{11} tail, with volume 334 Å^3 (see Appendix F for information about molecular volume determination), and using the dividing surface location of 1.569 nm as the radius of a sphere, the calculated aggregation number (the sphere volume divided by the molecular tail volume) is 48.4, in good agreement with the value of 48 selected in the simulation. The distribution of individual atomic groups relative to the final dividing surface is captured in a seven-number box plot format in Figure 3-5, where the location of the red median lines above or below the orange dividing surface line indicates the groups that are identified as head or tail, respectively. The additional percentiles in the box plot are provided to give a sense of the distribution spread. For example, it is interesting to note that the terminal methyl group has a large spread, and samples the interfacial region about 10-20% of the time at equilibrium (recall that these results are time averages). This recalls the work of Menger and Doll,[186] who demonstrated that terminal methyl groups sample the interface, using olefin oxidation experiments.

In the case of $C_{12}E_8$ (see Figure 3-4), a total of 17 iterations is necessary to achieve convergence, due to the presence of the long ethoxylated head. The first iter-

ation (with only water considered in preparing the rdf) yields a head identification of just $CH_2OCH_2CH_2OH$. By the 17th iteration, the head is identified as comprising $CH_2(OCH_2CH_2)_8OH$ (i.e., a C_{11} tail identification). The shift in the dividing surface location is substantial: from 2.162 nm (blue dashed line) to 1.538 nm (orange dashed line), a difference of 0.624 nm, and a change in volume of approximately 27000 $Å^3$. The calculated aggregation number, 45.6, is slightly smaller than the simulated number of 48. The smaller core (relative to the SDS case) is due to the slight peak in the final rdf in Figure 3-4, which is due to the physical attachment of the head groups to the tails (such that they cannot freely mix in the bulk solution), and their difference in density compared to water groups (this is a comparison of CH_2, O, and OH groups with H_2O groups, since the rdfs are for groups, not molecules). When a peak is present, the determination of $\widetilde{g}$ is typically accomplished by taking an average of the values obtained beyond the peak. Since $\widetilde{g}$ will be less than the peak value, the peak introduces a slight negative contribution to the integral on the right-hand side of Eq. 3.3.3, which is offset with a shift of the dividing surface to the left, as observed. Note that a peak is not observed with SDS, because SDS does not possess a large head. Since the difference in aggregation number here is slight (less than 3 molecules), and $C_{12}E_8$ represents a fairly large head for a surfactant, my results suggest that the new technique for head/tail identification introduced here has the potential for broad applicability. The varied systems in Chapters 5 and 6 will provide an additional test. Finally, Figure 3-6 demonstrates the distribution of individual atomic group relative to the final dividing surface obtained for $C_{12}E_8$. As with SDS above, the location of the red median lines above or below the orange dividing surface line indicates the groups that are identified as head or tail, respectively. The additional percentiles are again provided to give a sense of spread.

Finally, it is worth notingwhile to note that the identifications for SDS and $C_{12}E_8$ determined here agree with the rule of thumb used in the original MT modeling of Puvvada and Blankschtein,[81] based on the recommendation by Tanford,[34] that the first methylene group be considered quite hydrated due to its adjacency to oxygen groups capable of hydrogen bonding.

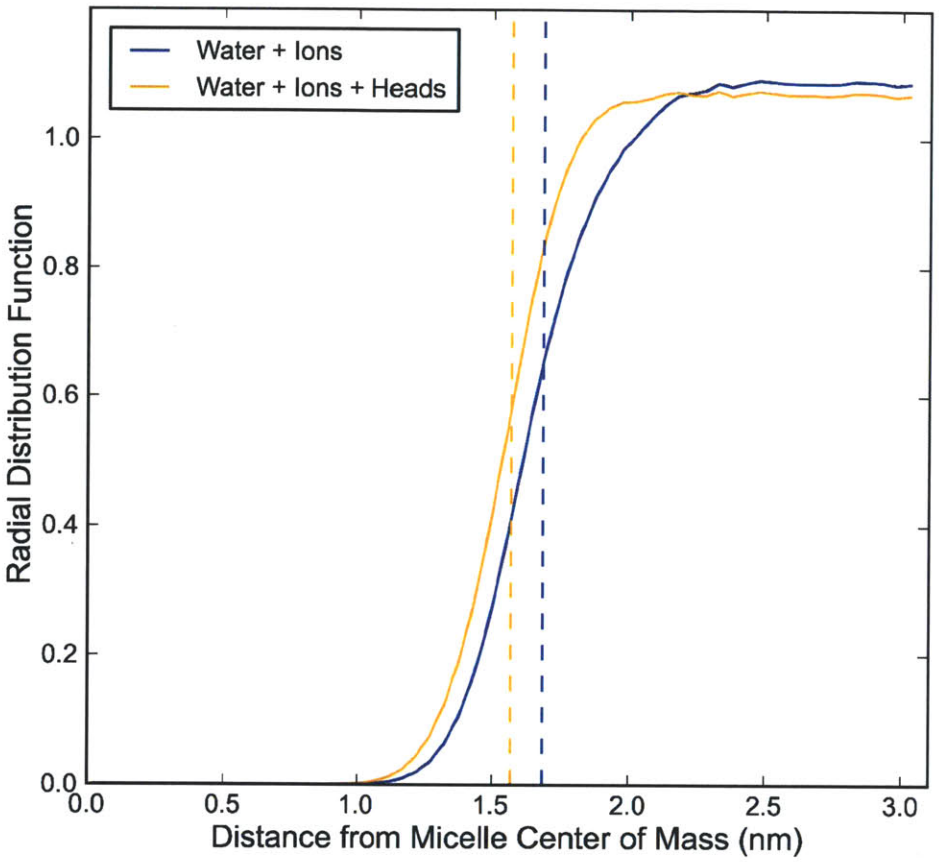


Figure 3-3: The dividing surface approach applied to a spherical sodium dodecyl sulfate (SDS) micelle of aggregation number 48 at $T = 298.15K$. The blue curve represents the radial distribution function (rdf) for water and ions relative to the center of mass of the micelle. The vertical, blue dashed line is the location of the dividing surface for this initial iteration. The orange curve represents the rdf for water, ions, and head groups, identified through an iterative process. The vertical, orange dashed line is the location of the final dividing surface (i.e., the location of the micelle core-water interface for the purposes of modeling).

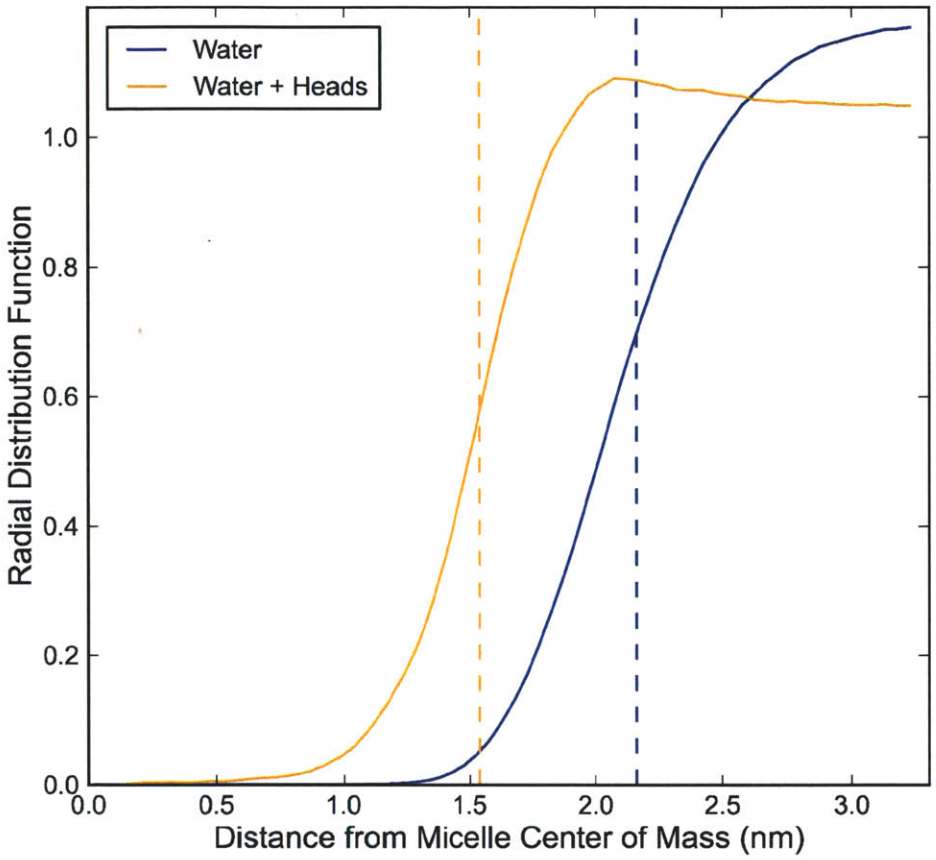


Figure 3-4: The dividing surface approach applied to a spherical octaethylene glycol monododecyl ether ($C_{12}E_8$) micelle of aggregation number 48 at $T = 298.15K$. The blue curve represents the radial distribution function (rdf) for water relative to the center of mass of the micelle. The vertical, blue dashed line is the location of the dividing surface for this initial iteration. The orange curve represents the rdf for water and head groups, identified through an iterative process. The vertical, orange dashed line is the location of the final dividing surface (i.e., the location of the micelle core-water interface for the purposes of modeling).

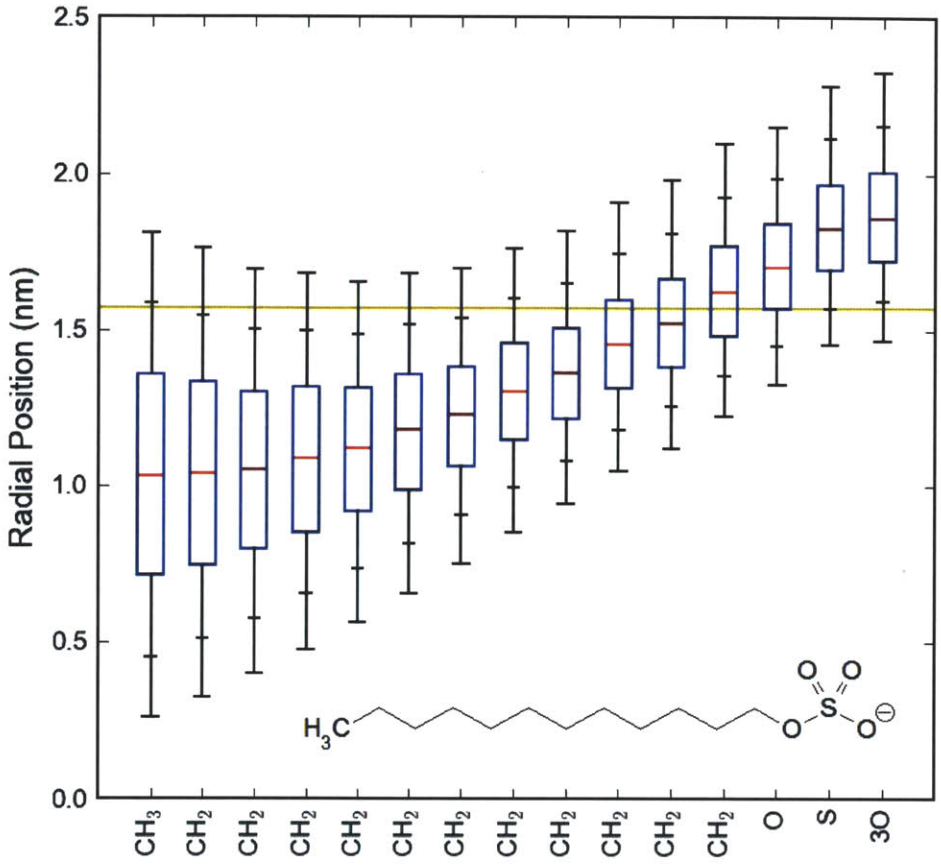


Figure 3-5: Average location of individual atomic groups within sodium dodecyl sulfate (SDS) relative to the dividing surface derived from water, ions, and heads in Figure 3-3 (orange line). Each box with associated lines represents a seven-number summary of the various percentiles of the probability distribution for a group: the red line is the median; the lower and upper bound of the blue box is the first and third quartile, respectively; the black lines extend down to the 2nd percentile and up to the 98th percentile, and the intermediate hatches are at the 9th and 91st percentiles. The head/tail identification is performed by assessing the location of each median relative to the dividing surface, as described in the text. The remaining percentiles in the box plot are provided to give a visual indication of the spread in the location of each group, with equidistant lines indicating an approximately normal distribution.

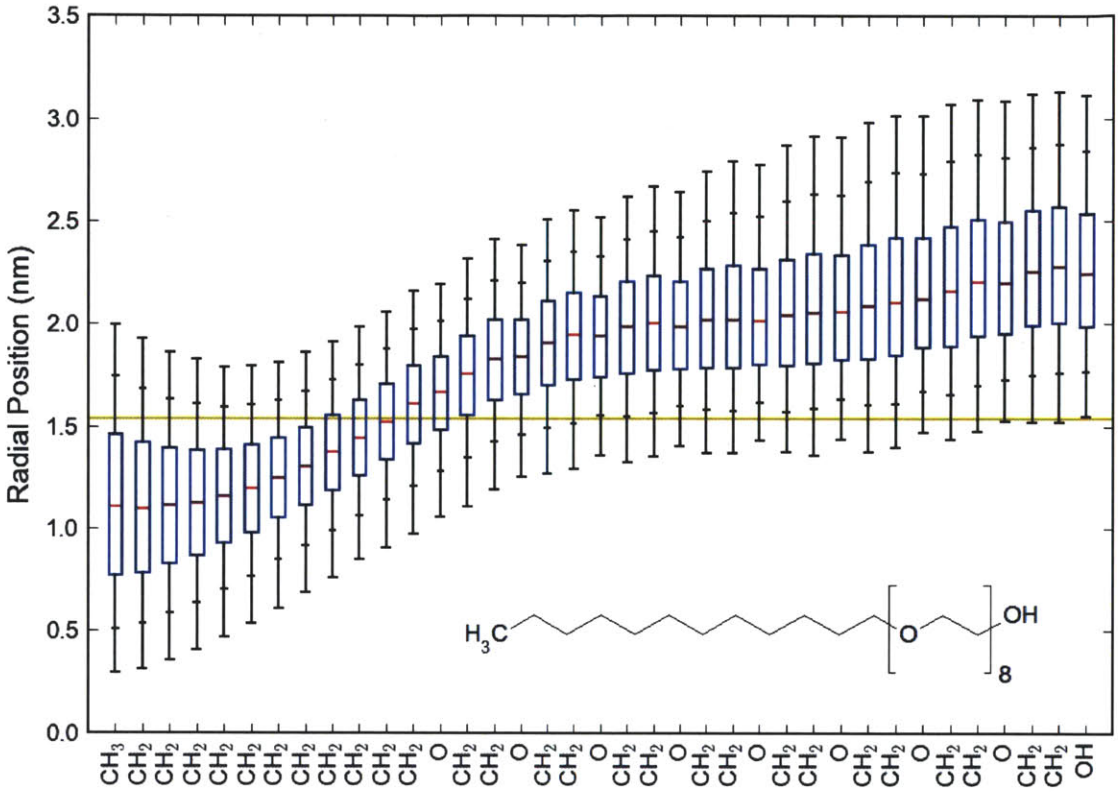


Figure 3-6: Average location of individual atomic groups within octaethylene glycol monododecyl ether ($C_{12}E_8$) relative to the dividing surface derived from water and heads in Figure 3-4 (orange line). Each box with associated lines represents a seven-number summary of the various percentiles of the probability distribution for a group: the red line is the median; the lower and upper bound of the blue box is the first and third quartile, respectively; the black lines extend down to the 2nd percentile and up to the 98th percentile, and the intermediate hatches are at the 9th and 91st percentiles. The head/tail identification is performed by assessing the location of each median relative to the dividing surface, as described in the text. The remaining percentiles in the box plot are provided to give a visual indication of the spread in the location of each group, with equidistant lines indicating an approximately normal distribution.

3.4 Molecular Dynamics Simulations Protocol for the CSMT Framework

As described in Section 3.2, modeling the micellization behavior of a particular surfactant using the CSMT framework requires two simulations: one of the surfactant in a monomer state, and one of the surfactant in an aggregated state. Since we are interested in the precise arrangement of water, ions, and other participants in the hydrogen-bonding network of water about surfactant groups, explicit solvent simulations are necessary.[108] Furthermore, since electrostatics considerations affect the arrangement of water, it makes sense to use an all-atom treatment of the surfactant, where each atom in the surfactant molecule has an explicit representation within the simulation box (as opposed to a united atom treatment which groups the hydrogens in methylene and methyl groups into a single bead along with their associated carbon[187]). An all-atom simulation has more atom centers which can be assigned a partial charge, thereby improving the fit of the electrostatic potential experienced by a point charge approaching the molecule to that computed using ab initio methods.[188] There is also an improvement in the representation of the torsional degrees of freedom in such simulations, due to the increased number of dihedrals introduced by the presence of the explicit hydrogens, which can be parameterized to again reproduce ab initio results.[188] (A discussion of force field parameterization approaches that I have used in modeling some of the surfactants considered in this Thesis can be found in Appendix E.)

The molecular dynamics (MD) software used by Stephenson and coworkers[108–110] and also used in this Thesis is the GROMACS (GROningen MAchine for Chemical Simulations) high performance molecular dynamics software package originally developed at the University of Groningen and maintained by van der Spoel, Hess, and Lindahl.[189,190] All the analysis presented in Chapters 5 and 6 was performed using GROMACS Version 4.5.1.

To run a simulation in GROMACS, four items at minimum are needed:[190] (i) a topology for each unique molecule in the system, which identifies the force field

to be used in modeling the surfactant, the type of each atom within that force field specification, and the specific connectivity of the atoms in terms of bond, bond angles, and bond dihedrals, (ii) a list of the molecules in the system and their counts, (iii) a file of initial coordinates for the system, and (iv) a file specifying the run parameters to be used, including solution conditions, time steps, and other simulation-related parameters.

I have devoted a substantial effort in the development of an automation framework which allows a user to rapidly assemble these four items to create the micelle and monomer simulations needed for CSMT predictions, based on input of the chemical structure of the surfactant of interest into a software package that I developed called Program BUILDER. A full discussion of this program is found in Appendix C. Here, I will proceed with a brief discussion of the settings used for all the runs in Chapters 5 and 6. Program BUILDER was used in the preparation of all these systems, and the study was limited to spherical micelles (the shape most commonly encountered at the critical micelle concentration for nearly all the surfactant systems studied). Simulation snapshots of the micelle simulations for two representative surfactants are presented in Figure 3-7 to illustrate the all-atom simulations used (water is not shown for clarity, but is present explicitly in the actual simulations).

The settings that I have used are as follows:

• The default integrator is used (a leap-frog algorithm) for integration of Newton's equations of motion through time.[190]

• The OPLS-AA force field is used exclusively,[188] except when custom parametrization is required due to missing parameters. The parameters for certain systems are taken from the literature (e.g., from the ionic liquids work of Canongia Lopes and coworkers[191–194]), and documented in Chapters 5 and 6 accordingly. Others were developed by me using the techniques discussed in Appendix E.

• All bonds are converted into constraints to allow an increase in the time step. Constraints are solved using the LINCS (linear constraint solver) algorithm, allowing the stable use of a 2 fs time step.[195]

• Translations of the center of mass of the simulation cell are removed at each

time step.

• The simulation cell is rectangular (i.e., possesses orthogonal box vectors) and is periodic in all directions.

• Coordinates are written every 10000 steps (20 ps of simulation time). All analyses described in Section 3.5 are based on these coordinates (the simulation trajectory).

• A nearest neighbor list utilizing a grid is used to determine which atoms must be considered for non-bonded interactions. This list is updated every 10 time steps.

• A three-dimensional particle mesh Ewald (PME) summation approach is used to handle the long-range Coulombic interactions,[196,197] which permits efficient and accurate computation of the Coulombic forces in a simulation cell with periodic boundary conditions[190] by dividing the Coulombic potential into direct space, reciprocal space, and correction terms, with a fast Fourier transform (FFT) used to compute the reciprocal space term.[197] Small cutoffs can then be used in both spaces without adversely impacting the accuracy of calculation (i.e., a small spatial cutoff in direct space, and a small cutoff for the maximum magnitude of the wave vectors in reciprocal space can be used).[198] I have used the settings recommended in the manual for GROMACS[190] of a direct space cutoff of 0.9 nm, an FFT grid spacing of 0.12 nm, and a cubic interpolation of charges to the FFT grid.

• Van der Waals interactions are modeled using a 12-6 Lennard-Jones potential,[199] and a relatively short cutoff of 0.9 nm with a long range dispersion correction to the energy and pressure,[190] as recommended by Stephenson and coworkers.[108]

• Initial velocities are generated using a Maxwell distribution for the desired temperature (see Section C.3 of Appendix C).

• System temperature is maintained using a velocity-rescaling thermostat which maintains the proper statistics for the canonical ensemble.[200] In this technique, velocities of individual particles are rescaled such that the total kinetic energy is set equal to a stochastic target rather than to a constant value, permitting fluctuations in the kinetic energy. A coupling time constant of 0.5 ps is used.

• System pressure is maintained isotropically using a Berendsen coupling algorithm,[201] an exponential relaxation technique which scales the coordinates and sim-

ulation box vectors at each time step of the simulation, with a time constant of 2.0 ps, a reference pressure of 1 bar, and a compressibility of 4.5×10^{-5} bar^{-1} (a default recommended value corresponding to the isothermal compressibility of water at atmospheric pressure and $T = 300K$[190]).

3.5 Simulation Analysis for the CSMT Framework

3.5.1 Overview

In this section, I present the various concepts and procedures that I have used to analyze the micelle and monomer simulations obtained using the protocols described in Section 3.4. The goal is to obtain from these simulations: (i) the surfactant head/tail identification, (ii) the fractional degrees of hydration, f_a, for each atomic group a in a surfactant of interest, (iii) the solvent accessible surface areas, $SASA_a$, for each group a, and (iv) the solvent accessible surface area for the micelle core, $SASA_{core}$. With knowledge of (i), (ii), (iii), and (iv) above, along with the specific details of the micelle simulation (aggregation number and system temperature and pressure conditions), one can determine the CSMT transfer free energy, $g_{tr,CSMT}$, according to the methodology described in Section 3.2.

Item (i) above is obtained according to the methodology described in Section 3.3. The approach therein requires determination of radial distribution functions of hydrated groups (including water itself) about the center of mass of the micelle simulated. This is described in Section 3.5.2.

Item (ii) above is obtained by counting the number of hydrating contacts that fall within a certain cutoff radius of a group of interest. This cutoff radius is motivated by the location of the first hydration shell about the group of interest, determined through analysis, again, of a radial distribution function. I present this analysis in Section 3.5.3.

Items (iii) and (iv) involve determination of solvent accessible surface areas. I briefly describe this concept in Section 3.5.4.

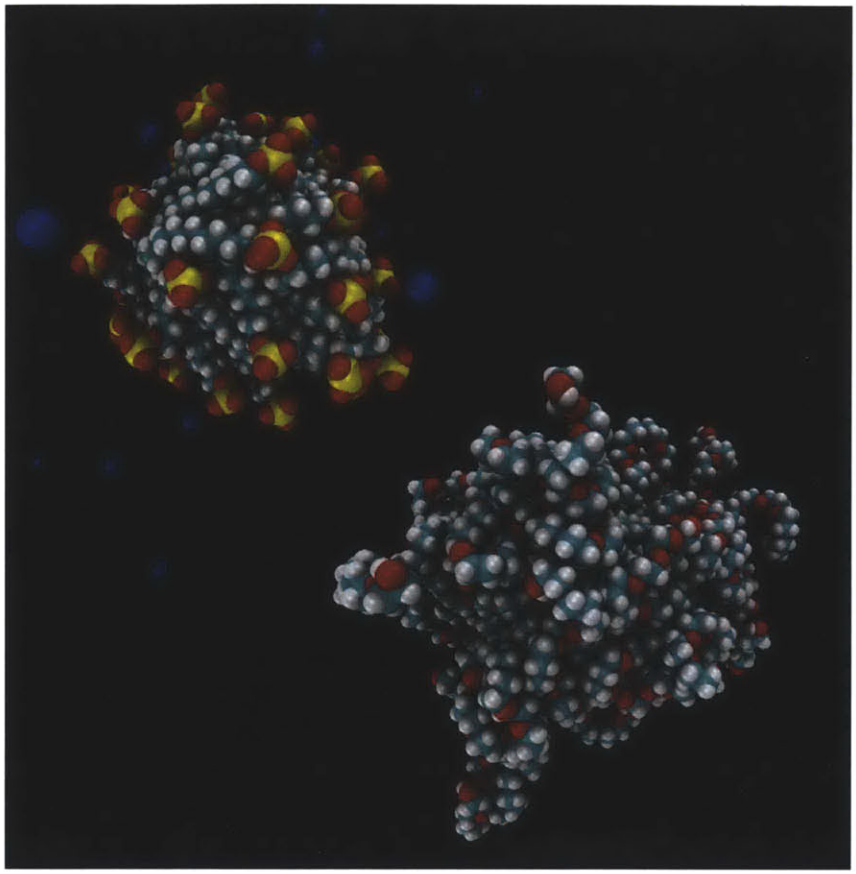


Figure 3-7: Simulation snapshots of equilibrated sodium dodecyl sulfate (SDS, top left) and octaethylene glycol monododecyl ether ($C_{12}E_8$, bottom right) micelles, representing two independent simulations and an aggregation number of 48 used in each. The SDS micelle is quite compact due to the small head groups, with a substantial fraction of the core exposed to contact with water (not shown for clarity). A cloud of counterions surrounds the micelle, with transient direct interactions between individual counterions and the charged sulfate heads. In the $C_{12}E_8$ micelle, the long ethoxylated heads substantially shield the core. Individual surfactants transiently demonstrate extension of the ethoxylated head into the bulk region. However, a majority of the ethoxylated chains remain associated with the outer region of the core. (Color code for atoms: red – oxygen, teal – carbon, white – hydrogen, yellow – sulfur, and dark blue – sodium.)

Finally, in Section 3.5.5, I describe the specific protocol followed in proceeding from completed monomer and micelle simulations through to the calculation of $g_{tr,CSMT}$, in a numbered list format.

3.5.2 Radial Distribution Functions

Radial distribution functions, generally denoted $g_{AB}(r)$, represent the non-uniform distribution of atoms of type B in space about centers of type A in a real system relative to an ideal gas possessing the same atom number density for B, $\rho_B = N_B/V$, where N_B is the average number of atoms of type B within the system volume V surrounding the centers in A.[202] When the distribution of B about a single point in space, labeled 0, is of interest in a simulation (e.g., the distribution of the oxygens and hydrogens of water about the center of mass of a micelle), one can determine the average number of atoms located in a spherical shell of thickness dr centered on the point using the radial distribution function $g_{0B}(r)$ as follows:[190]

$$n_B(r)\,dr = g_{0B}(r)\,\rho_B\left(4\pi r^2 dr\right), \tag{3.5.1}$$

where $4\pi r^2 dr$ is the volume of the spherical shell, and ρ_B is the time-averaged number density of B within a spherical volume V of radius equal to half of the simulation box width, R. Note that

$$N_B = \rho_B V = \rho_B\left(\frac{4}{3}\pi R^3\right) = \int_0^R g_{0B}(r)\,\rho_B\left(4\pi r^2 dr\right). \tag{3.5.2}$$

Since $\int_0^R \rho_B\left(4\pi r^2 dr\right) = \rho_B V$, it is clear that to the extent that $g_{0B}(r)$ is less than unity in any region of space, there must be another region in which it is greater than unity. If $g_{0B}(r)$ is uniformly equal to unity, the uniform ideal gas state is achieved. Van der Waals and Coulombic interactions in a simulation ensure that $g_{0B}(r)$ will always be non-uniform, although the extent of non-uniformity will be system dependent.

From Eq. 3.5.1, the functional form for $g_{0B}(r)$ is given by[190]

$$g_{0B}(r) = \frac{1}{\rho_B} \sum_{b=1}^{N_B} \frac{\delta(r_{0b} - r)}{4\pi r^2}, \tag{3.5.3}$$

where the substitution $n_B(r) = \sum_{b=1}^{N_B} \delta(r_{0b} - r)$ has been made, with δ being the Dirac delta distribution, r_{0b} the radial position of atom b in set B about the origin 0, and $\delta(r_{0b} - r)$ being an ensemble-average quantity. For molecular dynamics simulations, all such averages are time-averages, based on the simulation trajectory.[203]

Practical calculations of $g_{0B}(r)$ and, more generally, of $g_{AB}(r)$ involve histogram-binning approaches.[202] Bin widths are selected in order to balance the desire to see rapid changes in structure (e.g., narrow hydration shells) with the need for sufficient statistical accuracy. The bin width that I have used is 0.05 nm, a width which is less than half of the van der Waals radius of hydrogen (see Table F.1 in Appendix F). Examples of radial distribution function profiles $g_{0B}(r)$ for sets B containing water, ions (when present), and surfactant head groups (about micelle centers of mass) are shown in Figures 3-3 and 3-4. These profiles are used for the determination of dividing surface locations, as described in Section 3.3.

Radial distribution functions can also be calculated for a collection of atoms about another set of atoms. Equation 3.5.3 is then written as[190]

$$g_{AB}(r) = \frac{1}{\rho_B} \frac{1}{N_A} \sum_{a=1}^{N_A} \sum_{b=1}^{N_B} \frac{\delta(r_{ab} - r)}{4\pi r^2}, \tag{3.5.4}$$

where A indicates the set of atoms serving as centers, B is the set of atoms for which the distribution is intended, N_A and N_B are the total number of atoms in sets A and B, respectively, and r_{ab} is the distance between a particular atom a in A and an atom b in B. The functional form for $g_{AB}(r)$ in Eq. 3.5.4 is a simple average of the individual $g_{aB}(r)$ profiles centered on a. Radial distribution functions of the form $g_{AB}(r)$ are useful to determine appropriate cutoffs for counting hydrating contacts, as described in Section 3.5.3.

The implementation of $g_{AB}(r)$ within the GROMACS software package is ac-

cessed via the command **g_rdf**, using the "-rdf atom" option and the "-com" flag if performing a center of mass calculation.

3.5.3 Counting Hydrating Contacts and Determining Fractional Degrees of Hydration

In Figure 3-8, a sodium dodecyl sulfate (SDS) surfactant is depicted in both micelle and monomer environments. A key output of a simulation is the trajectory of the system – a series of positions for each atom in the system recorded as a function of time. Based on the selection of a proper cutoff radius, then, one can readily count the number of hydrating contacts experienced by each surfactant group at any instant of time by comparing the positions of atoms defined as hydrating to the positions of the atoms constituting the groups of interest. Following Stephenson and coworkers, hydrating atoms are considered as those capable of participating in the hydrogen bonding network of water (the hydrogens and oxygen of water itself; ions such as sodium; hydrogen bond donors, such as the hydrogens attached to electronegative atoms like oxygen or nitrogen; and hydrogen bond acceptors, such as the oxygens in ketones, ethers, and esters).[108] Furthermore, contacts are considered to be intermolecular in origin only (e.g., in the monomer simulation, only contacts with ions and water are considered, and not contacts with hydrating atoms within the same molecule).[204] Then, the fractional degree of hydration can be computed as

$$f_a = \frac{N^{SIH}_{hydr,mic}(a) - \left(N^{SIH}_{hydr,mon}(a) - N^{SI}_{hydr,mon}(a)\right)}{N^{SI}_{hydr,mon}(a)}, \tag{3.5.5}$$

where SIH indicates that all solvent (water), ion, and hydrogen bond acceptors or donors are considered, and SI indicates that only solvent and ions are considered. The purpose of subtracting the term $N^{SIH}_{hydr,mon}(a) - N^{SI}_{hydr,mon}(a)$ in the numerator of Eq. 3.5.5 is to approximately eliminate the counting of contacts of a group with a hydrogen bond acceptor or donor on the same molecule (recall that the counting of contacts is intended to capture intermolecular contacts). Note that this subtraction is not done directly due to limitations in the routine used by GROMACS to count

contacts (i.e., there is no mechanism for prohibiting intramolecular contacts in that program).

As shown in Figure 3-8, the process of counting hydrating contacts requires selection of a cutoff radius. Determination of an appropriate value for this radius requires defining the notion of contact itself. For the purposes of the CSMT framework, where the goal is to quantify the hydrophobic effect, contact between a group on the surfactant molecule and a hydrating group should indicate that the hydrating group has been in some way affected by the presence of the surfactant group. It is well known that both hydrophilic and hydrophobic groups affect the local structure of water, by inducing the formation of solvation shells,[205–208] in which water is in some sense immobilized relative to bulk conditions.[183,205] Accordingly, it makes sense to select a cutoff which encompasses the relatively immobilized water molecules and other hydrating groups.

In order to assess what an appropriate cutoff value might be, and whether a single cutoff value can be used across different group types and different surfactant molecules, I simulated radial distribution functions for water about various groups in sodium dodecyl sulfate (SDS) and octaethylene glycol monododecyl ether ($C_{12}E_8$), as shown in Figures 3-9 to 3-12. In Figures 3-9 and 3-11, carbon groups at the start, middle, and end of the alkyl chain in SDS and $C_{12}E_8$, respectively, were selected. In Figure 3-10, the various groups in the sulfate head of SDS were considered. In Figure 3-12, oxygens throughout the ethoxylate region in $C_{12}E_8$ were considered, including the first ether oxygen connected to the alkyl tail, an intermediate (the fourth) ether oxygen, and the terminal alcohol group. In all these figures, the presence of peaks and valleys in the rdf profiles indicate structuring of water (representing regions of higher and lower density characteristic of the solvation shell concept, with a more compact structuring about hydrophilic groups, such as oxygen). Based on these results, I have found that a cutoff of 0.55 nm provides a fairly consistent demarcation between the region inducing structure in water and the bulk region (where rdfs approach unity and do not exhibit large peaks and valleys). This cutoff appears to work well for both SDS and $C_{12}E_8$, and, furthermore, appears to be reasonably consistent with

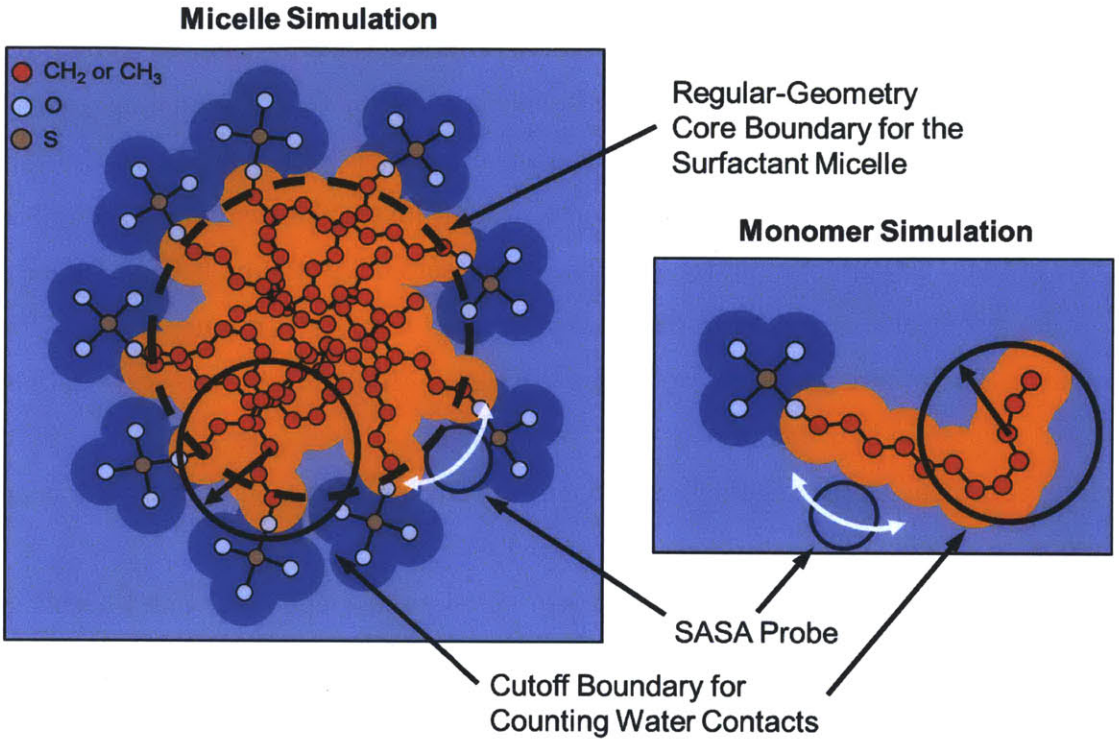


Figure 3-8: Schematic illustrating the concept of counting water contacts and determining solvent accessible surface areas (SASAs) in monomer and micelle computer simulations. The surfactant depicted is sodium dodecyl sulfate (SDS). The centers of individual atoms are denoted with small circles (see the diagram for the color code), while the large circle surrounding each center approximates the van der Waals radii of the atoms (sodium and water not shown for clarity). The SASA probes approximate the radius of water as 1.4Å, and can approach atom centers only as close as the sum of the probe radius and the atom van der Waals radius. Adjacent atoms can obstruct the probe, decreasing SASA accordingly. The monomer simulation is used to obtain $SASA_a$, while the micelle simulation is used to obtain $SASA_{core}$ (both $SASA$'s are displayed in orange). The obtained $SASA_{core}$ is always larger than the idealized core boundary (depicted as a dashed line, and determined using a dividing surface approach, as described in Section 3.3). Simulations also provide a means to determine the number of hydrating contacts that each group experiences in the two environments. The cutoff boundary for counting contacts is determined as described in the text.

consideration to both hydrophobic and hydrophilic groups. In Chapters 5 and 6, I will make use of this cutoff in making micellization predictions.

3.5.4 Group and Core Solvent Accessible Surface Areas

The concept of a solvent accessible surface area (SASA) was first introduced by Lee and Richards[184] in their study of the structural characteristics of proteins. Computing the SASA requires knowledge of the van der Waals radii, R_{VdW}, of all atoms in the molecule (see Table F.1 in Appendix F), and the selection of a spherical probe of radius R_{probe}, which approximates the size of the solvent of interest. For water, Lee and Richards used the value $R_{probe} = 1.4\mathring{A}$,[184] which I have also used in this Thesis. Conceptually, the SASA is the area of the surface obtained by rolling the probe about the molecule (as shown in Figure 3-8 of Section 3.5.3), treating both the probe and the molecule as hard surfaces which cannot interpenetrate and are always in contact. The center of the probe traces out the surface, and consequently the surface is always located, at minimum, a distance $R_{VdW} + R_{probe}$ from the nearest atom center. Note that the notion of rolling the probe is for illustrative purposes only – the surface itself may be discontinuous. In GROMACS, the actual calculation of SASA utilizes the Double Cubic Lattice Method developed by Eisenhaber and coworkers,[209] which uses two lattices containing atom center and surface dot information, respectively. This algorithm is a variation of the original algorithm due to Shrake and Rupley,[210] which also used dots to represent three-dimensional surfaces.

Computation of SASA in GROMACS requires specification of a calculation group and an output group. The calculation group consist of all groups which are considered to contribute to the full surface being probed. The output group can be any subset of the calculation group. The resulting SASA value is that associated with this subset, where groups which are not in the output group, but are still in the calculation group, can impact the SASA calculation by shielding the output groups. For $SASA_a$ calculations in the monomer state, the calculation group is the entire surfactant molecule, and the output group is the group a of interest. For $SASA_{core}$ calculations, both the calculation group and the output group comprise all the sur-

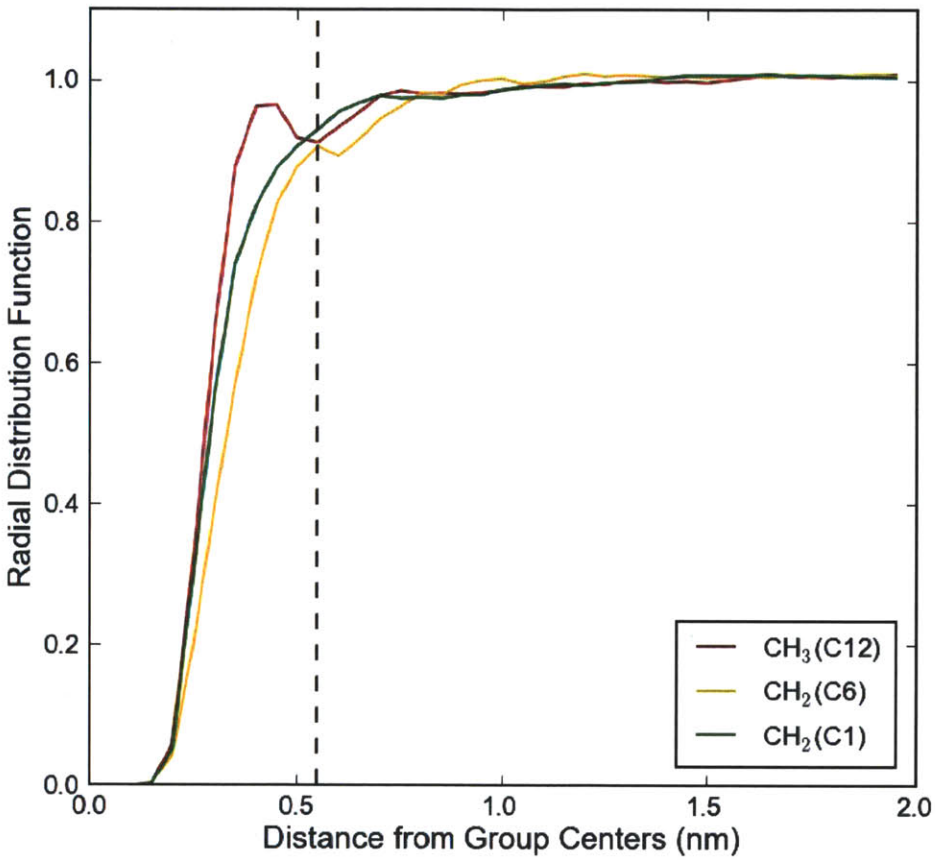


Figure 3-9: Radial distribution functions of water about various alkyl groups in sodium dodecyl sulfate (SDS) in the monomer state. Three carbon groups at different positions in the alkyl chain are selected for comparison purposes, ranging from the terminal methyl group (C12) to the methylene group alpha to the sulfate (C1). The vertical dashed line indicates a proposed cutoff of 0.55 nm to determine the onset of the bulk disordered state for water (i.e., the boundary of the hydration shells surrounding the selected groups). Note that there are few, broad peaks and valleys in the rdf of water, indicating a looser structuring of water about these hydrophobic groups. The proposed cutoff appears to demarcate fairly well the start of the disordered region for water, as indicated by an rdf profile that rises to unity.

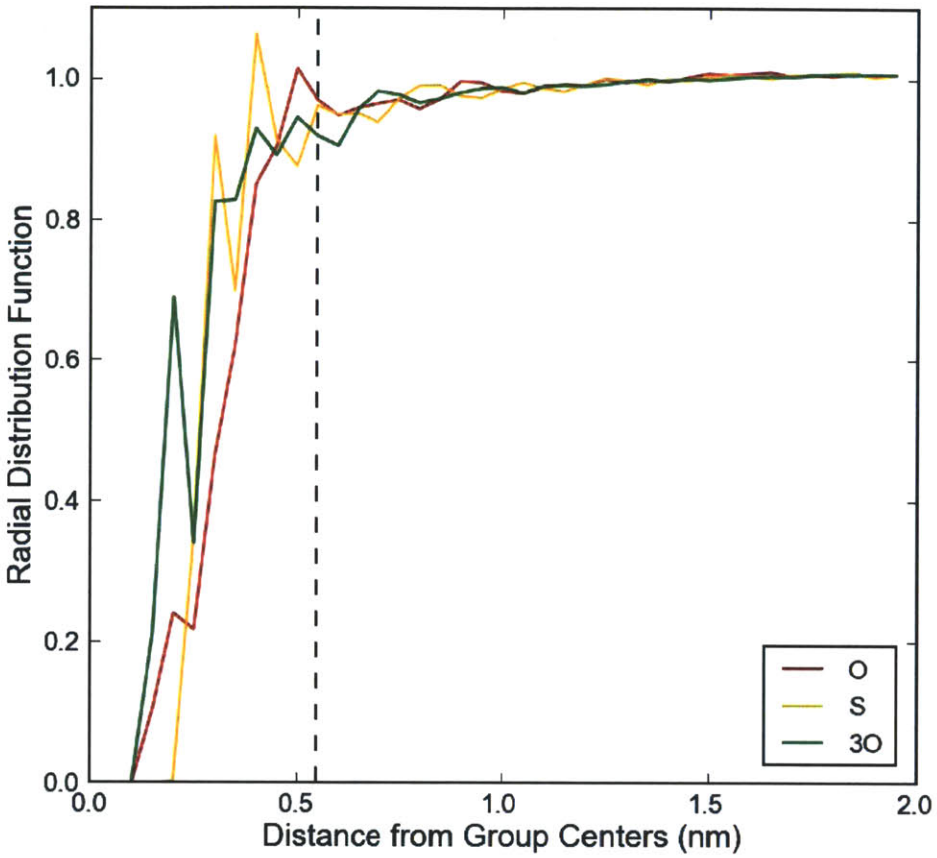


Figure 3-10: Radial distribution function of water about various atoms in the sulfate head group of sodium dodecyl sulfate (SDS) in the monomer state. The oxygen connecting the sulfur to the alkyl tail, the sulfur itself, and the three oxygens bearing the majority of the sulfate charge are considered separately. The vertical dashed line indicates a proposed cutoff of 0.55 nm to determine the onset of the bulk disordered state for water (i.e., the boundary of the hydration shells surrounding the selected groups). Note that there are numerous peaks and valleys in the rdf of water, indicating multiple shells. The proposed cutoff appears to demarcate fairly well the start of the disordered region for water, as indicated by an rdf profile that rises to unity.

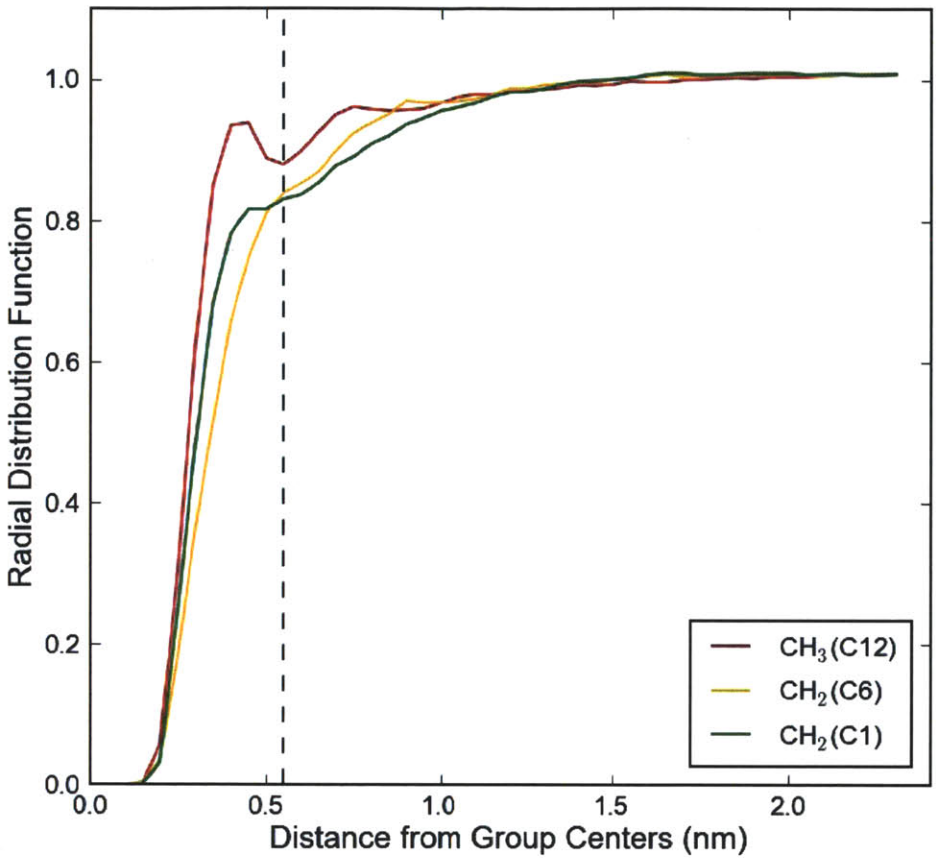


Figure 3-11: Radial distribution function of water about various alkyl groups in octaethylene glycol monododecyl ether ($C_{12}E_8$) in the monomer state. Three carbon groups at different positions in the alkyl chain are selected for comparison purposes, ranging from the terminal methyl group (C12) to the methylene group alpha to the first ether oxygen (C1). The vertical dashed line indicates a proposed cutoff of 0.55 nm to determine the onset of the bulk disordered state for water (i.e., the boundary of the hydration shells surrounding the selected groups). Note that there are few, broad peaks and valleys in the rdf of water, indicating a looser structuring of water about these hydrophobic groups. The proposed cutoff appears to demarcate fairly well the start of the disordered region for water, as indicated by an rdf profile that rises to unity.

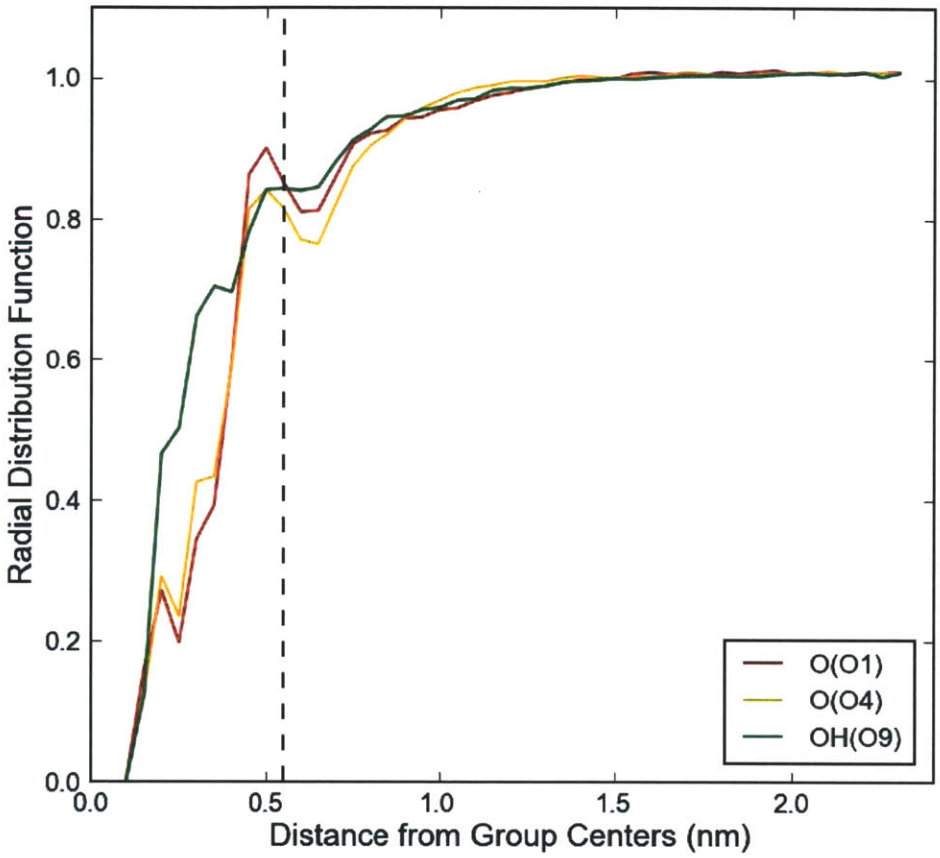


Figure 3-12: Radial distribution function of water about various oxygen groups in octaethylene glycol monododecyl ether ($C_{12}E_8$) in the monomer state. The first ether oxygen connected to the alkyl tail (O1), the fourth ether oxygen (O4), and the terminal alcohol group (O9) are considered. The vertical dashed line indicates a proposed cutoff of 0.55 nm to determine the onset of the bulk disordered state for water (i.e., the boundary of the hydration shells surrounding the selected groups). Note that there are numerous peaks and valleys in the rdf of water, indicating multiple shells. The proposed cutoff appears to demarcate fairly well the start of the disordered region for water, as indicated by an rdf profile that rises to unity.

factant groups identified as tail. Note that heads are not considered because these are not introduced in the molecular-thermodynamic thought process until the steric free-energy step (see Section 2.4 of Chapter 2 for more information). However, the amount of surface shielded by the head-tail bond must be subtracted, just as is done to calculate the interfacial area of the idealized micelle geometry exposed to water for the purposes of assessing the interfacial free-energy contribution to the free energy of micellization. Denoting the output of GROMACS as $SASA'_{core}$, the value of $SASA_{core}$ to be used in calculations like Eq. 3.2.10 is given by

$$SASA_{core} = SASA'_{core} - \underline{A}_0, \tag{3.5.6}$$

where $\underline{A}_0/n$ (with n the micelle aggregation number) is typically taken to be $21\mathring{A}^2$.[148]

3.5.5 Analysis Protocols

Analysis of Monomer Simulations

In the steps which follow, all GROMACS commands are accompanied with the specification of various files pertaining to the simulation and with a file containing indices of the atoms of interest for any particular calculation (an index group file). The full details of these commands can be found in the GROMACS manual.[190]

(1) Center the simulation on the monomer using **g_trjconv** with the "-pbc nojump" option and the "-center" flag.

(2) Determine the time-dependent behavior of the solvent accessible surface area of the monomer, $SASA_{monomer}$, using **g_sas** with the "-probe 0.14" option (for a probe of $1.4\mathring{A}$).

(3) Select the time window for analysis, based on convergence in the $SASA_{monomer}$ value (allowing a minimum start time of 5 ns to avoid initial simulation conditions). Note that some monomers exhibit self-association between different regions of the monomer (e.g., in ethoxylated surfactants, some period of time is spent with the ethoxylated head group interacting with the alkyl tail, reducing $SASA_{monomer}$ accordingly; in twin-tailed surfactants one can observe the two tails interacting) – in

these cases, the convergence of $SASA_{monomer}$ to a steady distribution is very important. For example, although there may be approximately two discrete values for $SASA_{monomer}$ observed at equilibrium (associated and not-associated), the convergence in the relative frequency of these two states is what is important.

(4) Determine $SASA_a$ for each group a in the surfactant. The same option for **g_sas** as in Step (2) above is used.

(5) Determine the number of hydrating contacts excluding hydrogen-bonding atoms in the surfactant (i.e., contact with just water and ions), and include them using **g_mindist** with the "-d 0.55" option (for a cutoff radius of 0.55 nm, writing to a file specified with the "-on" flag. Use **g_analyze** to perform block-averaging of the transient data obtained.

Analysis of Micelle Simulations

As with the monomer simulation analysis protocol above, full details for the commands mentioned below are found in Ref. [190].

(1) Center the simulation on the micelle using **g_trjconv** with the "-pbc nojump" option and the "-center" flag.

(2) Determine the time-dependent behavior of the solvent accessible surface area of the micelle, $SASA_{micelle}$, using **g_sas** with the "-probe 0.14" option (for a probe of 1.4$\mathring{A}$).

(3) Select the time window for analysis, based on convergence in the $SASA_{micelle}$ value (allowing a minimum start time of 5 ns to avoid initial simulation conditions).

(4) Determine the radial distribution functions (rdfs) for each surfactant group relative to the center of mass of the micelle using **g_rdf** with the "-rdf atom" option, the "-com" flag and the "-bin 0.05" option (for a bin width of 0.5$\mathring{A}$).

(5) Determine the time-averaged, median radial positions of all surfactant groups based on the rdfs obtained in Step (4) above. This is accomplished by determining the fraction of time that a group b of type B spends in each histogram bin of the rdf for the full set B. This is the probability, $p_{b,bin}$, of finding a group b in the bin,

computed from the set of $g_{0B,bin}$ values as follows:

$$\begin{aligned} p_{b,bin} &= \frac{n_{bin}}{\sum_{bins} n_{bin}} \\ &= \frac{g_{0B,bin}\rho_B 4\pi r_{bin}^2 \Delta r_{bin}}{\sum_{bins} g_{0B,bin}\rho_B 4\pi r_{bin}^2 \Delta r_{bin}} \\ &= \frac{g_{0B,bin} r_{bin}^2}{\sum_{bins} g_{0B,bin} r_{bin}^2}, \end{aligned} \tag{3.5.7}$$

where r_{bin} is the center of the bin, and $g_{0B,bin}$ is the value directly obtained from **g_rdf**. One then sums across bins, starting from the center of mass of the micelle, and proceeding outward radially, until the probability first reaches (or exceeds) 0.5. Linear interpolation is then used to determine the radial position corresponding to the median (the position where the cumulative probability is exactly 0.5). Other percentiles can be obtained in a similar fashion, with a cumulative probability of 0.25 corresponding to the first quartile, and a cumulative probability of 0.75 corresponding to the third quartile. This technique was used to prepare Figures 3-5 and 3-6 in Section 3.3.

(6) Sort the surfactant groups by their median radial positions from outermost to innermost. Develop rdfs about the center of mass of the micelle for the set of water plus ions, water plus ions plus all instances of the outermost group, water plus ions plus all instances of the two outermost groups, etc., working down the sorted list. The same options for **g_rdf** as in Step (4) above are used.

(7) Determine the radial location of the dividing surface according to the iterative scheme described in Section 3.3. Record the set of tail groups based on this dividing surface (the set of groups with median positions less than the radial location of the dividing surface).

(8) Determine $SASA_{core}$. The same option for **g_sas** as in Step (2) above is used.

(9) Determine the number of hydrating contacts (contacts with water, ions, and hydrogen-bonding atoms) using **g_mindist** with the "-d 0.55" option (for a cutoff radius of 0.55 nm, writing to a file specified with the "-on" flag. Use **g_analyze** to perform block-averaging of the transient data obtained.

CSMT Calculations

This section focuses on the computation of $g_{tr,CSMT}$, according to Eq. 3.2.10, which I reproduce below for convenience:

$$g_{tr,CSMT} = g_{tr,0} + \sum_a (1 - f_a + \zeta_a f_a)\, g_{tr,a} + \widehat{f}_{int} \left[-1 + \sum_a \zeta_a f_a n \frac{SASA_a}{SASA_{core}} \right] . \quad (3.5.8)$$

(1) Compute f_a for each group a in the surfactant, using the hydrating contact data prepared in Step (5) of the monomer simulation analysis protocol and Step (9) of the micelle simulation analysis protocol, according to Eq. 3.5.5.

(2) Determine ζ_a as 0, if group a is a head group, and as 1, if group a is a tail group, based on the head/tail identification obtained in Step (7) of the micelle simulation analysis protocol.

(3) Determine $\widehat{f}_{int}$ using the MT model for the interfacial free energy, as applied to the simulated aggregate with aggregation number n and the identified surfactant tail.

(4) Select a solubility model from which $g_{tr,0}$ and $g_{tr,a}$ can be readily calculated (see Section F.5 of Appendix F).

(5) Compute $g_{tr,CSMT}$ according to Eq. 3.5.8.

3.6 Conclusions

In this Chapter, I described the general CSMT framework developed by Stephenson and coworkers,[108–110] which aims to improve the quantification of the hydrophobic effect (which drives surfactant micellization behavior in aqueous media) using molecular dynamics computer simulations. I discussed how the use of only two simulations (that of a surfactant in a monomer state and a micelle state) is required in order to determine a transfer free energy that can be used in the original MT framework to make micellization predictions without further simulation effort required. This enables the development of libraries of simulation data for pure surfactants which can be used by users lacking advanced computational resources.

I then presented one of my key contributions to the CSMT approach: a method for identifying the head and tail of a surfactant using a dividing surface approach applied to the simulated surfactant micelle. Head and tail identification is critical to the success of both the MT and CSMT models. Previously, rules of thumb were used to make this determination. The dividing surface approach can be applied equally well to ionic and nonionic surfactants and is made self-consistent through an iterative scheme which yields the final location of the dividing surface, taken to be equivalent to the micelle core-water interface invoked in MT modeling. Practical examples involving sodium dodecyl sulfate (SDS) and octaethylene glycol monododecyl ether ($C_{12}E_8$) were presented.

Next, I discussed the simulations settings that I have used in applying the CSMT framework to the systems studied in Chapters 5 and 6. I also discussed briefly the automation framework that I developed to rapidly prepare simulation input files based on user input of surfactant chemical structures. Full details of this useful framework are presented in Appendix C.

Finally, I discussed the specific analysis protocols applied to the monomer and micelle simulation results in order to obtain fractional degrees of hydration, solvent accessible surface areas, and, ultimately, the CSMT transfer free energies needed for micellization property prediction.

In the next Chapter, I will describe a statistical-thermodynamic theory for evaluating the packing of surfactant tails within a micelle core. Specifically, this theory is a single-chain, mean-field theory which aims to quantify free-energy penalties associated with the confinement and constraint of these tails, provided an appropriate chain model for the internal degrees of freedom of each surfactant tail. I present the Rotational Isomeric State model as one such chain model, and describe how packing calculations for branched surfactant tails can be performed.

Chapter 4

Chain Packing Theory for Linear and Branched Surfactant Tails

4.1 Overview

The packing free-energy contribution to the free energy of micellization describes the entropic penalty associated with confining surfactant tails within a micelle core – a penalty which is a function of not only the geometry of the micelle core but also of the specific chemical structure of the surfactant tail.[81,82] This confinement is accompanied by a constraint which serves to distinguish the micelle core model from an oil droplet; namely, in contrast to the oil droplet, where hydrocarbon chains are free to translate throughout the droplet, at least one end of each surfactant tail within a micelle core is constrained to reside at the core-water interface. This constraint arises from the chemical bond between the surfactant tail and head, and the preferential localization of the head at the solvent-side of the interface.[35] Note that this constraint does not prevent lateral translation of the head-tail connection along the interface. In fact, the dynamic behavior of surfactant tails within a micelle core is such that the core environment is still quite disordered and liquid-like,[34] as shown by various alkene oxidation,[186] radioactive tracer,[211] and fluorescent probe experiments.[212] In analogy to the phase behavior of hydrocarbon oils at room temperature, this liquid-like state within the core is expected to be valid for linear alkyl tails up to 16 carbons in length,

beyond which some preferred alignment may begin to occur.[34] Since branching in alkanes generally lowers the melting point relative to that for a linear alkane of the same primary chain length,[11] it is not unlikely that a fluid-like environment in micelle cores comprising even longer-tailed branched surfactants may be found.

To model the behavior of the surfactant tails within the liquid-like environment of a micelle core (specifically, to determine the probability distribution of tail orientations and conformations), we first assume that the behavior of any single chain within a homogeneous, multicomponent micelle can be viewed as being representative of all the other chains of the same species.[82] Accordingly, we only need to analyze the energetic behavior of one representative chain for each species present, thereby reducing the complexity of the problem considerably.[148] To relate the problem of developing chain statistics for the packed state relative to an unperturbed state, for which chain models are typically developed,[213] the concept of a mean field is introduced, which aims to capture implicitly the impact of all the chains in the core on the representative chain to be studied.[214] In this thesis, I make use of the single-chain, mean-field (SCMF) theory of Ben-Shaul, Szleifer, and Gelbart,[82] which enables the calculation of any observable which depends on chain conformation, given a chain model and the specification of constraints appropriate to the system of interest. Examples of conformation-dependent observables include the packing free energy itself, various order parameters, and moments of the distribution of individual atomic groups of the chain in space. For example, using the SCMF theory one can determine the mean radial position of the terminal methyl group of a linear alkyl chain (see Chapter 5 for examples), thereby enabling comparison with alkene oxidation experiments, which appear to indicate that this group is often located quite close to the micelle core-water interface, in contrast to interior methylene groups in the chain.[186]

When the SCMF theory is applied to a micelle core, we will see that the mean field arises naturally from application of a constant density constraint throughout the micelle core, reflecting the experimental observation of core fluidity. The packing free energy can be accurately computed from: (a) knowledge of the internal energy of the central chain as a function of conformation, (b) possession of a scheme for sampling

chain conformations, and (c) a self-consistent solution for the mean field.[81,214]

In this Appendix, I first describe the equations involved in the SCMF theory, deriving the expressions relevant to chain packing in a micelle core. Then, since the SCMF theory requires a chain model to assess the internal energy of a chain as a function of its conformation, I will discuss one particularly useful chain model that I use in my chain packing calculations: the Rotational Isomeric State (RIS) model of Flory.[213] Specifically, I will present the RIS model applied by Mattice and Santiago to linear alkanes[215] (based on the earlier work of Abe, Jernigan, and Flory[216]), as well as the RIS model applied by Mattice to branched alkanes.[217] These two classes of chains constitute the main types of chains encountered in the surfactant systems discussed in this thesis. As needed, the specific values of the RIS parameters used may be modified to account for stiffer chains (e.g., fluorocarbons[161,218]).

4.2 Single-Chain, Mean-Field (SCMF) Theory

4.2.1 The System and Its Natural Variables

In the molecular-thermodynamic framework presented in Chapter 2, a surfactant solution is modeled as containing dispersed components (monomers) and aggregated components (micelles). The solution as a whole is kept at constant temperature (T), pressure (P_{ext}), and number of moles of each species ($\{N_{i,solution}\}$, where i indexes the species). Each type of micelle is parameterized by a set of aggregation numbers for each species ($\{n_{i,\xi}\}$, where i indexes the species and ξ indexes the micelle type), a geometry (typically a sphere, cylinder, or bilayer), and a size. Assuming that the micelles are incompressible, with a liquid-like core, the micelle core volume can be calculated as follows:

$$\underline{V}_{\xi} = \sum_{i} n_{i,\xi} v_{i,\xi,core}, \tag{4.2.1}$$

where $v_{i,\xi,core}$ is the molecular volume of species i present in the core of a micelle of type ξ. In the case of a surfactant, $v_{i,\xi,core}$ is the surfactant tail volume, computed from experimental data for the density of pure oils at T and P_{ext}. (The underbar

used with the micelle core volume, $\underline{V}_\xi$, denotes extensivity.) The oils are chosen to be analogous to the surfactant tail (e.g., a 12 carbon linear alkyl tail would have dodecane as an analog).[81] (In the case of a fully-encapsulated solubilizate, $v_{i,\xi,core}$ would be the molecular volume of the solubilizate, determined from the density of the pure solubilizate in a liquid state.)

Within the SCMF theory, the micelle core itself is defined as the system. It is modeled as containing a constant number of tails ($n_\xi = \sum_i n_{i,\xi}$), with constant core volume ($\underline{V}_\xi$) and temperature ($T_\xi = T$). From a statistical-thermodynamic standpoint, holding these natural variables constant corresponds to: (i) the use of a canonical ensemble in describing the thermodynamics of the system, and (ii) the use of a Helmholtz free energy in describing the energetics of the system. As a result, when discussing the packing free energy, I will use $\underline{F}_{pack}$ for the total micelle core packing free energy (analogous to an extensive quantity) and f_{pack} for the per-surfactant packing free energy (analogous to an intensive quantity), where $\underline{F}$ and f indicate a Helmholtz free energy. As indicated in Chapter 2, a $P\underline{V}$ work term of the form $P_{ext}\underline{V}_\xi$ is needed to convert the Helmholtz free energy of the micelle core system to a Gibbs free energy for the purpose of evaluating reference-state chemical potentials within the micelle-containing solution as a whole. However, as also demonstrated in that Chapter, the $P_{ext}\underline{V}_\xi$ term is a negligible quantity for liquid systems. Therefore, although one will sometimes encounter the symbol g_{pack} in the literature in preference to f_{pack},[81,148,161] the two expressions can be viewed as essentially synonymous (in the references cited, they are equivalent – the $P_{ext}\underline{V}_\xi$ has been set to zero).

4.2.2 The Maximum Entropy Principle

With the system defined, one can proceed with the derivation of the equations of the SCMF theory. There are two equivalent approaches. The first, which I will not show, begins with a description of the full partition function of the system, and progresses through a series of approximations to the mean-field result.[219] The second makes use of a maximum entropy (MAXENT) formalism, and I will focus on this approach, due to its clarity of presentation.[82,219]

The MAXENT formalism, as applied to thermodynamic systems, was developed by Jaynes[220,221] from the seminal work of Shannon on information theory,[222] where the entropy of a probability distribution is defined as follows:

$$H\left(p_1, p_2, \ldots, p_M\right) = -K \sum_m p_m \ln p_m, \tag{4.2.2}$$

where p_m is the probability of the system being in one of M discrete microstates, indexed by m, and K is a constant.[220] Here, the term microstate refers to a description of the physical system at a molecular level – e.g., the specification of $6N$ positions and velocities in a system containing N atoms.[223] The function H is always greater than or equal to zero, is maximal when all microstates have the same probability, and is minimal when only one state has a non-zero probability of unity.[222] Note that, based on quantum-mechanical arguments, physical systems are always considered to have a finite value for M, in spite of the fact that many variables, such as atom position and internal energy, are often approximated as being continuous for convenience of calculation.[220,224]

The MAXENT formalism aims to determine a functional form for p_m (referred to as the probability distribution function, or pdf[219]) in an underspecified system through maximization of H/K subject to constraints which represent the knowledge that one possesses regarding the system at hand. The probability distribution derived in this manner maximizes the uncertainty in how the system partitions between states. That is, no additional information is imposed on the system beyond the constraints – a "principle of minimum prejudice".[223] When used to describe the thermodynamics of physical systems, H in Eq. 4.2.2 is $\underline{S}$, the entropy, and K is k_B, the Boltzmann constant.[219]

Using MAXENT, one can readily derive the probability distribution for a canonical ensemble by invoking just two constraints:

$$\sum_m p_m = 1, \tag{4.2.3}$$

representing the normalization of the pdf, and

$$\sum_m p_m \epsilon_m = \langle \epsilon_m \rangle = \underline{U}, \tag{4.2.4}$$

representing knowledge of the total system internal energy, $\underline{U}$, where ϵ_m is the internal energy associated with microstate m. The constraints of constant volume and number of particles are implicitly contained in the enumeration of microstates. That is, microstates which violate either of these two constraints (e.g., by proposing atom positions outside of the system volume) are simply not considered.

The constraints in Eqs. 4.2.3 and 4.2.4 are added to H/K using the two Lagrange multipliers $\psi - 1$ and β, respectively, prior to maximization.[224] (Note that the choice of $\psi - 1$ as the Lagrange multiplier, rather than simply ψ, affords a convenient cancellation later in the derivation.) The constrained maximization is then carried out using variational calculus to determine the functional form of p_m. That is, we write the variation

$$\delta \left[\sum_m p_m \ln p_m + (\psi - 1) \left(-1 + \sum_m p_m \right) + \beta \left(- \langle \epsilon_m \rangle + \sum_m p_m \epsilon_m \right) \right] = 0, \tag{4.2.5}$$

which, when evaluated using the chain rule, yields

$$\sum_m \delta p_m \left[\ln p_m + \psi + \beta \epsilon_m \right] = 0. \tag{4.2.6}$$

Since the variations δp_m are in general independent, the quantity in brackets in Eq. 4.2.6 must be equal to 0 for the relation to hold, such that

$$p_m = \frac{\exp\left(-\beta \epsilon_m\right)}{Z}, \tag{4.2.7}$$

where

$$Z \equiv \exp\left(\psi\right) \tag{4.2.8}$$

is a partition function.

Inserting Eqs. 4.2.7 and 4.2.8 into Eq. 4.2.3 yields

$$Z = \sum_m \exp\left(-\beta\epsilon_m\right), \tag{4.2.9}$$

and

$$\psi = \ln Z = \ln \sum_m \exp\left(-\beta\epsilon_m\right). \tag{4.2.10}$$

At this point the Lagrange multiplier β is still an unknown. Below, we proceed to ascertain its significance.

From Eq. 4.2.10, one can see that

$$-\frac{\partial\psi}{\partial\beta} = -\frac{\partial \ln Z}{\partial\beta} = \frac{1}{Z}\sum_m \exp\left(-\beta\epsilon_m\right)\epsilon_m = \sum_m p_m\epsilon_m = \langle\epsilon_m\rangle. \tag{4.2.11}$$

In fact, any constraint of the form

$$\sum_m p_m X_m = \langle X_m\rangle, \tag{4.2.12}$$

which expresses knowledge of the expected value of X_m through some physical measurement, and is included in the maximization procedure through a Lagrange multiplier λ_X, will produce a relation[220]

$$-\frac{\partial\psi}{\partial\lambda_X} = -\frac{\partial \ln Z}{\partial\lambda_X} = \langle X_m\rangle. \tag{4.2.13}$$

Another important relation can be found by taking the derivative of $\langle\epsilon_m\rangle$ with respect to β, which yields[224]

$$\begin{aligned}\frac{\partial\langle\epsilon_m\rangle}{\partial\beta} &= \frac{1}{Z}\frac{\partial}{\partial\beta}\left(\sum_m \exp\left(-\beta\epsilon_m\right)\epsilon_m\right) - \frac{1}{Z^2}\sum_m \exp\left(-\beta\epsilon_m\right)\epsilon_m\frac{\partial Z}{\partial\beta} \\ &= -\left\langle\epsilon_m^2\right\rangle + \langle\epsilon_m\rangle^2,\end{aligned} \tag{4.2.14}$$

or

$$\frac{\partial\langle\epsilon_m\rangle}{\partial\beta} = \frac{\partial U}{\partial\beta} = -\sigma^2\left(\epsilon_m\right), \tag{4.2.15}$$

where $\sigma^2(\epsilon_m)$ is the variance of ϵ_m. The variance is always a positive quantity, indicating that $\langle \epsilon_m \rangle$ (i.e., $\underline{U}$) always decreases with increasing β.

A general derivation relating β to the system temperature can be made by first showing that

$$\beta = \frac{N}{PV} \tag{4.2.16}$$

for the specific case of an ideal monatomic gas, using the MAXENT formalism to construct an expression for the pressure from a kinetic theory.[224] Relating Eq. 4.2.16 to the well-known ideal gas equation of state then yields

$$\beta = \frac{1}{k_B T}, \tag{4.2.17}$$

demonstrating that the Lagrange multiplier β is the thermodynamic beta $(1/k_B T)$ for this special case.[223,225] The constraint of constant temperature in a canonical ensemble thus determines the value of β.

Proving Eq. 4.2.17 for non-ideal systems involves the following thought process, outlined by Tribus.[224] (I) Select two rigid, closed systems of interest with possibly differing values of β, say β^I and β^{II}, such that $\beta^I \leq \beta^{II}$, without loss of generality. These two systems have corresponding initial energies $\underline{U}_0^I$ and $\underline{U}_0^{II}$, computed as in Eq. 4.2.4. (II) Allow the two systems to interact through energy exchange only (i.e., both systems remain closed). Conservation of energy requires that the combined system has an energy $\underline{U}^T = \underline{U}_0^I + \underline{U}_0^{II}$, but information is lost regarding the energies of the original two systems after equilibration. If these final energies of the two systems are denoted $\underline{U}_f^I$ and $\underline{U}_f^{II}$, one knows only that $\underline{U}_f^I + \underline{U}_f^{II} = \underline{U}^T$ holds. This implies

$$\left(\underline{U}_0^I - \underline{U}_f^I\right) + \left(\underline{U}_0^{II} - \underline{U}_f^{II}\right) = 0, \tag{4.2.18}$$

which is another formulation of energy conservation. From a MAXENT standpoint, knowledge of only the total energy $\underline{U}^T$ yields only one energy constraint, with a single Lagrange multiplier, say β^T, shared by the two connected systems. We can now evaluate three possibilities related to the change in energy of the systems after

contact. (1) $\underline{U}_f^I > \underline{U}_0^I$, which implies that $\beta^T < \beta^I$ (via Eq. 4.2.15), and $\underline{U}_f^{II} < \underline{U}_0^{II}$ (via Eq. 4.2.18), such that $\beta^T > \beta^{II}$ (via Eq. 4.2.15). However, $\beta^I \leq \beta^{II}$ by choice of system labels, and $\beta^T < \beta^I$ and $\beta^{II} < \beta^T$ leads to a contradiction, precluding the assumption that $\underline{U}_f^I > \underline{U}_0^I$ is possible. (2) $\underline{U}_f^I < \underline{U}_0^I$, which implies $\beta^T > \beta^I$, $\underline{U}_f^{II} > \underline{U}_0^{II}$, and $\beta^T < \beta^{II}$ (again, through application of Eqs. 4.2.15 and 4.2.18). In this case, $\beta^I < \beta^T < \beta^{II}$, implying that $\beta^I \neq \beta^{II}$. This is a valid outcome. (3) $\underline{U}_f^I = \underline{U}_0^I$, which implies $\beta^T = \beta^I$, $\underline{U}_f^{II} = \underline{U}_0^{II}$, and $\beta^T = \beta^{II}$ (applying the same relations used in (1) and (2)), implying that $\beta^I = \beta^{II}$. This is also a valid outcome.

To summarize these three cases, energy flows in only one direction: from a system with a low value of β to a system with a high value of β, and does not flow when the β values are equal (described as "thermal equilibrium"). As a result, β can be referred to as the "temper" of a system, such that comparison of the β value of two systems immediately provides an indication for the direction of energy flow when the two systems are brought into contact.[224]

It now remains to construct a meter to measure β, which will provide insight into the functional form of β in the general case. One can select a small, closed, and rigid system of monatomic ideal gas, such that Eq. 4.2.17 is already known to hold generally, and Eq. 4.2.16 can be evaluated experimentally through measurement of the pressure.[224] By the zeroth law of thermodynamics, if the meter reads the same value of β when connected to each of the two original systems in turn, those two systems are in equilibrium with each other, and each possesses that value of β. Accordingly, at thermal equilibrium, regardless of the composition or non-ideality of two systems brought into contact, Eq. 4.2.17 ($\beta = 1/k_B T$) is always a measurable outcome. Since Eq. 4.2.17 is independent of the extensive quantities of the meter, it is used in preference to Eq. 4.2.16.

The net result of the discussion above is that the MAXENT formalism yields the probability distribution for the canonical ensemble when the only constraints specified are Eqs. 4.2.3 and 4.2.4, and Eq. 4.2.17 is applied to the probability distribution

function that results (Eq. 4.2.7):

$$p_m = \frac{\exp\left(-\frac{\epsilon_m}{k_B T}\right)}{Z}. \tag{4.2.19}$$

One important consequence of the above arguments identifying the nature of β (and the reason for that presentation here) is that one can now write

$$\begin{aligned} \underline{F} &= \underline{U} - T\underline{S} \\ &= \sum_m p_m \left(\epsilon_m + k_B T \ln p_m\right) \end{aligned} \tag{4.2.20}$$

and show that *minimization* of $\underline{F}$ with respect to the single constraint of Eq. 4.2.3 yields the same result of Eq. 4.2.7:[219]

$$\delta\left[\sum_m p_m \left(\epsilon_m + k_B T \ln p_m\right) + \left(k_B T \left(\psi' - 1\right)\right)\left(-1 + \sum_m p_m\right)\right] = 0, \tag{4.2.21}$$

where $k_B T \left(\psi' - 1\right)$ has been chosen as a Lagrange multiplier for convenience, leading to

$$\sum_m \delta p_m \left[\epsilon_m + k_B T \ln p_m + k_B T \psi'\right] = 0, \tag{4.2.22}$$

and

$$p_m = \frac{\exp\left(-\frac{\epsilon_m}{k_B T}\right)}{Z}, \tag{4.2.23}$$

where

$$Z \equiv \exp\left(\psi'\right) = \sum_m \exp\left(-\frac{\epsilon_m}{k_B T}\right). \tag{4.2.24}$$

Equation 4.2.24 is clearly equal to Eq. 4.2.9 since Eq. 4.2.17 holds.

This form of the MAXENT formalism is generally more convenient for use, since one can propose any free energy expression $\underline{F}$ which depends on chain conformation, and follow the same minimization procedure to readily obtain the appropriate pdf p_m. As an example, if one wanted to incorporate the effect of enthalpic mixing between two dissimilar chains into the packing model (e.g., fluorocarbon and hydrocarbon chains),

one need only propose a functional form for the mixing free energy (e.g., a regular solution expression involving local volume fractions and an interaction parameter[139]) and combine that with the free energy in Eq. 4.2.20 to obtain a new free energy functional which can be minimized with respect to the usual constraints to obtain p_m.

4.2.3 Derivation of the Multicomponent Packing Probability Distribution Functions

The Helmholtz free energy of a system under constant $N\underline{V}T$ conditions was presented in general form in Eq. 4.2.20, repeated here for completeness:

$$\underline{F} = \sum_m p_m \left(\epsilon_m + k_B T \ln p_m\right), \tag{4.2.25}$$

where m is a microstate of the system corresponding to full specification of its degrees of freedom. In the SCMF theory, the Helmholtz free energy, $\underline{F}_{core,\xi}$, related to packing chains into a micelle core of type ξ, is simplified considerably by assuming that it can be calculated from the free energies associated with individual chains.[226] In other words,

$$\underline{F}_{core,\xi} = n_\xi \sum_i \alpha_{i,\xi} \left[\sum_{\omega_i} p_{\omega_i,\xi} \left(\epsilon_{\omega_i} + k_B T \ln p_{\omega_i,\xi}\right)\right], \tag{4.2.26}$$

where ω_i is a microstate of a central chain of species i, characterized by a full specification of its conformation and orientation,[226] $p_{\omega_i,\xi}$ is its associated probability distribution function in the specified core, ϵ_{ω_i} is its associated internal energy (independent of core type), $\alpha_{i,\xi}$ is the mole fraction of species i in the core, and n_ξ is the total core aggregation number. As discussed in Section 4.2.2, $k_B T = 1/\beta$. In the discussion that follows, I will drop all reference to the index ξ. It will be assumed that we are working with only one micelle type at a time in determining chain properties and packing free energies.

To apply the MAXENT formalism presented in the previous section, we now specify the system constraints. As usual, we require normalization of the probability

distributions. In this case we have such a constraint for each species i, as follows:

$$\sum_{\omega_i} p_{\omega_i} = 1. \tag{4.2.27}$$

We also have a new constraint imposed on the system due to the recognized fluidity of the core. Specifically, this information implies constant density throughout the core, which can be written as:

$$\sum_i \alpha_i \sum_{\omega_i} p_{\omega_i} v_{\omega_i}(r) = \frac{A(r)}{n}, \tag{4.2.28}$$

where $v_{\omega_i}(r)$ is the volume of species i (per length) at radial position r in microstate ω_i, and $A(r)$ is the area of the micelle core geometry at position r. For spheres, r is the usual radial position measured from the center of the sphere, with $A(r) = 4\pi r^2$. For cylinders, r is the cross-sectional radial position normal to the central axis, with $A(r) = 2\pi r L$, and L the length of the cylinder. For bilayers, r is the normal distance from the bilayer mid-plane and $A(r) = 2LW$, where L is the bilayer length, and W is the bilayer width (there are two planes for any given r, except at $r = 0$). Although various models for $v_{\omega_i}(r)$ may be considered, a simple model assigns a volume for each atomic group in a chain to the point in space corresponding to its center of mass.[219] That is,

$$v_{\omega_i}(r) = \sum_s \delta(r - r_s(\omega_i)) v_s, \tag{4.2.29}$$

where δ is the delta function (providing the per-length units to $v_{\omega_i}(r)$), $r_s(\omega_i)$ is the radial position of group s in species i in microstate ω_i, v_s is the volume ascribed to group s (computed from density data or van der Waals radii), and the summation is over all the groups in the molecule of interest (implicitly dependent on i). Note that

$$\int_{r=0}^{r=R} v_{\omega_i}(r)\, dr = \int_{r=0}^{r=R} \sum_s \delta(r - r_s(\omega_i)) v_s dr = \sum_s v_s = v_{i,core}, \tag{4.2.30}$$

where R is the position of the micelle core-water interface.

Since the micelle is modeled as being incompressible, Eq. 4.2.1 holds. This equation can be rearranged as

$$n = \frac{\underline{V}}{\sum_i \alpha_i v_{i,core}}, \tag{4.2.31}$$

after substituting $n_i = \alpha_i n$. Inserting Eq. 4.2.31 into Eq. 4.2.28 then yields

$$\sum_i \alpha_i \sum_{\omega_i} p_{\omega_i} v_{\omega_i}(r) = \frac{A(r)}{\underline{V}} v_{avg}, \tag{4.2.32}$$

where $v_{avg} \equiv \sum_i \alpha_i v_{i,core}$ is the composition-weighted average of the core species volumes (a quantity fixed by the specification of micelle composition in the case of a strict surfactant head/tail assignment). Since the volumes of the three regular geometries are $\frac{4}{3}\pi R^3$ for a sphere, $\pi R^2 L$ for a cylinder, and $2LWR$ for a bilayer, the ratio of radially-dependent area to volume becomes

$$\frac{A(r)}{\underline{V}} = \frac{s r^{s-1}}{R^s}, \tag{4.2.33}$$

where s is a geometry-dependent shape constant, given by

$$s = \begin{cases} 3 \text{ for a sphere} \\ 2 \text{ for a cylinder} \\ 1 \text{ for a bilayer} \end{cases}. \tag{4.2.34}$$

Note that Eq. 4.2.33 has no dependence on L for a cylinder, or on L or W for a bilayer.

Inserting Eq. 4.2.33 into Eq. 4.2.32 yields a final form for the constant density constraint. Specifically,

$$\sum_i \alpha_i \sum_{\omega_i} p_{\omega_i} v_{\omega_i}(r) = \frac{s r^{s-1}}{R^s} v_{avg}. \tag{4.2.35}$$

We now proceed to minimize Eq. 4.2.26, subject to the constraints in Eqs. 4.2.27

and 4.2.35, using Lagrange multipliers. Specifically,

$$\delta\left[\begin{array}{l} n\sum_i \alpha_i \left[\sum_{\omega_i} p_{\omega_i}\left(\epsilon_{\omega_i}+k_BT\ln p_{\omega_i}\right)\right] \\ +\sum_i \left[n\alpha_i k_BT\left(\psi_i-1\right)\right]\left(-1+\sum_{\omega_i} p_{\omega_i}\right) \\ +\int_{r=0}^{r=R}\left[n\pi\left(r\right)\right]\left(-\frac{sr^{s-1}}{R^s}v_{avg}+\sum_i \alpha_i \sum_{\omega_i} p_{\omega_i} v_{\omega_i}\left(r\right)\right) \end{array}\right] = 0, \tag{4.2.36}$$

where $n\alpha_i k_BT\left(\psi_i-1\right)$ is a convenient Lagrange multiplier for the normalization constraints, and $n\pi\left(r\right)$ is a convenient Lagrange multiplier for the constant density constraints (a continuous set of constraints, requiring satisfaction at each point r). Carrying out the variation in Eq. 4.2.36 yields

$$n\sum_i \alpha_i \sum_{\omega_i} \delta p_{\omega_i}\left[\epsilon_{\omega_i}+k_BT\left(\psi_i+\ln p_{\omega_i}\right)+\int_{r=0}^{r=R}\pi\left(r\right)v_{\omega_i}\left(r\right)dr\right]=0. \tag{4.2.37}$$

Since the individual δp_{ω_i} may vary independently, the argument in brackets must be equal to zero. That is,

$$p_{\omega_i}=\frac{\exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\int_{r=0}^{r=R}\pi\left(r\right)v_{\omega_i}\left(r\right)dr\right)}{Z_i}, \tag{4.2.38}$$

where

$$Z_i\equiv\exp\left(\psi_i\right), \tag{4.2.39}$$

is the partition function for the central chain representing species i, and

$$\psi_i=\ln\sum_{\omega_i}\exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\int_{r=0}^{r=R}\pi\left(r\right)v_{\omega_i}\left(r\right)dr\right). \tag{4.2.40}$$

In practice, the continuous Lagrange multiplier function $\pi\left(r\right)$ is often replaced with a discrete set of multipliers $\{\pi_l\}$, where π_l is the Lagrange multiplier corresponding to application of the constant density constraint in a layer of finite thickness, indexed by l. In that case, the integral in Eq. 4.2.38 becomes

$$\int_{r=0}^{r=R}\pi\left(r\right)v_{\omega_i}\left(r\right)dr\approx\sum_l \pi_l v_{\omega_i,l}, \tag{4.2.41}$$

where $v_{\omega_i,l}$ is the volume of chain i in microstate ω_i found in layer l. The probability distribution function p_{ω_i} is then given by

$$p_{\omega_i} = \frac{\exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\sum_l \pi_l v_{\omega_i,l}\right)}{Z_i}, \tag{4.2.42}$$

where

$$Z_i = \sum_{\omega_i} \exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\sum_l \pi_l v_{\omega_i,l}\right). \tag{4.2.43}$$

The constant density constraint in Eq. 4.2.35 becomes

$$\sum_i \alpha_i \sum_{\omega_i} p_{\omega_i} v_{\omega_i,l} = \frac{s v_{avg}}{R^s} \int_{r=r_{0,l}}^{r=r_{1,l}} r^{s-1} dr, \tag{4.2.44}$$

where the integral bounds $r = r_{0,l}$ and $r = r_{1,l}$ specify the thickness of the layer. A layer-based approach typically involves layers of small widths $r_{1,l} - r_{0,l} \approx 1.5\mathring{A}$, or the length of a C-C bond. This represents a suitable trade-off between accuracy and performance in solving for the π_l values.

To carry out practical calculations, one needs to select a suitable model for each chain from which a representative sampling of the microstates ω_i can be performed, and ϵ_{ω_i} and $v_{\omega_i,l}$ values can be assigned. Initial guesses are provided for $\{\pi_l\}$ (typically setting these uniformly to zero), and a non-linear solver is used to find the values of $\{\pi_l\}$ which allow the fulfillment of Eq. 4.2.44 in all the layers of the micelle core. Once these Lagrange multipliers are determined, all the expectation values of quantities that are functions of the chain microstate can be computed as follows:

$$\langle X_{\omega_i} \rangle = \sum_{\omega_i} p_{\omega_i} X_{\omega_i}. \tag{4.2.45}$$

An effective non-linear solver that I have used to find $\{\pi_l\}$ involves the modified implementation of the Powell Hybrid method found in MINPACK.[156]

An additional subtle feature of the packing problem allows one to reduce the number of equations to be solved by 1. Since the layer volumes sum up to yield the fixed micelle core volume, only $L - 1$ $\{\pi_l\}$ values are independent. This can be seen

by manipulating Eq. 4.2.42, including setting $\pi_l = \widehat{\pi}_l + \pi'$, where π' is an arbitrary constant. We find that

$$\begin{aligned} p_{\omega_i} &= \frac{\exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\sum_l\left(\widehat{\pi}_l+\pi'\right)v_{\omega_i,l}\right)}{\sum_{\omega_i}\exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\sum_l\left(\widehat{\pi}_l+\pi'\right)v_{\omega_i,l}\right)} \\ &= \frac{\exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\sum_l\widehat{\pi}_l v_{\omega_i,l}\right)\exp\left(+\beta\pi' v_{i,core}\right)}{\sum_{\omega_i}\exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\sum_l\widehat{\pi}_l v_{\omega_i,l}\right)\exp\left(+\beta\pi' v_{i,core}\right)} \\ &= \frac{\exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\sum_l\widehat{\pi}_l v_{\omega_i,l}\right)}{\sum_{\omega_i}\exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\sum_l\widehat{\pi}_l v_{\omega_i,l}\right)}, \end{aligned} \tag{4.2.46}$$

because $\sum_l v_{\omega_i,l} = v_{i,core}$, and $\exp\left(+\beta\pi' v_{i,core}\right)$, as a constant, can be factored out of the summation and cancelled from the numerator and the denominator. From the standpoint of numerical solution of the constant density constraints, this probability distribution function is identical to Eq. 4.2.42. That is, the symbols $\{\pi_l\}$ or $\{\widehat{\pi}_l\}$ are functionally equivalent, indicating that the offset π' may be selected freely. Freedom of choice of the value for the constant π' allows setting any one particular π_l (but only one) to a fixed value, say 0, thereby demonstrating the reduction in the number of independent equations from L to $L-1$.

With Eq. 4.2.42 in hand, we can substitute it into Eq. 4.2.26 to obtain a final expression for the core Helmholtz free energy. Specifically,

$$\begin{aligned} \underline{F}_{core} &= n\sum_i \alpha_i \left[\sum_{\omega_i} p_{\omega_i}\left(\epsilon_{\omega_i} + k_B T \ln\frac{\exp\left(-\beta\epsilon_{\omega_i}\right)\exp\left(-\beta\sum_l \pi_l v_{\omega_i,l}\right)}{Z_i}\right)\right] \\ &= n\sum_i \alpha_i \left[\sum_{\omega_i} p_{\omega_i}\left(-k_B T\ln Z_i - \sum_l \pi_l v_{\omega_i,l}\right)\right], \end{aligned} \tag{4.2.47}$$

or

$$\underline{F}_{core} = n\sum_i \alpha_i \left[-k_B T\ln Z_i - \sum_l \pi_l \left\langle v_{\omega_i,l}\right\rangle\right] \tag{4.2.48}$$

where $\langle v_{\omega_i,l}\rangle$ is the expectation value for the volume of the central chain representing species i found in layer l of the micelle core – that is, averaged across all microstates ω_i.

4.2.4 The Packing Free Energy

As a free-energy contribution to the free energy of micellization, the packing free energy conceptually links the states of free, unconfined, unconstrained surfactant tails in an oil-like phase with the same tails packed within the confinement of the micelle core geometry. The Helmholtz free energy described by Eq. 4.2.48 in Section 4.2.3 represents the final state. It now remains to evaluate the Helmholtz free energy of the initial, free state, which I will denote as $\underline{F}_{free}$, and calculate as follows:

$$\underline{F}_{free} = n \sum_i \alpha_i \left[\sum_{\omega_i} p_{\omega_i,free} \left(\epsilon_{\omega_i} + k_B T \ln p_{\omega_i,free} \right) \right], \tag{4.2.49}$$

where a dependence on the micelle type, ξ (index suppressed), arises from the composition of ξ (i.e., the mole fractions $\{\alpha_i\}$) and its total aggregation number (i.e., n). We need to determine, as before, the functionality represented by $p_{\omega_i,free}$, the probability distribution function of a central chain representing species i in microstate ω_i. The MAXENT formalism can again be used, where the notion that each central chain has an average internal energy, $\langle \epsilon_{\omega_i} \rangle_{free}$, allows us to minimize $\underline{F}_{free,\xi}$ with respect to any constraints. Since the free environment represents one in which each chain is considered to be in an unperturbed state, and the unperturbed state (i.e., the isotropic, bulk state) is well-characterized by the use of an RIS chain model,[227] only a single constraint is required, related to normalization of $p_{\omega_i,free}$. That is,

$$\sum_{\omega_i} p_{\omega_i,free} = 1. \tag{4.2.50}$$

Following the derivations of Section 4.2.3, the minimization procedure yields

$$p_{\omega_i,free} = \frac{\exp\left(-\beta \epsilon_{\omega_i}\right)}{Z_{i,free}}, \tag{4.2.51}$$

where $Z_{i,free}$ is the partition function for the central chain representing species i in

the free state, defined as

$$Z_{i,free} \equiv \sum_{\omega_i} \exp\left(-\beta \epsilon_{\omega_i}\right). \tag{4.2.52}$$

Note that $p_{\omega_i,free}$ has no dependence on the micelle core type ξ, as expected for a free chain. Furthermore, there are no Lagrange multipliers to solve for! Substituting Eq. 4.2.51 into Eq. 4.2.49 yields

$$\underline{F}_{free} = n \sum_i \alpha_i \left(-k_B T \ln Z_{i,free}\right). \tag{4.2.53}$$

The packing free energy can now be computed as a difference between the final (core) and initial (free) states. Specifically,

$$\underline{F}_{pack} = \underline{F}_{core} - \underline{F}_{free}, \tag{4.2.54}$$

where substituting Eqs. 4.2.48 and 4.2.53 into Eq. 4.2.54 yields the following useful working expression:

$$\underline{F}_{pack} = n \sum_i \alpha_i \left[-k_B T \ln \frac{Z_i}{Z_{i,free}} - \sum_l \pi_l \left\langle v_{\omega_i,l} \right\rangle\right]. \tag{4.2.55}$$

In contrast to the individual core and free states, where the partition functions grow unbounded as the sampling increases, Eq. 4.2.55 crucially contains a ratio of partition functions, which will converge to a finite value. It is important to recognize this distinction between computing an expectation value and computing a partition function. Expectation values can be computed using Eq. 4.2.45 for any number of samples, and their numerical values will typically converge as the number of samples is increased. Partition functions will never converge, and their absolute calculation would rely on a full enumeration of all possible microstates (finite, due to quantum mechanical principles, but inaccessibly large). Fortunately, one is always interested in differences between states and the associated free energies which exhibit ratios of the type present in Eq. 4.2.55.

Self-consistent evaluation of $\underline{F}_{pack}$, and $Z_i/Z_{i,free}$ in particular, can be accom-

plished using several strategies. For longer chains, use of the Monte Carlo approach described in Appendix B is effective. For shorter chains, a more efficient strategy involves: (i) fully enumerating the conformations of each central chain (e.g., in the sense of specifying dihedral angles and using an RIS model to evaluate internal energies), and (ii) randomly sampling external orientations of the chain in space in a manner which samples the whole of phase space (see Appendix A for details – in the case of three-state RIS models, this is computationally feasible for no more than around 18 to 20 rotatable bonds). Subsequently, for the purposes of calculating Z_i, chains which are found to possess atom groups that lie outside the core volume are assigned an infinite internal energy (zero probability of occurrence). Otherwise, if all atom groups are within the core volume, the internal energy ϵ_{ω_i} determined from the chain model is used to calculate the energy-related Boltzmann factors. For the purposes of calculating $Z_{i,free}$, in contrast, the internal energy ϵ_{ω_i} determined from the chain model is *always* used in calculating the energy-related Boltzmann factors, irrespective of chain conformation. That is, an infinite internal energy penalty is never applied, since there is no concept of a confining geometry in the free state. In this manner, a consistent set of chain samples is used to evaluate the energies of both environments, in spite of the fact that a large number of chains will have a zero probability of occurring in the packed state.

4.3 Rotational Isomeric State Models

4.3.1 Background

A chain molecule with N atoms has $3N$ positional (i.e., configurational) degrees of freedom within a fixed frame-of-reference. These can be viewed as 3 translational and 3 rotational degrees of freedom (the chain's external orientation), and $3N - 6$ internal degrees of freedom (the chain's internal conformation). Any particular chain conformation can be fully described either through the use of internal coordinates (i.e., bond lengths, angles, and dihedrals – also referred to as Z-matrix coordinates[228])

or Cartesian coordinates (see Appendix A for conversion from internal to Cartesian coordinates using a two-step placement algorithm, and Appendix D for conversion from Cartesian to internal coordinates through the definitions of bond length, angle, and dihedral). Internal coordinates are especially convenient when modeling the bonded contribution to the potential energy of the chain, which typically involves bond-stretching, angle-bending, and bond-torsion contributions.[188]

Given a model for the potential energy of a chain as a function of its conformation (see Appendix D for the example of the OPLS-AA force field[188]), one can construct a (high-dimensional) potential energy surface for a chain as a function of its internal coordinates, from which the location of local energy minima can be probed. For linear alkanes, these minima turn out to be quite narrow and deep, corresponding to conformations which possess at or near-equilibrium values for all bond lengths and angles, and have individual bond dihedrals which lie within one of three narrow ranges of values,[216] discussed in more detail below. Similar observations have been made with regards to other types of chains, although the specifics of the various bond lengths, angles, and dihedrals which induce a local minimum varies.[229]

In Section 4.2, it was demonstrated that the probability of a chain configuration in the context of packing (Eq. 4.2.42) depends on two Boltzmann factors: one due to the chain internal energy (i.e., its configurational potential energy), and one due to the distribution of the chain volume in space within a pressure field which serves to maintain constant density constraints. For chains with high internal energy, the first Boltzmann factor will yield a very low value in comparison to a chain possessing a conformation corresponding to a local internal energy minimum state. For our purposes, we will assume that the pressure-volume term cannot counterbalance a high internal energy penalty. That is, for the purpose of packing, we will consider only chain conformations exhibiting a local internal energy minimum. Since it appears that such conformations typically have equilibrium bond lengths and angles,[216] we will always use these values, such that specification of the most-likely chain conformations will only require knowledge of the dependence of the chain internal energy on the remaining, unspecified bond dihedral angles.

We now identify three important points regarding the sets of dihedral angles that lead to local minimization of the chain internal energy for linear alkanes. Similar considerations arise when dealing with other types of chains (e.g., fluorinated chains or polymers such as polystyrene), although specific values for the dihedral angles and associated internal energies will differ.[229]

The first point is that each *C-C-C-C* dihedral angle in a linear alkane chain will be found primarily in one of three states, conventionally labeled *trans* (t, such that $\phi_t = 180°$), *gauche*$^+$ (g^+, such that $\phi_{g^+} = 180° + \widehat{\phi}$), or *gauche*$^-$ (g^-, such that $\phi_{g^-} = 180° - \widehat{\phi}$), where the *gauche* states are offset from the *trans* state by some offset angle $\widehat{\phi}$, where $\widehat{\phi}$ falls within the range of 110°–120°[230] and is typically taken to be 120° for both *gauche* states (see Hoeve,[231] Patterson and Flory,[230] and Mattice and Santiago[215]). The two *gauche* states have been shown experimentally to be equipotential with each other, but the two possess a higher energy than the *trans* state, with an energy difference of approximately 500 ± 100 cal/mol.[230,231]

The second point is that the terminal methyl groups of a linear alkane chain exhibit a three-fold symmetry in the rotational profile of the *H-C-C-C* (the first chain *C-C* bond) or *C-C-C-H* (the last chain *C-C* bond) dihedrals. Here, the angles yielding energy minima are also taken to be $\phi_{g^-} = 180° - 120° = 60°$, $\phi_t = 180°$, and $\phi_{g^+} = 180° + 120° = 300°$, but there is no difference in energy between these states.[216]

The third point is that, although a change to a dihedral angle in one part of a molecule could potentially displace distant atoms in a manner which causes highly unfavorable excluded-volume interactions (i.e., steric overlap, which can be a highly non-local effect), it appears that this effect is minor in terms of calculating the chain internal energy.[216] However, the dihedral states of nearest neighbor bonds *do*, in certain combinations, cause such unfavorable steric interactions for hydrogen atoms in methylene groups separated by four consecutive bonds. Specifically, these unfavorable combinations are identified as pairings of *gauche* bonds of opposite sense (i.e., either g^+g^- or g^-g^+ sequential pairs). The smallest chain in which this effect must be considered is n-pentane,[216] as illustrated in Fig. 4-1, and this is therefore referred to as the "pentane effect".[227] For linear alkanes, the characterization of pairs of neighboring

bond states appears sufficient for the purpose of reconstructing the internal energy of the chain.[213]

The Rotational Isomeric State (RIS) model of Flory affords a compact mathematical representation of the three points discussed above, in the form of statistical weight matrices corresponding to the torsional state of each C-C bond, and, when relevant, its immediate predecessor.[213] These matrices, usually labeled $\mathbf{U}_b$, for bond $b = 1, ..., B$ with B the total number of C-C bonds in the chain, have components $U_{b,ij}$ (i.e., for row i, column j). These components are Boltzmann factors of the form

$$U_{b,ij} = \exp\left(-\frac{E_{b,ij}}{k_B T}\right), \tag{4.3.1}$$

where $E_{b,ij}$ is a temperature-independent internal energy, k_B is the Boltzmann constant, and T is the absolute system temperature.

For bonds 1 and B, $\mathbf{U}_1 = \mathbf{U}_B = \mathbf{I}_{3\times 3}$, the identity matrix, reflecting the equipotential torsional nature of these bonds due to the attached methyl groups.

For bond 2, whose immediate predecessor (bond 1) is symmetric, $\mathbf{U}_2$ is also symmetric, but has weightings, σ, which are smaller than unity for the gauche states:[216]

$$\mathbf{U}_2 = \mathbf{diag}\,(1, \sigma, \sigma) \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & \sigma \end{bmatrix}. \tag{4.3.2}$$

For bonds $3, ..., B-1$, the states of each bond and its predecessor are taken into account as follows:[215]

$$\mathbf{U}_b = \begin{array}{c} \\ t \\ g^+ \\ g^- \end{array} \begin{array}{c} \begin{array}{ccc} t & g^+ & g^- \end{array} \\ \begin{bmatrix} \tau & \sigma & \sigma \\ 1 & \sigma\psi & \sigma\omega \\ 1 & \sigma\omega & \sigma\psi \end{bmatrix} \end{array}, \tag{4.3.3}$$

where the rows correspond to the states of the dihedral centered on the predecessor bond $(b-1)$, the columns correspond to the states of the dihedral centered on the current bond, b, as shown, the term σ is the three-bond (centered on bond b) *gauche*

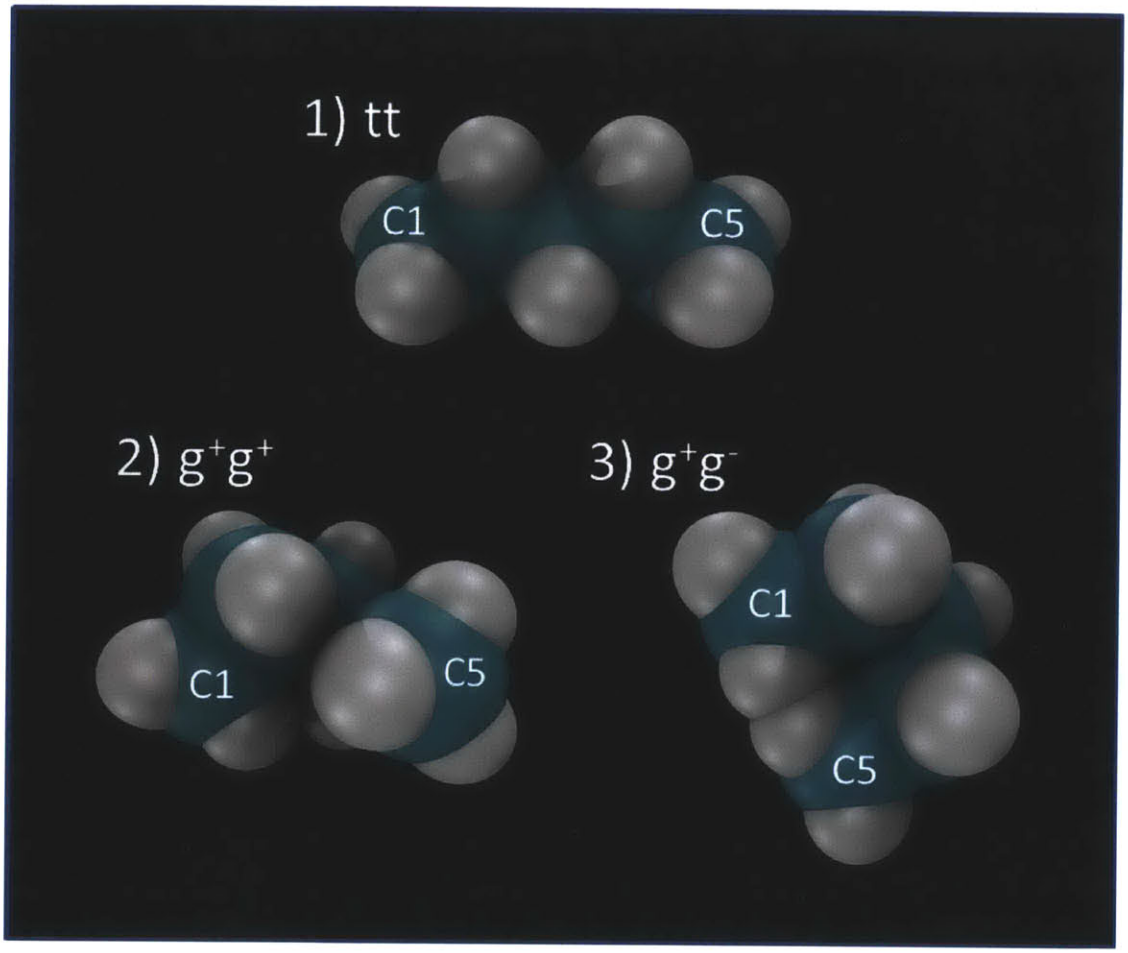


Figure 4-1: Three illustrative molecular simulations snapshots of pentane in various internal conformations of importance to the Rotational Isomeric State (RIS) theory. The snapshots represent three of the nine possibilities for the pairs of dihedral states characterizing the interior C1-C2-C3-C4 and C2-C3-C4-C5 dihedrals: 1) an all-*trans* conformation, representing the minimum-energy state, 2) a gauche$^+$/gauche$^+$ conformation, representing a higher energy state (observe that the hydrogens are brought into closer contact), and 3) a gauche$^+$/gauche$^-$ conformation, representing the highest energy chain conformation allowed in the RIS model, due to significant overlap of the hydrogens attached to the C1 and C5 carbons. (Color code: white – hydrogen, teal – carbon.)

weighting, and the terms τ, ψ, and ω are four-bond (centered on the bond pair $b-1$ and b) weightings that account for interactions of the sort pictured in Figure 4-1. The order (t, g^+, g^-) in the rows and columns is by convention,[216,217] and will not be shown explicitly again. For linear alkanes, there is no ambiguity about which dihedral one is referring to when specifying a dihedral centered on a bond, since each interior *C*-*C* bond of a chain corresponds to only one *C*-*C*-*C*-*C* dihedral. In branched chains, however, this is not the case near the branch nodes. Therefore, I will be careful to specify the precise dihedral meant in each situation.

The weightings presented in Eq. 4.3.3 are in their most general form for a linear alkane. Setting $\tau = 1$ and $\psi = 1$ recovers the form used by Abe et al.,[216] and the additional choice of $\omega = 0$ yields what our group has used in the past.[119] Specific values for the weightings σ, τ, ψ, and ω are selected to reproduce the statistical behavior of unperturbed chains in the bulk.[215] See Section 4.3.2 below for a full specification of these values for linear alkanes.

The matrix $\mathbf{U}_b$ in Eq. 4.3.3 can also be formulated in an equivalent form as follows:

$$\mathbf{U}_b = \mathbf{V}_b \mathbf{D}_b, \tag{4.3.4}$$

$$\mathbf{D}_b = \mathbf{diag}\,(1, \sigma, \sigma)\,, \tag{4.3.5}$$

$$\begin{aligned} \mathbf{V}_b &= \begin{bmatrix} w\,(t,t) & w\,(t,g^+) & w\,(t,g^-) \\ w\,(g^+,t) & w\,(g^+,g^+) & w\,(g^+,g^-) \\ w\,(g^-,t) & w\,(g^-,g^+) & w\,(g^-,g^-) \end{bmatrix} \\ &= \begin{bmatrix} \tau & 1 & 1 \\ 1 & \psi & \omega \\ 1 & \omega & \psi \end{bmatrix}, \end{aligned} \tag{4.3.6}$$

where $\mathbf{D}_b$ encodes the weightings due to the three-bond interactions centered on bond b (in the order of t, g^+, g^-, corresponding to the RIS states for the dihedral centered on bond b), and $\mathbf{V}_b$ encodes the weightings due to the four-bond interactions centered

on the pair of bonds $b-1$ and b, with these weightings symbolically represented by a function w, which assigns a weighting to pairs of consecutive bond states as shown. For linear alkanes, as in Eq. 4.3.6, the first argument of w is the state of the dihedral centered on bond $b-1$ and the second argument is the state of the dihedral centered on bond b, which happens to correspond in every case to the row and column states presented in Eq. 4.3.3. For branched alkanes, this will not always be the case. Note that $\mathbf{U}_2$ can also be formulated in this manner, if $\mathbf{V}_2$ is set equal to the identity matrix, such that $\mathbf{U}_2 = \mathbf{D}_2$. This simply reflects the fact that only three-bond interactions are considered for this bond.

We will find that the formulation of $\mathbf{U}_b$ in Eq. 4.3.4 is especially convenient in terms of understanding the matrices required for the modeling of branched alkanes (see Section 4.3.3 below).

4.3.2 Linear Chains

In my linear alkane modeling work (see Chapter 5), I use the values prescribed by Mattice and Santiago as follows:[215] $\tau = 1$, $\psi = 1$, $E_\sigma = 500\ cal/mol$, and $E_\omega = 2000\ cal/mol$ (corresponding to $\sigma = \exp(-E_\sigma/RT) = 0.43$ and $\omega = \exp(-E_\omega/RT) = 0.0342$ for $R = 1.987\ cal/mol\,K$ and $T = 298.15\ K$). The bond lengths and angles of all the bonds are set at $l_{CC} = 1.53\mathring{A}$ and $\theta_{CCC} = 112°$. The *gauche* angle offset (from the *trans* angle of $\phi_t = 180°$) is $\widehat{\phi} = 120°$, such that $\phi_{g^+} = 180° + 120° = 300°$ and $\phi_{g^-} = 180° - 120° = 60°$. For the purpose of calculating the distribution of chain volume in space, methylene (CH_2) groups and the terminal methyl (CH_3) group are treated as united atom, with $v_{CH_2} = 27\mathring{A}^3$ and $v_{CH_3} = 54\mathring{A}^3$.

4.3.3 Branched Chains

In my branched alkane modeling work (see Chapter 6), I use the approach of Mattice[217] in considering chains which exhibit branching, by enumerating all the three-bond and four-bond interactions that occur when each bond in the chain is rotated. Although this specific approach can be applied to branched chains possessing an

arbitrary number of branches at least two bonds apart, or even to chains with tetra-functionality at a given carbon atom, in this thesis I focus mainly on systems with a single branch. Therefore, I will present the derivation of Mattice's $\mathbf{U}_b$ matrices for only this last case (that of single branching), using the decomposition shown in Eqs. 4.3.4-4.3.5 and several diagrams to illustrate the geometry of the conformations under consideration. Fig. 4-2 shows the general labeling scheme that I use in the discussion which follows. For bonds far from the node of interest, a linear alkane chain model is appropriate (specifically, bonds A3-A2, B2-B3, and C2-C3 in Fig. 4-2, and others not shown which are even further away from the nodal carbon N, are considered to qualify for this treatment[217]). In Fig. 4-2, the specific stereochemistry of the side chain is left intentionally ambiguous – this position will be specified in practice by reporting an offset dihedral angle, $\phi_{offset} = \phi_{A2\text{-}A1\text{-}N\text{-}C1} - \phi_{A2\text{-}A1\text{-}N\text{-}B1}$, of either $+120°$ or $-120°$, reflecting an assumed perfectly tetrahedral nature of the nodal carbon N.

In Mattice's approach,[217] the $\mathbf{U}_b$ matrices are developed systematically by first considering the rotational states and torsional energies of the main chain bonds in the presence of the hindrance introduced by the first carbon of the side chain (in three-bond interactions) and its hydrogens (in four-bond interactions), once the first bond connecting to the nodal carbon is added to the chain. Then, once the main chain is fully specified, the side chain bonds are considered, with the new requirement of a three-bond RIS matrix (a "three-layer" matrix) not typically found in linear chains, representing the states of the three neighboring bonds sharing a connection with the nodal carbon. Using the specific labels in Fig. 4-2 as a reference, this approach implies that we build up the main A chain as we would build up a linear chain, until the A1-N bond is introduced, whose torsional states reflect placement of both B1 and C1 in space. One then continues down the B chain, considering bonds N-B1 and B1-B2 in the presence of the placed C1. As already mentioned, bond B2-B3 and bonds further down the main B chain are considered equivalent to bonds in a linear alkane. Finally, once the main chain is constructed in this manner, one considers the rotational states associated with the N-C1 bond (giving rise to the "three-layer" matrix due to its having two neighbors: A1-N and N-B1). Bond C1-C2 is the final bond affected by

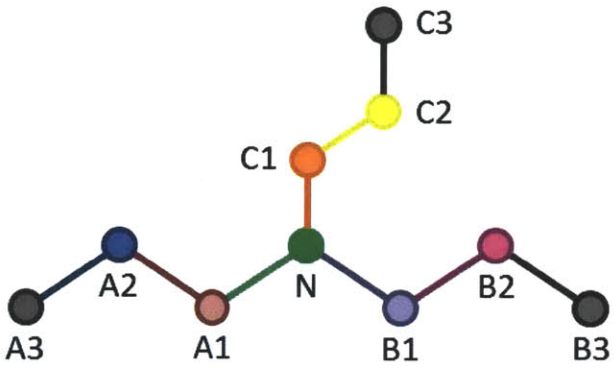


Figure 4-2: Labeling used in the text and subsequent figures in the discussion of the RIS matrix development for branched alkyl chain packing. Each filled circle represents a carbon-based united atom group (e.g., a CH_2 group). The label "N" identifies the CH nodal group, while all the other labels signify a branch letter and united atom group number representing position in the chain relative to group N. The main chain proceeds down branch A (descending numbers), through N, and up branch B (ascending numbers). Here, the main chain is represented in an all-*trans* configuration. The side chain is identified as branch C. The definition of main and side chains is such that branch B contains a number of united atom groups equal to or greater in number than that in branch C. Finally, note that the side chain is depicted here in an ambiguous manner with regards to the stereochemistry of N. In subsequent figures, the A2-A1-N-C1 dihedral relative to the A2-A1-N-B1 dihedral will always be specified. (Color code: colors other than black indicate atom groups that participate in three-bond interactions of relevance to branched packing near the nodal carbon N; atom groups in black participate only through four-bond interactions, and the coloring is otherwise arbitrary and meant to serve only to visually distinguish between groups.)

the node, after which the linear alkane model can again be employed. Note that, if branch C is of length 1, the rotational states of N-C1 need not be considered (due to the symmetry of a methyl group). If branch C is of length 2, the rotational states of C1-C2 need not be considered, and so forth. Similar considerations arise for short A or B branches.

I now consider the first bond affected by the branch at N: the A1-N bond, whose rotational states involve placement of B1 and C1. The presence of the C1 carbon, in contrast to a linear alkane, will ensure that there is always a *gauche* interaction of some type when considering the three-bond interactions centered on A1-N (i.e., the interactions involving A2 and B1 and A2 and C1). As seen in Fig. 4-3, the precise nature of the hindrance introduced by the side chain will depend on the stereochemistry.

For case (1), where $\phi_{offset} = +120°$, it follows that

$$\begin{aligned} \mathbf{D}^{(1)}_{A1-N} &= diag\,(1\cdot\sigma, \sigma\cdot\sigma, \sigma\cdot 1) \\ &= diag\,(\sigma, \sigma^2, \sigma)\,, \end{aligned} \tag{4.3.7}$$

corresponding to the three-bond (A2-B1,A2-C1) interactions (t, g^+), (g^+, g^-), and (g^-, t) experienced as the A1-N bond is rotated through its three allowable RIS states (the states are for $\phi_{A2-A1-N-B1}$ and $\phi_{A2-A1-N-C1}$). In Eq. 4.3.7, the three-bond interaction Boltzmann factors are considered multiplicative (i.e., the energies are additive, such that when both the main chain and the side chain are in *gauche* states, as a result of rotation of the A1-N bond, the energy is $2E_\sigma$, which yields a weighting of σ^2).

The four-bond interactions are those centered on A1-N and its predecessor bond A2-A1, due to interactions between the hydrogens on A3 and either B1 or C1 (refer to Fig. 4-1). For case (1), in order to construct $\mathbf{V}^{(1)}_{A1-N}$, I consider the positions of B1 and C1 relative to A3 as a result of rotations of the two central bonds. For example, when bond A2-A1 is t (i.e., $\phi_{A3-A2-A1-N} = 180°$) and bond A1-N is t (i.e., $\phi_{A2-A1-N-B1} = 180°$), this implies that C1 is effectively in a g^+ position (i.e., $\phi_{A2-A1-N-C1} = 300°$).

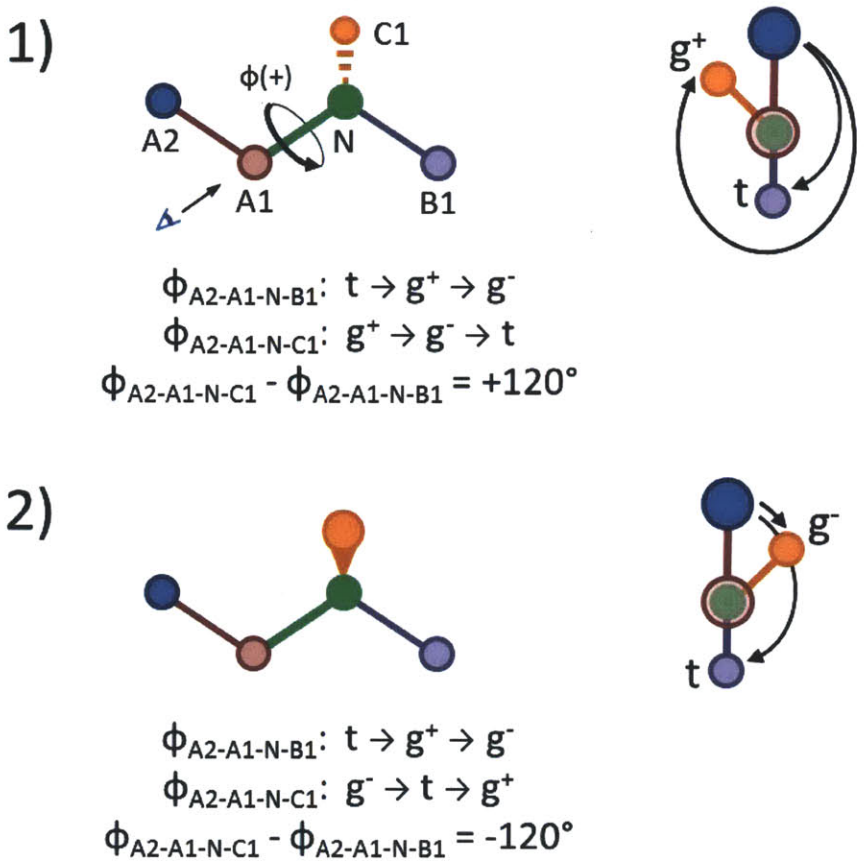


Figure 4-3: Diagram illustrating the atoms adjacent to, or participating in, the rotation about the A1-N bond in a branched chain. In 1), the branch is located in the g^+ position relative to the main (A-B) chain in a t configuration, while in 2), the position is g^- (refer to the diagrams on the right, which show the layout of the bonds as seen when looking down the A1-N bond, depicted by the eye symbol and arrow on the left). As the main chain rotates through the states indicated for $\phi_{A2-A1-N-B1}$, the side chain rotates in an offset manner, as indicated for $\phi_{A2-A1-N-C1}$.

If one examines Eq. 4.3.6 and assumes that the same weightings hold here, one must then extract the $w(t,t)$ and $w(t,g^+)$ entries to model this set of states, multiplying them (i.e., summing their corresponding energies) to obtain the weighting appropriate for the (1,1) position of $\mathbf{V}^{(1)}_{A1-N}$. Implementing this approach as one changes the state of A2-A1 (rows of $\mathbf{V}^{(1)}_{A1-N}$: $\phi_{A3-A2-A1-N} = (t, g^+, g^-)$) and A1-N (columns of $\mathbf{V}^{(1)}_{A1-N}$: $\phi_{A2-A1-N-B1} = (t, g^+, g^-)$) systematically, yields

$$\begin{aligned}
\mathbf{V}^{(1)}_{A1-N} &= \begin{bmatrix} w(t,t)\,w(t,g^+) & w(t,g^+)\,w(t,g^-) & w(t,g^-)\,w(t,t) \\ w(g^+,t)\,w(g^+,g^+) & w(g^+,g^+)\,w(g^+,g^-) & w(g^+,g^-)\,w(g^+,t) \\ w(g^-,t)\,w(g^-,g^+) & w(g^-,g^+)\,w(g^-,g^-) & w(g^-,g^-)\,w(g^-,t) \end{bmatrix} \\
&= \begin{bmatrix} \tau & 1 & \tau \\ \psi & \psi\omega & \omega \\ \omega & \omega\psi & \psi \end{bmatrix}, \qquad (4.3.8)
\end{aligned}$$

where each entry of $\mathbf{V}^{(1)}_{A1-N}$ is the total statistical weighting due to four-bond interactions, and the products of w correspond to

$$w\left(\phi_{A3-A2-A1-N}, \phi_{A2-A1-N-B1}\right) w\left(\phi_{A3-A2-A1-N}, \phi_{A2-A1-N-C1}\right). \qquad (4.3.9)$$

Applying Eq. 4.3.4 yields

$$\begin{aligned}
\mathbf{U}^{(1)}_{A1-N} &= \mathbf{V}^{(1)}_{A1-N}\mathbf{D}^{(1)}_{A1-N} \\
&= \sigma \begin{bmatrix} \tau & \sigma & \tau \\ \psi & \psi\omega\sigma & \omega \\ \omega & \omega\psi\sigma & \psi \end{bmatrix} \qquad (4.3.10)
\end{aligned}$$

Note that Eq. 4.3.10 is equivalent to Mattice's Eq. (21),[217] recognizing that $\mathbf{U}^{(1)}_{A1-N} = {}_1U_{n_1}$ (in Mattice's notation) and setting $\tau = 1$.

For case (2), where $\phi_{offset} = -120°$, one has instead

$$\mathbf{D}^{(2)}_{A1-N} = diag\left(\sigma, \sigma, \sigma^2\right), \qquad (4.3.11)$$

$$\mathbf{V}^{(2)}_{A1-N} = \begin{bmatrix} w(t,t)\,w(t,g^-) & w(t,g^+)\,w(t,t) & w(t,g^-)\,w(t,g^+) \\ w(g^+,t)\,w(g^+,g^-) & w(g^+,g^+)\,w(g^+,t) & w(g^+,g^-)\,w(g^+,g^+) \\ w(g^-,t)\,w(g^-,g^-) & w(g^-,g^+)\,w(g^-,t) & w(g^-,g^-)\,w(g^-,g^+) \end{bmatrix}$$

$$= \begin{bmatrix} \tau & \tau & 1 \\ \omega & \psi & \omega\psi \\ \psi & \omega & \psi\omega \end{bmatrix}, \tag{4.3.12}$$

and

$$\mathbf{U}^{(2)}_{A1-N} = \sigma \begin{bmatrix} \tau & \tau & \sigma \\ \omega & \psi & \omega\psi\sigma \\ \psi & \omega & \psi\omega\sigma \end{bmatrix}. \tag{4.3.13}$$

Note that Eq. 4.3.13 is equivalent to Mattice's Eq. (4),[217] recognizing that $\mathbf{U}^{(2)}_{A1-N} = {}_1U_{n_1}$ (in Mattice's notation) and setting $\tau = 1$.

The second bond affected by the branch at N is the N-B1 bond. The presence of the C1 carbon, as with the A1-N case, always ensures at least one *gauche* interaction of some type when considering the three-bond interactions centered on N-B1 (i.e., the interactions between A1 and B2 and between C1 and B2). The two possible stereoisomers are presented in Fig. 4-4, corresponding to the same cases considered in Fig. 4-3.

For case (1) ($\phi_{offset} = +120°$), it follows that

$$\begin{aligned} \mathbf{D}^{(1)}_{N-B1} &= diag\,(1\cdot\sigma, \sigma\cdot 1, \sigma\cdot\sigma) \\ &= diag\,(\sigma, \sigma, \sigma^2), \end{aligned} \tag{4.3.14}$$

corresponding to the three-bond (A1-B2,C1-B2) interactions (t, g^-), (g^+, t), and (g^-, g^+) experienced as the N-B1 bond is rotated through its three allowable RIS states (the states are for $\phi_{A1-N-B1-B2}$ and $\phi_{C1-N-B1-B2}$). As before, the three-bond interaction Boltzmann factors are considered multiplicative.

The four-bond interactions are those centered on N-B1 and its predecessor bond A1-N, and due to interactions between the hydrogens on A2 and B2. It is important

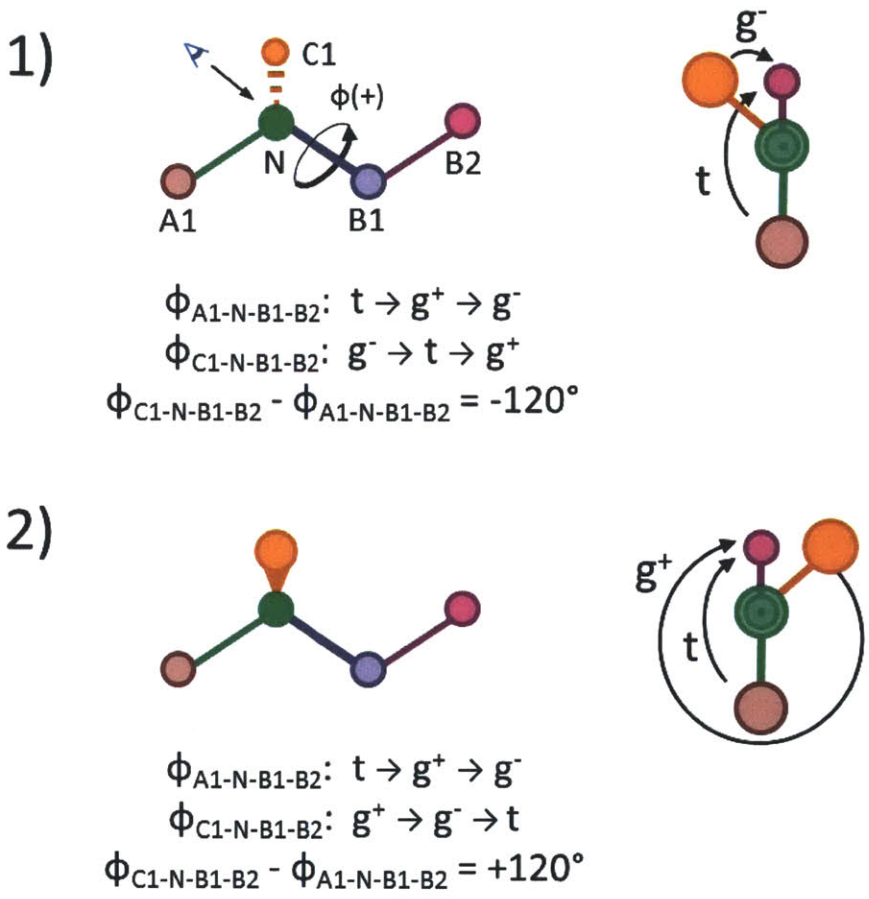


Figure 4-4: Diagram illustrating the atoms adjacent to, or participating in, the rotation about the N-B1 bond in a branched chain. Cases 1) and 2) correspond to the same cases considered in Fig. 4-3. In 1), the side chain is situated in a g^- position when considering the C1-N-B1-B2 dihedral. In 2), the side chain is situated in a g^+ position. As the main chain rotates through the states indicated for $\phi_{A1-N-B1-B2}$, the side chain rotates in an offset manner, as indicated for $\phi_{C1-N-B1-B2}$. (Refer to the diagrams on the right for examples, which show the layout of the bonds as seen when looking down the N-B1 bond, depicted by the eye symbol and arrow on the left.)

to note that, at this stage in building up the chain, there is no atom group C2 present. Therefore, this case is exactly identical to the case of a linear alkane. Therefore, it follows that

$$\begin{aligned} \mathbf{V}^{(1)}_{N-B1} &= \begin{bmatrix} w(t,t) & w(t,g^+) & w(t,g^-) \\ w(g^+,t) & w(g^+,g^+) & w(g^+,g^-) \\ w(g^-,t) & w(g^-,g^+) & w(g^-,g^-) \end{bmatrix} \\ &= \begin{bmatrix} \tau & 1 & 1 \\ 1 & \psi & \omega \\ 1 & \omega & \psi \end{bmatrix}, \end{aligned} \tag{4.3.15}$$

where the rows of $\mathbf{V}^{(1)}_{N-B1}$ correspond to $\phi_{A2-A1-N-B1} = (t, g^+, g^-)$, and the columns of $\mathbf{V}^{(1)}_{N-B1}$ correspond to $\phi_{A1-N-B1-B2} = (t, g^+, g^-)$, with the total statistical weightings

$$w\left(\phi_{A2-A1-N-B1}, \phi_{A1-N-B1-B2}\right). \tag{4.3.16}$$

Applying Eq. 4.3.4 yields

$$\begin{aligned} \mathbf{U}^{(1)}_{N-B1} &= \mathbf{V}^{(1)}_{N-B1}\mathbf{D}^{(1)}_{N-B1} \\ &= \sigma \begin{bmatrix} \tau & 1 & \sigma \\ 1 & \psi & \omega\sigma \\ 1 & \omega & \psi\sigma \end{bmatrix} \end{aligned} \tag{4.3.17}$$

Note that Eq. 4.3.17 is equivalent to Mattice's Eq. (22),[217] recognizing that $\mathbf{U}^{(1)}_{N-B1} = {}_2U_1$ (in Mattice's notation) and setting $\tau = 1$.

For case (2) ($\phi_{offset} = -120°$), it follows that

$$\mathbf{D}^{(2)}_{N-B1} = diag\left(\sigma, \sigma^2, \sigma\right), \tag{4.3.18}$$

$$\mathbf{V}^{(2)}_{N-B1} = \begin{bmatrix} \tau & 1 & 1 \\ 1 & \psi & \omega \\ 1 & \omega & \psi \end{bmatrix}, \tag{4.3.19}$$

and

$$\mathbf{U}^{(2)}_{N-B1} = \sigma \begin{bmatrix} \tau & \sigma & 1 \\ 1 & \psi\sigma & \omega \\ 1 & \omega\sigma & \psi \end{bmatrix}, \tag{4.3.20}$$

where $\mathbf{V}^{(2)}_{N-B1} = \mathbf{V}^{(1)}_{N-B1}$, because C2 has not yet been placed.

Note that Eq. 4.3.20 is equivalent to Mattice's Eq. (7),[217] recognizing that $\mathbf{U}^{(2)}_{N-B1} = {}_2U_1$ (in Mattice's notation) and setting $\tau = 1$.

The third bond affected by the branch at N is the B1–B2 bond. Although the C1 carbon no longer impacts three-bond interactions, it does participate in four-bond interactions. The two possible stereoisomers are presented in Fig. 4-5, corresponding to the same cases considered in Figs. 4-3 and 4-4.

For case (1) ($\phi_{offset} = +120°$), it follows that

$$\mathbf{D}^{(1)}_{B1-B2} = diag\,(1, \sigma, \sigma)\,, \tag{4.3.21}$$

corresponding to the three-bond (N-B2) interactions (t), (g^+), and (g^-) experienced as the B1-B2 bond is rotated through its three allowable RIS states (the states are for $\phi_{N-B1-B2-B3}$). This is identical to the linear alkane case, as expressed in Eq. 4.3.5.

The four-bond interactions are those centered on B1-B2 and its predecessor bond N-B1 and due to interactions between the hydrogens of A1 and C1 and those of B3. If one considers the rotational states of the N-B1 bond, depicted in Fig. 4-4, and summarized on the right in Fig. 4-5, one can construct $\mathbf{V}^{(1)}_{B1-B2}$ from the appropriate weightings. Specifically,

$$\begin{aligned} \mathbf{V}^{(1)}_{B1-B2} &= \begin{bmatrix} w\,(t,t)\,w\,(g^-,t) & w\,(t,g^+)\,w\,(g^-,g^+) & w\,(t,g^-)\,w\,(g^-,g^-) \\ w\,(g^+,t)\,w\,(t,t) & w\,(g^+,g^+)\,w\,(t,g^+) & w\,(g^+,g^-)\,w\,(t,g^-) \\ w\,(g^-,t)\,w\,(g^+,t) & w\,(g^-,g^+)\,w\,(g^+,g^+) & w\,(g^-,g^-)\,w\,(g^+,g^-) \end{bmatrix} \\ &= \begin{bmatrix} \tau & \omega & \psi \\ \tau & \psi & \omega \\ 1 & \omega\psi & \psi\omega \end{bmatrix}. \end{aligned} \tag{4.3.22}$$

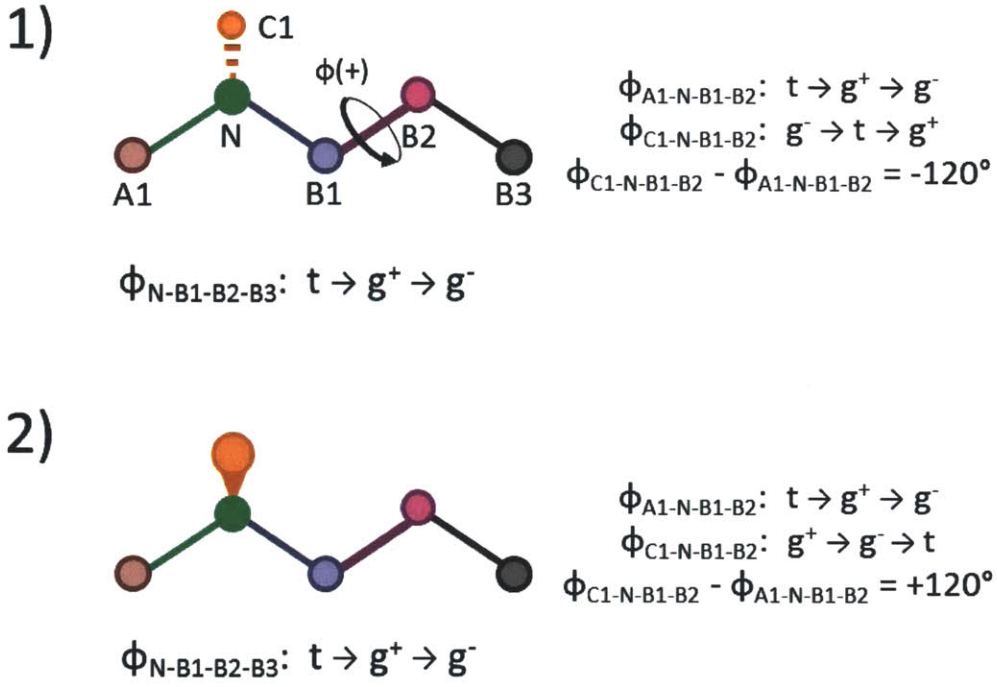


Figure 4-5: Diagram illustrating the atoms adjacent to, or participating in, the rotations about the B1-B2 bond in a branched chain, and the distant atoms (A1 and C1) participating in four-bond interactions with B3. Cases 1) and 2) correspond to the same cases considered in Figs. 4-3 and 4-4. For the three-bond interactions, only $\phi_{N-B1-B2-B3}$ is relevant (identical to a linear alkane case), while the four-bond interactions must consider the position of the distant atoms, as specified through $\phi_{A1-N-B1-B2}$ and $\phi_{C1-N-B1-B2}$.

where the rows of $\mathbf{V}^{(1)}_{B1-B2}$ correspond to $\phi_{A1-N-B1-B2} = (t, g^+, g^-)$, and the columns of $\mathbf{V}^{(1)}_{B1-B2}$ correspond to $\phi_{N-B1-B2-B3} = (t, g^+, g^-)$, with the total statistical weightings

$$w\left(\phi_{A1-N-B1-B2}, \phi_{N-B1-B2-B3}\right) w\left(\phi_{C1-N-B1-B2}, \phi_{N-B1-B2-B3}\right). \tag{4.3.23}$$

These are the weightings due to the interactions between A1 and B3 multiplied by the weightings due to the interactions between C1 and B3.

Applying Eq. 4.3.4 yields

$$\begin{aligned} \mathbf{U}^{(1)}_{B1-B2} &= \mathbf{V}^{(1)}_{B1-B2}\mathbf{D}^{(1)}_{B1-B2} \\ &= \begin{bmatrix} \tau & \omega\sigma & \psi\sigma \\ \tau & \psi\sigma & \omega\sigma \\ 1 & \omega\psi\sigma & \psi\omega\sigma \end{bmatrix}. \end{aligned} \tag{4.3.24}$$

Note that Eq. 4.3.24 is equivalent to Mattice's Eq. (23),[217] recognizing that $\mathbf{U}^{(1)}_{B1-B2} = {}_2U_2$ (in Mattice's notation) and setting $\tau = 1$.

For case (2) ($\phi_{offset} = -120°$), it follows that

$$\mathbf{D}^{(2)}_{B1-B2} = diag\left(1, \sigma, \sigma\right), \tag{4.3.25}$$

$$\begin{aligned} \mathbf{V}^{(2)}_{B1-B2} &= \begin{bmatrix} w(t,t)\,w(g^+,t) & w(t,g^+)\,w(g^+,g^+) & w(t,g^-)\,w(g^+,g^-) \\ w(g^+,t)\,w(g^-,t) & w(g^+,g^+)\,w(g^-,g^+) & w(g^+,g^-)\,w(g^-,g^-) \\ w(g^-,t)\,w(t,t) & w(g^-,g^+)\,w(t,g^+) & w(g^-,g^-)\,w(t,g^-) \end{bmatrix} \\ &= \begin{bmatrix} \tau & \psi & \omega \\ 1 & \psi\omega & \omega\psi \\ \tau & \omega & \psi \end{bmatrix}, \end{aligned} \tag{4.3.26}$$

and

$$\mathbf{U}^{(2)}_{B1-B2} = \begin{bmatrix} \tau & \psi\sigma & \omega\sigma \\ 1 & \psi\omega\sigma & \omega\psi\sigma \\ \tau & \omega\sigma & \psi\sigma \end{bmatrix}, \tag{4.3.27}$$

where $\mathbf{D}^{(2)}_{B1-B2} = \mathbf{D}^{(1)}_{B1-B2}$, because C1 does participate in the three-bond interactions in this case.

Note that Eq. 4.3.27 is equivalent to Mattice's Eq. (17),[217] recognizing that $\mathbf{U}^{(2)}_{B1-B2} = {}_2U_2$ (in Mattice's notation) and setting $\tau = 1$.

The fourth bond affected by the branch at N is the N–C1 bond. Since this bond has two neighboring bonds, A1-N and N-B1, the $\mathbf{U}_b$ matrix for this bond has three "layers", or sub-matrices, arising from the four-bond interactions. The two possible stereoisomers are presented in Fig. 4-6, corresponding to the same cases considered in Figs. 4-3-4-5.

For case (1) ($\phi_{offset} = +120°$), it follows that

$$\mathbf{D}^{(1)}_{N-C1} = diag\,(1\cdot\sigma, \sigma\cdot\sigma, \sigma\cdot 1) \tag{4.3.28}$$

$$= diag\,(\sigma, \sigma^2, \sigma)\,, \tag{4.3.29}$$

corresponding to the three-bond (A1-C2,B1-C2) interactions (t, g^+), (g^+, g^-), and (g^-, t) experienced as the N-C1 bond is rotated through its three allowable RIS states (the states are for $\phi_{A1-N-C1-C2}$ and $\phi_{B1-N-C1-C2}$).

The four-bond interactions are those centered on N-C1 and its predecessor bonds A1-N and N-B1, and due to the interactions between the hydrogens of A2 and B2 and those of C2. Following Mattice,[217] one considers each state of the N-B1 bond (i.e., the state of $\phi_{A1-N-B1-B2}$) as yielding a separate "layer" (3×3 submatrix) of the $3 \times 3 \times 3$ matrix $\mathbf{V}^{(1)}_{N-C1}$. We follow the usual ordering in describing the layers – (t, g^+, g^-), as illustrated in Fig. 4-6. We construct $\mathbf{V}^{(1)}_{B1-B2}$ one layer at a time.

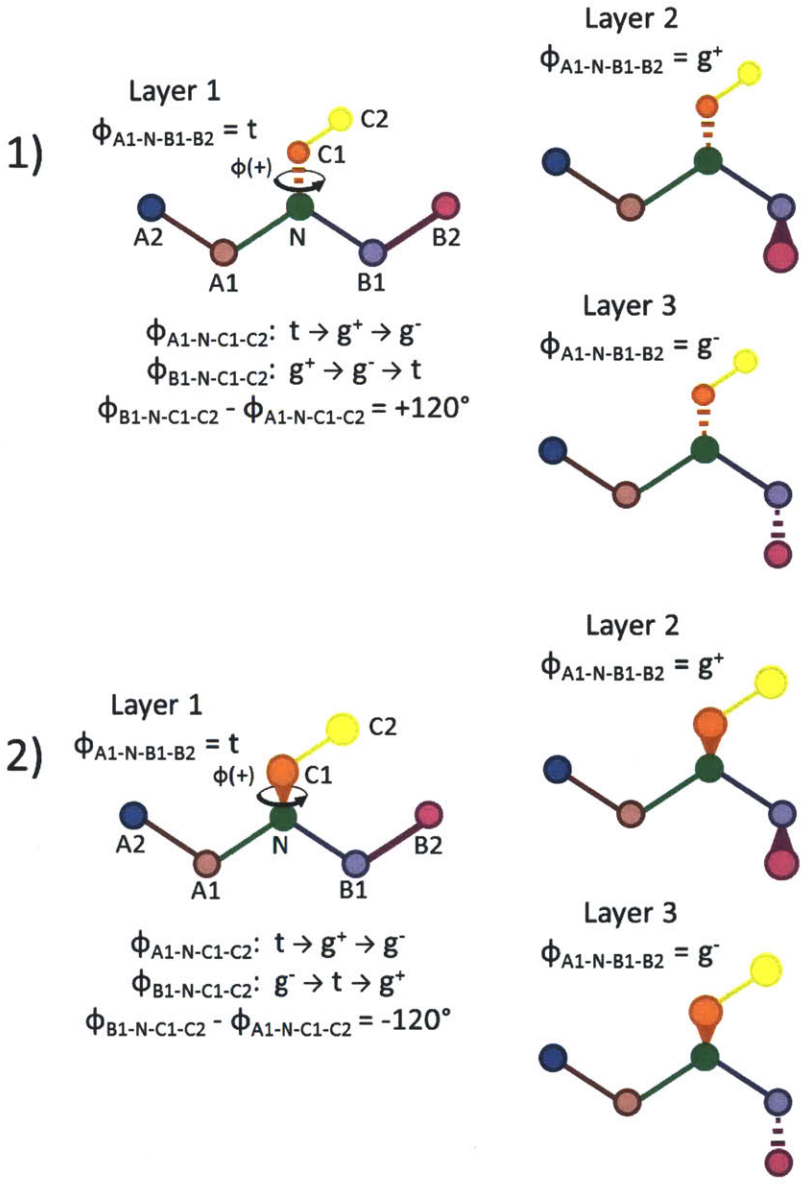


Figure 4-6: Diagram illustrating the atoms adjacent to, or participating in, the rotation about the N-C1 bond in a branched chain, and the distant atoms (A2 and B2) participating in four-bond interactions with C2. Cases 1) and 2) correspond to the same cases considered in Figs. 4-3-4-5. For the three-bond interactions, both $\phi_{A1-N-C1-C2}$ and $\phi_{B1-N-C1-C2}$ must be considered. For the four-bond interactions, the states of two predecessors must be considered: $\phi_{A2-A1-N-C1}$ and $\phi_{B2-B1-N-C1}$. Setting $\phi_{A1-N-B1-B2}$ to its various possible RIS states gives rise to a "three-layer" matrix.

Specifically,

$$\mathbf{V}_{N-C1}^{(1,1)} = \begin{bmatrix} w(g^+,t)\,w(g^-,g^+) & w(g^+,g^+)\,w(g^-,g^-) & w(g^+,g^-)\,w(g^-,t) \\ w(g^-,t)\,w(g^-,g^+) & w(g^-,g^+)\,w(g^-,g^-) & w(g^-,g^-)\,w(g^-,t) \\ w(t,t)\,w(g^-,g^+) & w(t,g^+)\,w(g^-,g^-) & w(t,g^-)\,w(g^-,t) \end{bmatrix}$$

$$= \begin{bmatrix} \omega & \psi^2 & \omega \\ \omega & \omega\psi & \psi \\ \tau\omega & \psi & 1 \end{bmatrix}, \tag{4.3.30}$$

$$\mathbf{V}_{N-C1}^{(1,2)} = \begin{bmatrix} w(g^+,t)\,w(t,g^+) & w(g^+,g^+)\,w(t,g^-) & w(g^+,g^-)\,w(t,t) \\ w(g^-,t)\,w(t,g^+) & w(g^-,g^+)\,w(t,g^-) & w(g^-,g^-)\,w(t,t) \\ w(t,t)\,w(t,g^+) & w(t,g^+)\,w(t,g^-) & w(t,g^-)\,w(t,t) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \psi & \omega\tau \\ 1 & \omega & \psi\tau \\ \tau & 1 & \tau \end{bmatrix}, \tag{4.3.31}$$

$$\mathbf{V}_{N-C1}^{(1,3)} = \begin{bmatrix} w(g^+,t)\,w(g^+,g^+) & w(g^+,g^+)\,w(g^+,g^-) & w(g^+,g^-)\,w(g^+,t) \\ w(g^-,t)\,w(g^+,g^+) & w(g^-,g^+)\,w(g^+,g^-) & w(g^-,g^-)\,w(g^+,t) \\ w(t,t)\,w(g^+,g^+) & w(t,g^+)\,w(g^+,g^-) & w(t,g^-)\,w(g^+,t) \end{bmatrix}$$

$$= \begin{bmatrix} \psi & \psi\omega & \omega \\ \psi & \omega^2 & \psi \\ \tau\psi & \omega & 1 \end{bmatrix}, \tag{4.3.32}$$

where the rows of $\mathbf{V}_{N-C1}^{(1,l)}$ (the submatrix for layer $l = 1, 2, 3$) correspond to $\phi_{A2-A1-N-B1} = (t, g^+, g^-)$, and the columns of $\mathbf{V}_{N-C1}^{(1,l)}$ correspond to $\phi_{A1-N-C1-C2} = (t, g^+, g^-)$, with the total statistical weightings

$$w\left(\phi_{A2-A1-N-C1}, \phi_{A1-N-C1-C2}\right) w\left(\phi_{B2-B1-N-C1}, \phi_{B1-N-C1-C2}\right). \tag{4.3.33}$$

These are the weightings due to the interactions between A2 and C2 multiplied by the weightings due to the interactions between B2 and C2.

Applying Eq. 4.3.4 yields

$$\begin{aligned} \mathbf{U}_{N-C1}^{(1,1)} &= \mathbf{V}_{N-C1}^{(1,1)}\mathbf{D}_{N-C1}^{(1,1)} \\ &= \sigma \begin{bmatrix} \omega & \psi^2\sigma & \omega \\ \omega & \omega\psi\sigma & \psi \\ \tau\omega & \psi\sigma & 1 \end{bmatrix}, \end{aligned} \tag{4.3.34}$$

$$\begin{aligned} \mathbf{U}_{N-C1}^{(1,2)} &= \mathbf{V}_{N-C1}^{(1,2)}\mathbf{D}_{N-C1}^{(1,2)} \\ &= \sigma \begin{bmatrix} 1 & \psi\sigma & \omega\tau \\ 1 & \omega\sigma & \psi\tau \\ \tau & \sigma & \tau \end{bmatrix}, \end{aligned} \tag{4.3.35}$$

$$\begin{aligned} \mathbf{U}_{N-C1}^{(1,3)} &= \mathbf{V}_{N-C1}^{(1,3)}\mathbf{D}_{N-C1}^{(1,3)} \\ &= \sigma \begin{bmatrix} \psi & \psi\omega\sigma & \omega \\ \psi & \omega^2\sigma & \psi \\ \tau\psi & \omega\sigma & 1 \end{bmatrix}. \end{aligned} \tag{4.3.36}$$

Note that Eqs. 4.3.34-4.3.36 are equivalent to Mattice's Eqs. (25)-(27),[217] recognizing that $\mathbf{U}_{N-C1}^{(1,l)} = {}_3U_1$ (lth layer) (in Mattice's notation) and setting $\tau = 1$.

For case (2) ($\phi_{offset} = -120°$), it follows that

$$\mathbf{D}_{N-C1}^{(2)} = diag\left(\sigma, \sigma, \sigma^2\right), \tag{4.3.37}$$

$$\mathbf{V}_{N-C1}^{(1,1)} = \begin{bmatrix} w(g^-,t)\,w(g^+,g^-) & w(g^-,g^+)\,w(g^+,t) & w(g^-,g^-)\,w(g^+,g^+) \\ w(t,t)\,w(g^+,g^-) & w(t,g^+)\,w(g^+,t) & w(t,g^-)\,w(g^+,g^+) \\ w(g^+,t)\,w(g^+,g^-) & w(g^+,g^+)\,w(g^+,t) & w(g^+,g^-)\,w(g^+,g^+) \end{bmatrix}$$

$$= \begin{bmatrix} \omega & \omega & \psi^2 \\ \tau\omega & 1 & \psi \\ \omega & \psi & \omega\psi \end{bmatrix}, \tag{4.3.38}$$

$$\mathbf{V}_{N-C1}^{(1,2)} = \begin{bmatrix} w(g^-,t)\,w(g^-,g^-) & w(g^-,g^+)\,w(g^-,t) & w(g^-,g^-)\,w(g^-,g^+) \\ w(t,t)\,w(g^-,g^-) & w(t,g^+)\,w(g^-,t) & w(t,g^-)\,w(g^-,g^+) \\ w(g^+,t)\,w(g^-,g^-) & w(g^+,g^+)\,w(g^-,t) & w(g^+,g^-)\,w(g^-,g^+) \end{bmatrix}$$

$$= \begin{bmatrix} \psi & \omega & \psi\omega \\ \tau\psi & 1 & \omega \\ \psi & \psi & \omega^2 \end{bmatrix}, \tag{4.3.39}$$

$$\mathbf{V}_{N-C1}^{(1,3)} = \begin{bmatrix} w(g^-,t)\,w(t,g^-) & w(g^-,g^+)\,w(t,t) & w(g^-,g^-)\,w(t,g^+) \\ w(t,t)\,w(t,g^-) & w(t,g^+)\,w(t,t) & w(t,g^-)\,w(t,g^+) \\ w(g^+,t)\,w(t,g^-) & w(g^+,g^+)\,w(t,t) & w(g^+,g^-)\,w(t,g^+) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \omega\tau & \psi \\ \tau & \tau & 1 \\ 1 & \psi\tau & \omega \end{bmatrix}, \tag{4.3.40}$$

$$\mathbf{U}_{N-C1}^{(2,1)} = \sigma \begin{bmatrix} \omega & \omega & \psi^2\sigma \\ \tau\omega & 1 & \psi\sigma \\ \omega & \psi & \omega\psi\sigma \end{bmatrix}, \tag{4.3.41}$$

$$\mathbf{U}_{N-C1}^{(2,2)} = \sigma \begin{bmatrix} \psi & \omega & \psi\omega\sigma \\ \tau\psi & 1 & \omega\sigma \\ \psi & \psi & \omega^2\sigma \end{bmatrix}, \tag{4.3.42}$$

and

$$\mathbf{U}_{N-C1}^{(2,3)} = \sigma \begin{bmatrix} 1 & \omega\tau & \psi\sigma \\ \tau & \tau & \sigma \\ 1 & \psi\tau & \omega\sigma \end{bmatrix}. \tag{4.3.43}$$

Note that Eqs. 4.3.41-4.3.43 are equivalent to Mattice's Eq. (12)-(14),[217] recognizing that $\mathbf{U}_{N-C1}^{(2,l)} = {}_3U_1$ (lth layer) (in Mattice's notation) and setting $\tau = 1$.

The fifth, and final, bond affected by the branch at N is the C1-C2 bond, which is quite similar to the B1-B2 bond. Specifically, this bond also exhibits three-bond interactions similar to linear alkanes, and four-bond interactions which must take into account the presence of both A1 and B1 relative to C3. The two possible stereoisomers are presented in Fig. 4-7, corresponding to the same cases considered in Figs. 4-3-4-6.

For case (1) ($\phi_{offset} = +120°$), it follows that

$$\mathbf{D}_{C1-C2}^{(1)} = diag\,(1, \sigma, \sigma)\,, \tag{4.3.44}$$

corresponding to the three-bond (N-C2) interactions (t), (g^+), and (g^-) experienced as the C1-C2 bond is rotated through its three allowable RIS states. As indicated earlier, this is identical to the linear alkane case, as expressed in Eq. 4.3.5.

The four-bond interactions are those centered on C1-C2 and its predecessor bond N-C1, and due to the interactions between the hydrogens of A1 and B1 and those of C3. If one considers the rotational states of the N-C1 bond, depicted in Fig. 4-6 and summarized on the right in Fig. 4-7, one can construct $\mathbf{V}_{C1-C2}^{(1)}$ from the appropriate weightings. Specifically,

$$\begin{aligned} \mathbf{V}_{C1-C2}^{(1)} &= \begin{bmatrix} w\,(t,t)\,w\,(g^+,t) & w\,(t,g^+)\,w\,(g^+,g^+) & w\,(t,g^-)\,w\,(g^+,g^-) \\ w\,(g^+,t)\,w\,(g^-,t) & w\,(g^+,g^+)\,w\,(g^-,g^+) & w\,(g^+,g^-)\,w\,(g^-,g^-) \\ w\,(g^-,t)\,w\,(t,t) & w\,(g^-,g^+)\,w\,(t,g^+) & w\,(g^-,g^-)\,w\,(t,g^-) \end{bmatrix} \\ &= \begin{bmatrix} \tau & \psi & \omega \\ 1 & \psi\omega & \omega\psi \\ \tau & \omega & \psi \end{bmatrix}. \end{aligned} \tag{4.3.45}$$

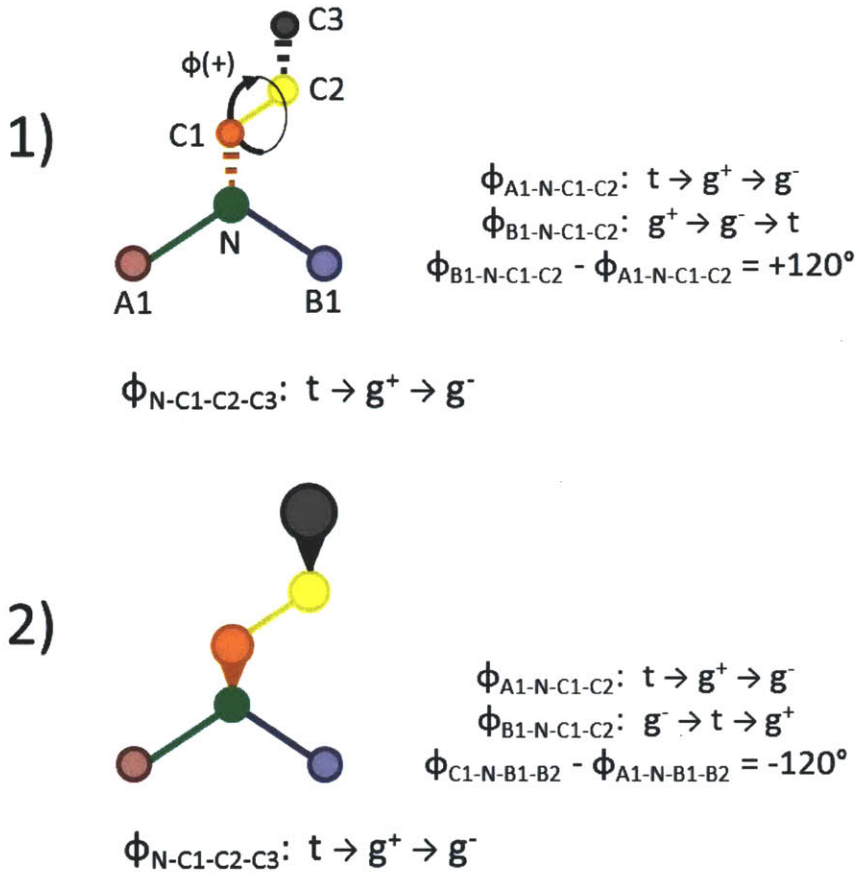


Figure 4-7: Diagram illustrating the atoms adjacent to, or participating in, the rotation about the C1-C2 bond in a branched chain, and the distant atoms (A1 and B1) participating in four-bond interactions with C3. Cases 1) and 2) correspond to the same cases considered in Figs. 4-3-4-6. For the three-bond interactions, only $\phi_{N-C1-C2-C3}$ is relevant (identical to a linear alkane case), while the four-bond interactions must consider the position of the distant atoms, as specified through $\phi_{A1-N-C1-C2}$ and $\phi_{B1-N-C1-C2}$.

where the rows of $\mathbf{V}^{(1)}_{C1-C2}$ correspond to $\phi_{A1-N-C1-C2} = (t, g^+, g^-)$, and the columns of $\mathbf{V}^{(1)}_{C1-C2}$ correspond to $\phi_{N-C1-C2-C3} = (t, g^+, g^-)$, with the total statistical weightings

$$w\left(\phi_{A1-N-C1-C2}, \phi_{N-C1-C2-C3}\right) w\left(\phi_{B1-N-C1-C2}, \phi_{N-C1-C2-C3}\right). \tag{4.3.46}$$

These are the weightings due to the interactions between A1 and C3 multiplied by the weightings due to the interactions between B1 and C3.

Applying Eq. 4.3.4 yields

$$\begin{aligned} \mathbf{U}^{(1)}_{C1-C2} &= \mathbf{V}^{(1)}_{C1-C2}\mathbf{D}^{(1)}_{C1-C2} \\ &= \begin{bmatrix} \tau & \psi\sigma & \omega\sigma \\ 1 & \psi\omega\sigma & \omega\psi\sigma \\ \tau & \omega\sigma & \psi\sigma \end{bmatrix}. \end{aligned} \tag{4.3.47}$$

Note that Eq. 4.3.47 is equivalent to Mattice's Eq. (24),[217] recognizing that $\mathbf{U}^{(1)}_{C1-C2} = {}_3U_2$ (in Mattice's notation) and setting $\tau = 1$.

For case (2) ($\phi_{offset} = -120°$), it follows that

$$\mathbf{D}^{(2)}_{C1-C2} = diag\left(1, \sigma, \sigma\right), \tag{4.3.48}$$

$$\begin{aligned} \mathbf{V}^{(2)}_{C1-C2} &= \begin{bmatrix} w(t,t)\,w(g^-,t) & w(t,g^+)\,w(g^-,g^+) & w(t,g^-)\,w(g^-,g^-) \\ w(g^+,t)\,w(t,t) & w(g^+,g^+)\,w(t,g^+) & w(g^+,g^-)\,w(t,g^-) \\ w(g^-,t)\,w(g^+,t) & w(g^-,g^+)\,w(g^+,g^+) & w(g^-,g^-)\,w(g^+,g^-) \end{bmatrix} \\ &= \begin{bmatrix} \tau & \omega & \psi \\ \tau & \psi & \omega \\ 1 & \omega\psi & \psi\omega \end{bmatrix}. \end{aligned} \tag{4.3.49}$$

and

$$\mathbf{U}^{(2)}_{C1-C2} = \begin{bmatrix} \tau & \omega\sigma & \psi\sigma \\ \tau & \psi\sigma & \omega\sigma \\ 1 & \omega\psi\sigma & \psi\omega\sigma \end{bmatrix}, \tag{4.3.50}$$

where $\mathbf{D}^{(2)}_{C1-C2} = \mathbf{D}^{(1)}_{C1-C2}$, because C1 does participate in the three-bond interactions in this case.

Note that Eq. 4.3.50 is equivalent to Mattice's Eq. (20),[217] recognizing that $\mathbf{U}^{(2)}_{B1-B2} = {}_3U_2$ (in Mattice's notation) and setting $\tau = 1$.

4.4 Conclusions

In this Chapter, I presented a chain packing theory for linear and branched surfactants that can be used to obtain the packing free-energy contribution to the free energy of micellization (which is discussed in Section 2.3 of Chapter 2). This single-chain, mean-field (SCMF) theory is based on the work of Ben-Shaul and coworkers.[82] It involves consideration of the probability distribution of conformations of a central, representative chain which is dependent on: (i) a model for the internal energy of the chain as a function of its conformation, (ii) the specific nature of the confinement in which the chain resides (i.e., the micelle core shape and size), and (iii) a mean pressure field, which results in a pressure-volume contribution to the probability distribution function arising from the application of constant density (incompressibility) constraints. This mean pressure field is viewed as representing the effect of all the other chains on the central chain, in an implicit manner.[232] Chain models of various types can be considered. The Rotational Isomeric State (RIS) model of Flory[77] is of particular relevance to this Thesis. Specifically, the RIS model replaces continuous torsional potential energy profiles with a series of discrete dihedral states and associated energies representing the minima in the original profile. The RIS model is particularly effective at reducing the number of degrees of freedom of the system, while maintaining good agreement with experimental measurement of chain configurational properties under bulk conditions (such as the end-to-end distance).[229] Additionally,

RIS models for a wide variety of chain types have been developed, as summarized in the compilation of Rehahn, Mattice, and Suter.[229] This provides a promising avenue for future expansion of this packing approach to include increasingly-complex surfactant tails.

In Section 4.2, I derived the probability distribution function (pdf) for the central chain using an application of the Maximum Entropy Principle.[220,221] This procedure involved minimization of the free energy of packing subject to the incompressibility constraints, using a variational approach. After obtaining the pdf, I discussed the procedure to calculate the packing free energy of short chains, which involves generating a full set of internal chain conformations and randomly-sampled external orientations from which the pressure field that enforces the constraint can be computed. Given the pressure field and the set of chain conformations, I discussed how the packing free energy is calculated. For longer or highly-branched chains, I refer the reader to the discussion in Appendix B, where I discuss an alternative sampling approach, using a Monte Carlo methodology, which, together with a decomposition of the packing free energy into confinement and constraint terms, enables efficient calculation of packing free energies above 16 rotatable bonds relative to the full-conformation approach.

In Section 4.3, I provided a full description of the RIS approach, including a summary of its application to linear chains. I then derived an RIS model for singly-branched chains based on the work of Mattice.[217] This approach involves enumeration of all three-bond and four-bond interactions as a function of neighboring bond dihedral states, avoiding double-counting by systematically building up the chain along the primary branch, followed by the secondary branch. For applications such as chain packing, there is no need to use the matrix formalism of Mattice. In fact, the enumeration can be performed algorithmically to generate RIS models for arbitrarily-branched chains. Although not presented in this Thesis, I have implemented this type of algorithmic approach for future chain packing applications involving highly-branched surfactants. The weighting results presented in Section 4.3.3 are fully replicated by this approach.

In this Chapter, along with Chapters 2 and 3, I have presented the major details

of the theoretical and simulations-based aspects of my computational framework for the prediction of the micellization behavior of linear and branched surfactants. In the next Chapter, I apply this computational framework to practical surfactant systems, initially focusing on linear surfactants. Application to branched surfactants follows, in Chapter 6.

Chapter 5

Modeling Results for Linear Alkyl Surfactants

5.1 Overview

In the preceding Chapters, I have laid out the details of a computational framework for predicting the micellization properties of surfactants in dilute aqueous solution. This framework involves a molecular-thermodynamic (MT) theory (see Chapter 2 for details), which has the additional capability of accepting inputs from molecular dynamics computer simulations (which, when utilized, represents a combined computer-simulation/molecular-thermodynamic, or CSMT, framework, as described in Chapter 3). The MT framework is sensitive to the specific chemical structure of a surfactant, through the functional forms of the various free-energy contributions to the modified free energy of micellization, g_m, which it aims to predict. These contributions are described at length in Section 2.3 of Chapter 2 and throughout Chapter 4, which addresses the packing free energy penalty associated with confining the tail (or dehydrated) portion of a surfactant within a micelle core.

In this Chapter, I implement the MT framework presented in Chapters 2 to 4 to predict critical micelle concentrations (cmc's) of linear alkyl surfactants of various types, in order to rigorously test the theory. Additionally, I investigate the impact of using computer simulations within the related CSMT framework, both for surfactant

head/tail identification (as described in Section 3.3 of Chapter 3), and as a means for modifying the transfer free energy to reflect microstructural hydration information obtained from computer simulation (see Section 3.2 of Chapter 3). As discussed in Section 2.2.4 of Chapter 2, the cmc is exponentially-dependent on the free energy of micellization. Hence, small errors in this free energy will be amplified in calculating the cmc, and comparison to experiment will yield important insight into the strengths and limitations of the MT and CSMT frameworks, enabling recommendations to be made for future use of the theory. Optimal micelle shapes, sizes, and degrees of counterion binding have also been predicted for all these systems. However, I have chosen to wait until Chapter 6 to provide some representative results and comparison with experiment. With the exception of systems exhibiting cylindrical growth, these properties do not yield as direct an insight into the quality of free-energy predictions as the cmc does.

Since this Chapter is focused on linear alkyl surfactants, the majority of my attention here will be on the effect of the selection of the surfactant head group on the predicted cmc. The surfactant head influences the value of g_m through two key contributions: the steric contribution and the electrostatic contribution (for charged surfactants).[81] Surfactant heads are typically classified into four main categories: (i) anionic surfactants, possessing a negative charge, and a positively-charged counterion, (ii) cationic surfactants, possessing a positive charge, and a negatively-charged counterion, (iii) zwitterionic surfactants, possessing a dipole and no counterion, and (iv) nonionic surfactants, possessing neither charge nor counterion.[13] In Section 5.2, I select common examples from each of these four categories, and provide a brief discussion of the modeling process associated with each example, including specific molecular descriptors used for the surfactant head groups, a discussion of the appropriate force field parameters for use with each surfactant considered, and a brief summary of simulation-related results. In Section 5.3, I discuss the cmc results for charged surfactants (i.e., the anionic, cationic, and zwitterionic surfactants), highlighting the role of the position of the charge relative to the micelle core-water interface in quantifying the electrostatic free-energy contribution to g_m. In Section 5.4, I discuss the

cmc results for nonionic ethoxylated and sugar-based surfactants, where the steric free-energy contribution to g_m is key, since there are no electrostatic considerations. Finally, in Section 5.5, I summarize the key results obtained in this Chapter.

5.2 Discussion of Systems

5.2.1 Charged Surfactants

Anionic surfactants make up the bulk of industrial surfactant usage (upwards of 60% in 2000), followed by nonionics (about 25%), cationics (10%), and zwitterionics and other minor classes (about 5%).[12] For this study, I have selected four types of anionic surfactants, which represent over 30% of industrial surfactant usage by net tonnage: the alkyl sulfonates, alkyl sulfates, alkyl benzene sulfonates, and alkyl ethoxysulfates,[12] whose general chemical structures are depicted in Figure 5-1. Note that linear alkyl benzene sulfonates are one of the most widely-used types of surfactants,[13] but that this term also is meant to encompass the structures which I refer to as branched in Chapter 6.

Although I have chosen to focus on the common sulfate and sulfonate anionic surfactants in this work, the class of anionic surfactants also includes carboxylates and phosphates.[13] The MT framework can be used with these surfactants, and I have studied fluorinated phosphate surfactants in particular, but this work is outside the scope of this Thesis. The reader interested in fluorinated materials is directed to the work of Srinivasan and Blankschtein.[161]

For cationic surfactant class, I have selected three types: the alkyl trimethylammoniums, alkyl pyridiniums, and 1-methyl,4-alkyl pyridiniums. For the zwitterionic surfactant class, I have selected a single type: the sulfobetaines. I have chosen to place more emphasis on anionic and cationic surfactants than zwitterionic surfactants since they are more common, as already mentioned, and, in several cases, have branched forms, which I study in the next Chapter. Figure 5-2 shows the chemical structures for the cationic and zwitterionic types considered.

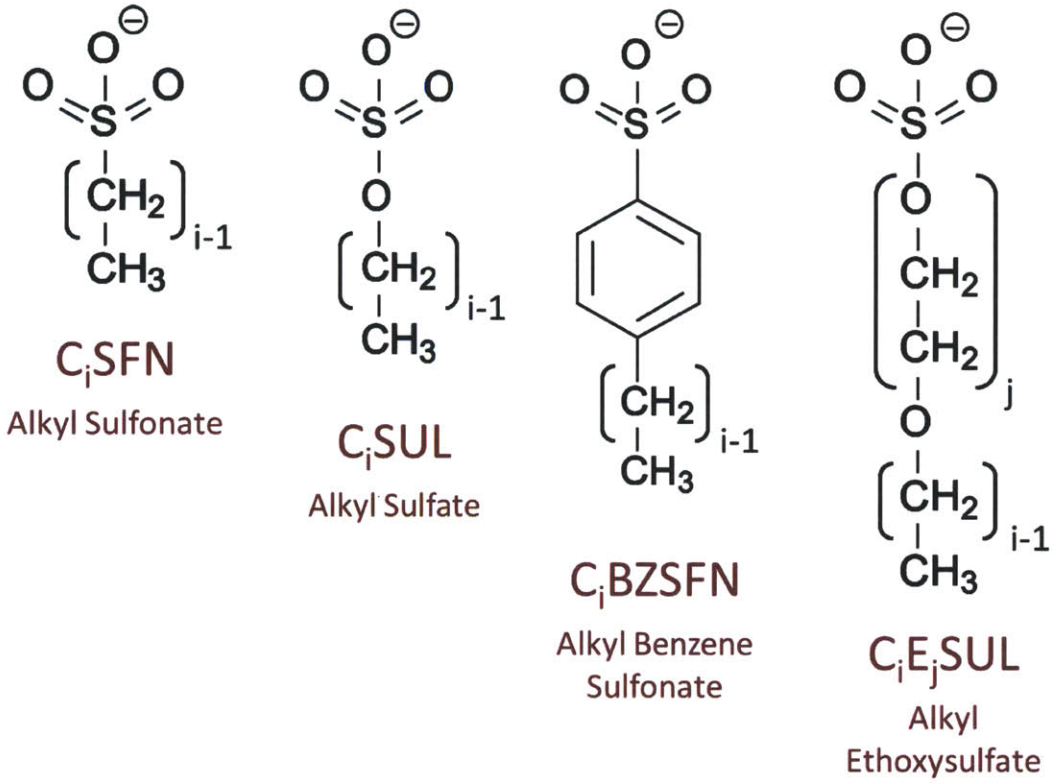


Figure 5-1: Chemical structures and labels for the linear anionic surfactants investigated in this Thesis. Selection of i in C_i is a selection of the length of the attached alkyl chain, where i is the number of carbon atoms in the chain. For the ethoxysulfate type, one must additionally specify j, the number of ethoxylate units present. (Note that these structures are intended to depict atom connectivity only, and are not intended to convey accurate bond angle and dihedral information.)

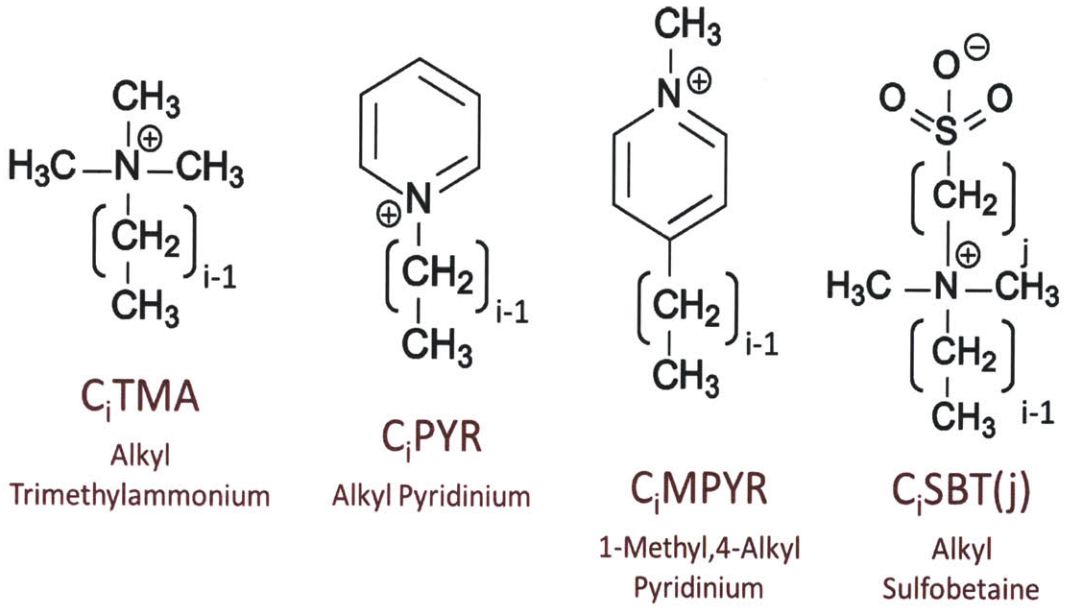


Figure 5-2: Chemical structures and labels for the linear cationic and zwitterionic surfactants investigated in this Thesis. Selection of i in C_i is a selection of the length of the attached alkyl chain, where i is the number of carbon atoms in the chain. For the sulfobetaine type, one must additionally specify j, the length of the alkyl spacer between the ammonium and sulfonate functional groups. (Note that these structures are intended to depict atom connectivity only, and are not intended to convey accurate bond angle and dihedral information.)

MT modeling of the surfactants depicted in Figures 5-1 and 5-2 requires a set of descriptors associated with the surfactant head and tail regions. The descriptors associated with the tail region include group contributions to the transfer-free energy, the tail-water interfacial tension, the Tolman length, the tail volume, and polynomials for the packing free energy of the tails as a function of micelle shape and size. Calculation of these descriptors has been described in great detail in Appendix F and will not be repeated here. Note that the alkyl chains considered in this section range from C_8 to C_{16}. Depending on the head/tail identification, the tail may consist of all or part of these alkyl chains and adjacent groups (e.g., part of the arene ring in the case of the alkyl benzene sulfonates in Figure 5-1). The descriptors associated with the head region include the size of the head (head area and length), and the location of any charges in the surfactant (one charge in the case of anionic and cationic surfactants, two charges in the case of zwitterionic surfactants) relative to the micelle core-water interface. These descriptors may also be sensitive to the precise head/tail identification, obtained either via rules of thumb, such as Tanford's suggestion that the alkyl carbon alpha to a charged group such as sulfate be considered hydrated (i.e., part of the head),[34] or using simulations, as described in Section 3.3 of Chapter 3. Table 5.1 contains the tail identifications used in this Chapter for MT modeling of charged surfactants. I use the Tanford rule of thumb for the sulfonates, sulfates, trimethylammoniums, pyridiniums, and sulfobetaines, whose CSMT tail identifications either agree with this rule of thumb, or contain additional groups in the head (see Table 5.2, which I will describe shortly). For the benzene sulfonates, ethoxysulfates, and 1-methyl pyridiniums, I use an identification based on a consensus value from the set of CSMT tail identifications presented in Table 5.2.

CSMT modeling of the surfactants depicted in Figures 5-1 and 5-2 requires a complete force field specification for each surfactant in order to conduct molecular dynamics simulations. The OPLS-AA force field of Jorgensen and coworkers, described in Section D.2 of Appendix D is appropriate for the alkyl chains, but the charged head groups are often incompletely specified. As mentioned in Appendix E, the force field development work of Canongia Lopes and coworkers in the ionic

Table 5.1: Charged Surfactant Tail Identifications Used for MT Modeling

Surfactant	Tail Identification[a]	Surfactant	Tail Identification[a]
C_iSFN	C_{i-1}	C_iTMA	C_{i-1}
C_iSUL	C_{i-1}	C_iPYR	C_{i-1}
C_iBZSFN	B_3C_i	C_iMPYR	B_0C_i
C_iE_jSUL	C_i	$C_iSBT(j)$	C_{i-1}

[a] The formula B_k means k aromatic carbon groups are included in the tail in a symmetric fashion ($k = 0$ indicates the full aromatic ring should be considered in the head region for packing calculation purposes).

liquids field is useful in completing these models. Specifically, Canongia Lopes and coworkers have computed atomic charges and bond length stretching, bond angle bending, and bond dihedral parameters for sulfonates and sulfates,[194] trimethylammoniums,[192] and pyridiniums[193] (specifically: 1-alkyl pyridiniums: it was necessary for me to compute a set of atomic charges for the 1-methyl,4-alkyl pyridiniums using the CHELPG algorithm[233] in Gaussian 03,[234] described in Section E.2.5 of Appendix E). The sulfobetaines $C_iSBT(j)$ can be constructed from sulfonate and dimethylammonium[192] parameters, if $j > 2$. The ethoxysulfates make additional use of the force field parameters for ethoxylates developed by Anderson and Wilson.[235] For the benzene sulfonates, I developed several new force field parameters. Development of force field parameters for this type is presented as a specific example throughout Section E.2 of Appendix E.

CSMT modeling requires simulation of a surfactant in monomer and micelle states, in order to compare the hydration states of atomic groups in each environment and ultimately compute a refined transfer free energy for the surfactant. The full protocol for simulations – used for all simulations carried out in generating the data presented in this Chapter – is described in Section 3.4 of Chapter 3. For each surfactant and specific micelle aggregation number considered, a minimum of five independent simulations of micelles with that aggregation number were run to assess the variation in predictions with starting conditions. For this purpose, Program BUILDER, described in Appendix C, was useful in creating unique initial micelle configurations. For cer-

tain surfactants of particular interest, I additionally explored the effect of changing the aggregation number, to assess the sensitivity of calculated quantities to the extent of the micelle. Finally, the specific aggregation numbers selected were based on consideration of the fully extended tail length and tail volume for C_i surfactants. Since I use a tail identification of C_{i-1} for most charged C_i surfactants, as shown in Table 5.1, a C_{12}SUL surfactant will have a tail of C_{11}. According to the group contributions to molecular volume presented in Table F.3 in Appendix F, a C_{11} chain will have a volume of 334 Å^3. The fully-extended tail length is taken to be $1.54+1.265(i-1) = 15.455$ Å for $i = 11$. For a spherical micelle, this results in an aggregation number of around 46. Looking forward to future investigations of mixed surfactant systems, it is convenient to select an aggregation number that permits evenly spaced composition fractions. Accordingly, I select multiples of 12, which are divisible by 3 and 4. The C_{12}SUL surfactant then has a natural aggregation number selection of 48. Similarly, 36 is convenient for C_{10} surfactants, 24 is convenient for C_8 surfactants, and 60 and 72 are appropriate values for C_{14} and C_{16} surfactants, respectively. When investigating multiple aggregation numbers for a single surfactant, I select from these numbers. As a result, for C_{12}SUL in particular, I have investigated aggregation numbers of 36, 48, 60, and even 72. There is, of course, a tendency towards ellipsoid formation at the higher aggregation numbers to accommodate the additional chains, and lower aggregation numbers tend to be more uniformly spherical.

Table 5.2 contains a summary of all the simulations conducted for charged surfactants, together with the head/tail identification results, presented as an identification of the tail region. What can be observed is a fairly consistent identification, within an atomic group or two, for each surfactant type across chain lengths. For example, the sulfonates C_iSFN have tails that are either C_{i-1} or C_i. The sulfates are even more uniformly C_i, especially for the important example of C_{12}SUL (sodium dodecyl sulfate). The benzene sulfonates, interestingly tend to exhibit half of the benzene ring in the head and half in the tail. I have used this fact in making MT predictions as well, since it was not clear in advance what the identification should be. In a sense, this is similar to the rule of thumb for purely alkane-based surfactants, in that some

groups are hydrated by the presence of the charge in the sulfonate head group and its interaction with water. The ethoxysulfates exhibit more of the chain in the tail, at times including the first ether group. As a result, the behavior is very much more like an ethoxylate (usually predicted in an MT setting using a C_i identification[85]), which is perhaps not surprising, since that is the group directly attached to the alkyl chain. For the trimethylammoniums, the identification is uniformly C_{i-4}, which is quite different from the usual assumption of C_{i-1}, but very much in line with fits of a micelle shape model to small-angle neutron scattering data for C_iTMA surfactants of various lengths, where between 3-4 methylene groups are consistently identified as "wet" (i.e., in the head region).[236] The two pyridinium cases exhibit that the location of the nitrogen in the pyridinium ring relative to the alkyl tail is important in determining the tail: the 1-alkyl pyridinium surfactants (C_iPYR) exhibit a C_{i-3} preference, while the 1-methyl,4-alkyl pyridinium surfactants (C_iMPYR) exhibit a C_i preference (in Table 5.2 this is marked as B_0C_i to indicate that the aromatic pyridinium ring should be considered in the head region when conducting packing calculations according to the mean field theory in Chapter 4). Finally, the sulfobetaines (C_iSBT(j)) exhibit behavior similar to the trimethylammoniums, since it is an ammonium group that is directly adjacent to the tail, as seen in Figure 5-2.

Given the tail identifications provided in Table 5.1 for MT runs, and in Table 5.2 for CSMT runs, and the details regarding tail modeling in Chapter 4 and Appendix F, it remains to compute the head properties necessary for modeling. I summarize a small selection of these in Table 5.3, based on calculations of the sort portrayed in Figure 5-3 for C_iSUL, where force field parameters involving equilibrium bond lengths and bond angles are used to calculate the distance between the micelle core-water interface and average location of charge, d_z, and head lengths, l_h, and head areas, a_h, are similarly computed, with the additional input of the van der Waals radii of the distal atoms.

Table 5.2: Charged Surfactant Tail Identifications from Computer Simulation

Surfactant	Ion	Tail Identification[a]
C_{10}SFN	Na+	S24: C_9 [1], C_{10} [6]; S36: C_9 [8], C_{10} [1]
C_{12}SFN	Na+	S36: C_{11} [5]; S48: C_{11} [5]
C_{14}SFN	Na+	S60: C_{13} [3]; C_{14} [2]
C_{16}SFN	Na+	S72: C_{15} [5]
C_{10}SUL	Na+	S24: C_9 [6]; S36: C_9 [6]
C_{12}SUL	Na+	S36: C_{11} [4]; S48: C_{11} [5]; S60: C_{11} [5]; S72: C_{11} [5]
C_{14}SUL	Na+	S48: C_{13} [4], C_{14} [1]; S60: C_{13} [4], C_{14} [1]
C_{16}SUL	Na+	S60: C_{15} [3], C_{16} [2]; S72: C_{15} [2], C_{16} [3]
C_8BZSFN	Na+	S24: B_3C_8 [5]
C_{10}BZSFN	Na+	S24: B_3C_{10} [3], B_5C_{10} [1], B_6C_{10} [1]; S36: B_3C_{10} [5]
C_{12}BZSFN	Na+	S24: B_3C_{12} [2], B_5C_{12} [2], B_6C_{12} [1]; S36: B_3C_{12} [3], B_5C_{12} [1], B_6C_{12} [1]; S48: B_3C_{12} [5]
$C_{12}E_1$SUL	Na+	S48: C_{12} [4], C_{12}+O+CH_2 [1]
$C_{12}E_2$SUL	Na+	S48: C_{12}+O [3], C_{12}+O+CH_2 [2]
$C_{12}E_3$SUL	Na+	S48: C_{12}+O [3], C_{12}+O+CH_2 [1]
C_{12}TMA	Br-	S36: C_8 [5]; S48: C_8 [8]; S60: C_8 [4]
C_{14}TMA	Br-	S60: C_{10} [5]
C_{16}TMA	Br-	S72: C_{12} [5]
C_{12}PYR	Br-	S48: C_9 [3], C_{10} [2]
C_{14}PYR	Br-	S60: C_{11} [4], C_{12} [1]
C_{16}PYR	Br-	S72: C_{13} [5]
C_{10}MPYR	I-	S36: B_0C_{10} [5]
C_{12}MPYR	I-	S48: B_0C_{12} [7]
C_{12}SBT(3)		S36: C_9 [5]; S48: C_8 [5], C_9 [4]
C_{14}SBT(3)		S60: C_{10} [1], C_{11} [4]
C_{16}SBT(3)		S72: C_{13} [5]
C_{12}SBT(4)		S36: C_9 [5]; S48: C_9 [5]; S60: C_7 [1], C_8 [4]

[a] See Section 3.3 of Chapter 3 for details of the head/tail identification procedure. Entries show the spherical aggregation number simulated as SN, where N is the aggregation number. Tail identifications are followed by the number of simulations possessing that identification. The formula B_k means k aromatic carbon groups are included in the tail in a symmetric fashion ($k = 0$ indicates the full aromatic ring should be considered in the head region for packing calculation purposes).

Table 5.3: Head Group Properties for Charged Surfactants

Head	a_h (Å^2)[a]	z[a]	d_z (Å)[a]	l_h (Å)[a]
SFN	24.9	-1	0.362	1.762
SFN+CH_2	24.9	-1	2.154	3.554
SUL	24.9	-1	1.973	3.373
SUL+CH_2	24.9	-1	2.888	4.288
SFN+B3	24.9	-1	2.861	4.261
E_1SUL	24.9[b]	-1	5.293	6.693
E_2SUL	24.9[b]	-1	6.509	7.909
E_3SUL	24.9[b]	-1	7.693	9.093
TMA+3CH_2	28.4	+1	3.843	6.848
TMA+4CH_2	28.4	+1	5.021	8.026
PYR+2CH_2	35.0	+1	0.663	3.689
PYR+3CH_2	35.0	+1	3.857	8.717
MPYR	35.0	+1	2.712	5.738
SBT(3)+3CH_2	28.4	+1/-1	3.843/9.098	10.498
SBT(4)+3CH_2	28.4	+1/-1	3.843/10.443	11.843

[a] a_h: head cross-sectional area; z: valence; d_z: distance from micelle core-water interface to average charge location; l_h: average head length.

[b] Head areas for ethoxysulfates with 3 or fewer ethoxylate units are determined by the size of the sulfate group.

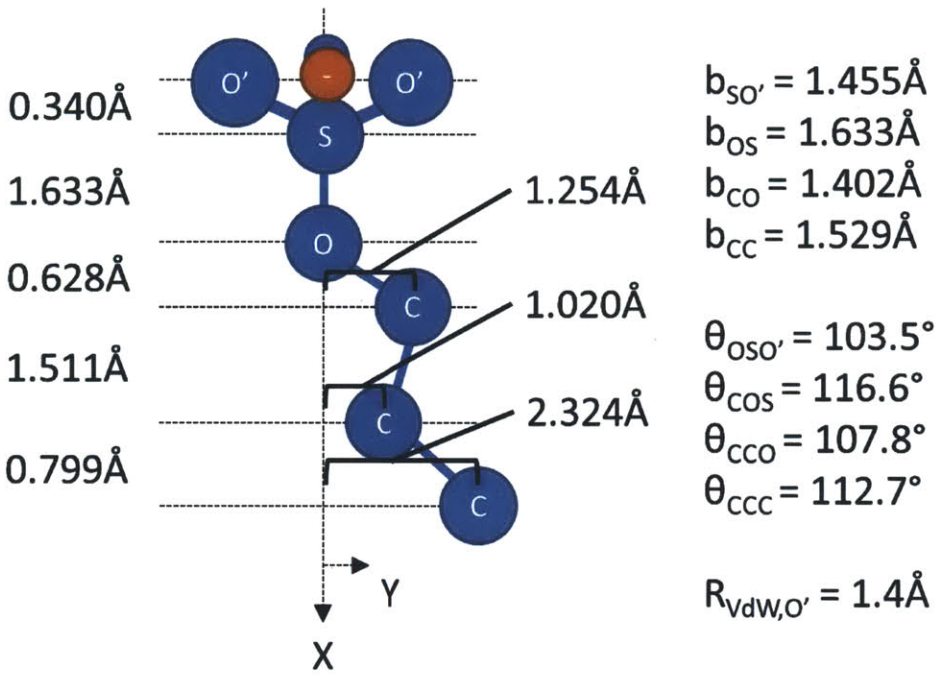


Figure 5-3: Distances used in computing surfactant head molecular descriptors for alkyl sulfate (C_iSUL) surfactants. The distance from the micelle core-water interface to the average location of charge (portrayed as the orange circle) depends on the specific tail identification. If the tail identification is C_{i-1}, then the distance from the first blue circle labeled "C" to the orange circle is d_z, which can be calculated from the geometry of the surfactant, as captured in the force field as equilibrium bond lengths and angles. Note that one typically assumes a fully extended, all-trans configuration when calculating d_z. Another possibility is to examine mean distances during simulations, if the head is larger and the surfactant is flexible (unlike the case with, say, an alkyl aryl surfactant such as benzene sulfonate, which is quite rigid).

5.2.2 Nonionic Surfactants

For nonionic surfactants, I have chosen to investigate the common alkyl ethoxylates and sugar-based surfactants (glycosides). The specific chemical structures selected are portrayed in Figure 5-4, and include both α and β anomeric forms of the alkyl glucoside and alkyl maltoside surfactants, which differ only in the stereochemistry of the chiral center attached to the ether oxygen bridging the sugar ring and the alkyl tail. It has been shown that this subtle difference can cause perceptible changes in the micelle structure and a slight elevation in cmc of the β form relative to the α form.[237] In the MT theory there is no way to distinguish between the two forms, if the head areas are considered to be the same. However, in the CSMT framework, differences in micelle structure will impact the local hydration states of individual groups, which may lead to differences ultimately in predicted cmc's. In fact, this will be confirmed in Section 5.4 below. β-Galactoside is included as an additional example of a single change to the stereochemistry of a chiral center, relative to glucoside. The alkyl methylglucamide surfactant type has been included as a test of the MT theory. CSMT results for this surfactant using my updated techniques are not yet available, although this surfactant has been previously studied by Stephenson and coworkers with an older approach.[109]

The tail identification used with the nonionic surfactants in Figure 5-4 in conventional MT modeling is taken to be C_i, as in Ref. [85]. For CSMT modeling, the OPLS-AA force field[188] can be applied to the alkyl tail, the ethoxylate model of Anderson and Wilson[235] can be applied to ethoxylate heads (with the addition of a missing set of dihedral parameters for the terminal ether oxygen-carbon-carbon-alcohol oxygen dihedral, which I have calculated according to the procedure described in Section E.2.4 of Appendix E), and the carbohydrate model of Damm and coworkers[238] can be applied to the glycosides. Aggregation numbers are selected for spherical micelle simulations as described in Section 5.2.1 above. The tail identifications obtained from these simulations are summarized in Table 5.4. The shorter chain ethoxylates tend to have at least one methylene group in the head, while the C_{12} ethoxylates are

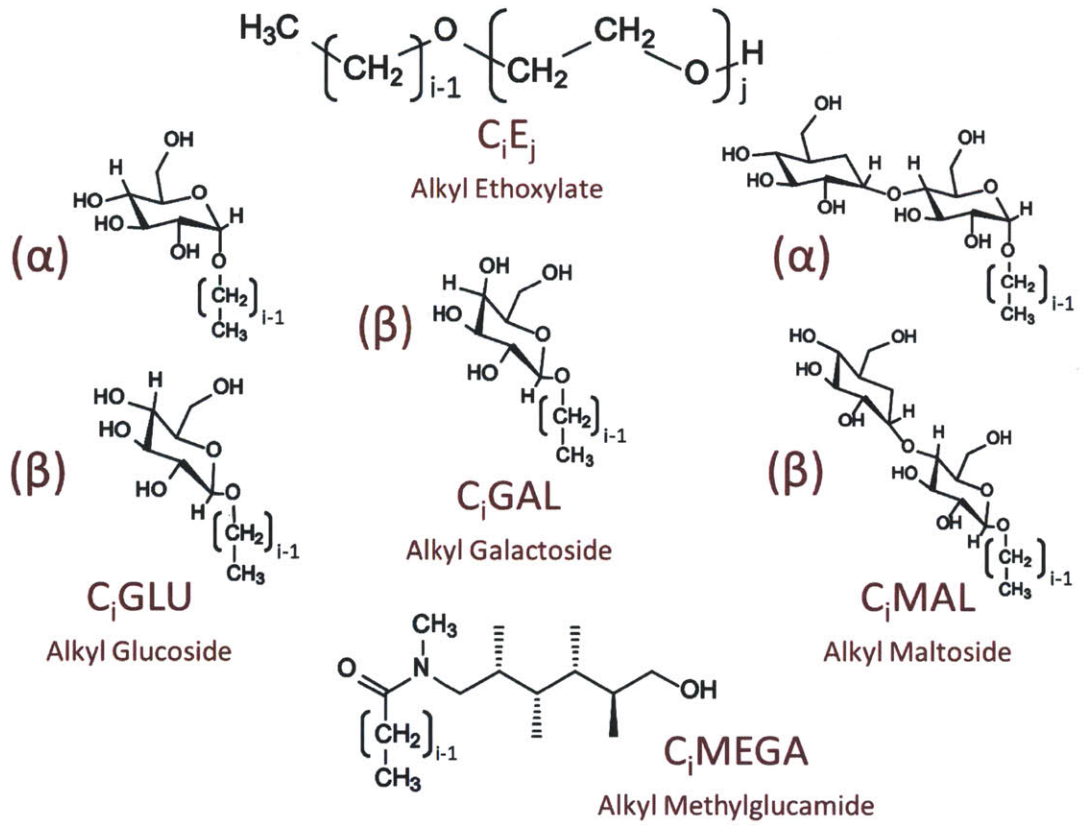


Figure 5-4: Chemical structures and labels for the linear nonionic surfactants investigated in this Thesis. Selection of i in C_i is a selection of the length of the attached alkyl chain, where i is the number of carbon atoms in the chain. For the ethoxylate type, one must additionally specify j, the number of ethoxylate units present. (Note that these structures are intended to depict atom connectivity only, and are not intended to convey accurate bond angle and dihedral information.)

Table 5.4: Nonionic Surfactant Tail Identifications from Computer Simulation

Surfactant	Tail Identification[a]
C_8E_4	S24: C_8 [9]
C_8E_6	S24: C_6 [2], C_7 [7]
C_8E_8	S24: C_6 [5], C_7 [3]
$C_{10}E_4$	S36: C_9 [5]
$C_{10}E_6$	S24: C_9 [4], C_{10} [1]; S36: C_9 [5]; S48: C_9 [5]
$C_{10}E_8$	S24: C_9 [4], C_{10} [1]; S36: C_8 [1], C_9 [4]; S48: C_8 [1], C_9 [4]
$C_{10}E_{10}$	S24: C_8 [2], C_9 [3]; S36: C_8 [1], C_9 [4]; C_8 [1], C_9 [4]
$C_{12}E_6$	S36: C_{12} [4], C_{12}+O [1]; S48: C_{12} [4], C_{12}+O [1]; S60: C_{12} [5]
$C_{12}E_8$	S36: C_{11} [3], C_{12} [2]; S48: C_{11} [2], C_{12} [3]; S60: C_{11} [4], C_{12} [1]
$C_{12}E_{10}$	S48: C_{12} [4], C_{12}+O [1]
C_8GALB(β)	S24: C_7 [5], C_8 [3]
C_8GLUB(α)	S24: C_7 [7]
C_8GLUB(β)	S24: C_7 [4], C_8 [1]
C_{10}GLUB(β)	S36: C_9 [3], C_{10} [2]
C_{12}GLUB(β)	S48: C_{11} [2], C_{12} [3]
C_{12}MAL(α)	S48: C_{11} [5]
C_{12}MAL(β)	S36: C_{11} [3], C_{12} [1], C_{12}+O [1]; S48: C_{11} [1], C_{12} [4]; S60: C_{11} [5]

[a] See Section 3.3 of Chapter 3 for details of the head/tail identification procedure. Entries show the spherical aggregation number simulated as SN, where N is the aggregation number. Tail identifications are followed by the number of simulations possessing that identification.

more in line with the MT tail identification. An early MT treatment of ethoxylates treated ethoxylated surfactants as having a tail of C_{i-1}.[81] From the tail identifications here, it appears that both identifications in the literature have some validity. Similar identifications are observed with the glucosides and maltosides: an identification of either C_i or C_{i-1} could be argued for. Note that, in CSMT calculations for a given micelle simulation, the identification associated with that micelle is used. As a result, averaged results for cmc's reported in Section 5.4 can contain contributions from differing identifications, according to Table 5.4. This provides no difficulty to the CSMT framework, which is designed such that the calculation of the transfer free energy is much less sensitive to the specific head-tail identification than in the original MT approach.

Finally, as in Section 5.2.1, based on the MT and CSMT tail identifications, one

Table 5.5: Head Group Properties for Nonionic Surfactants

Head	a_h (Å^2)[a]	Head	a_h (Å^2)[a]
E_4	28.3[b]	GAL	40.0
E_6	42.0[b]	GLU	40.0
E_8	52.9[b]	MAL	52.0
E_{10}	63.2[b]	MEGA	62.0

[a] a_h: head cross-sectional area
[b] Computed according to Eq. 5.2.1.

must compile the corresponding head properties. For nonionic surfactants, these are limited to the head area of the surfactants, which, for ethoxylates, can be computed from a scaling argument as follows:[90]

$$a_h(j) = 42.0\left(\frac{j}{6}\right)^{0.8}, \tag{5.2.1}$$

where j is the number of ethoxylate units in the C_iE_j surfactant. The head area for the sugar-based surfactants can be computed from geometric considerations. Table 5.5 summarizes these results.

5.3 Modeling Results for Charged Surfactants

5.3.1 Anionic Surfactants

In Table 5.6, I present experimental and predicted data for the cmc's of homologous series of alkyl sulfonates (C_iSFN), alkyl sulfates (C_iSUL), alkyl benzene sulfonates (C_iBZSFN), and alkyl ethoxysulfates (C_iE_jSUL), where i indicates the alkyl chain length, and j indicates the ethoxylate chain length (where applicable). The predicted data is divided into two categories: (i) predictions using the molecular-thermodynamic (MT) theory, and (ii) predictions using the combined computer-simulation/molecular-thermodynamic (CSMT) theory (involving computer simulations). Each category is subdivided into the four solubility models described in Section F.5 of Appendix F: (a) a direct group contribution method based on the correlations developed by

Abraham (MHA),[239] (b) a new fitting of group contribution terms to experimental solubilities for linear and branched oils, following a similar approach to the MHA model, which I introduced to explore whether linear and branched oils should be fit together or separately (MHA2), (c) a group contribution method with a constant term and specific adjustments based on the chemical nature of the surfactant tail, due to Klopman, Wang, and Balthasar (KWB),[5] and (d) a group contribution method with a constant term and separate handling of alkyl groups in alkanes (paraffins) and alkylarenes, due to Kühne, Ebert, Kleint, Schmidt, and Schüürmann (KEKSS).[6] Note that the MT predictions have a single value associated with each surfactant chemical structure and head/tail identification, and that only a single head/tail identification was selected for each series of surfactants, according to the identifications in Table 5.1 and mentioned in the text of Section 5.2.2.

Although direct comparison of predicted and experimental cmc's is useful to a formulator, who will design surfactant formulations based on this guidance, a comparison of the free energies associated with these cmc's is beneficial from the standpoint of gauging errors in the MT or CSMT model. In Section 2.2.4 of Chapter 2, I provided an expression relating the cmc (in mole fraction units) to the model free energy of micellization:

$$X_{cmc} \approx \exp\left(+\frac{g_m^*}{k_BT}\right), \tag{5.3.1}$$

where g_m^* is the free energy of micellization corresponding to an optimized micellar configuration (shape, size, and composition) at the solution conditions of interest and subject to mass balance constraints (k_BT is the thermal energy). If one is provided an experimental cmc in concentration units, an effective value for g_m^* can be determined by taking the natural logarithm of both sides of Eq. 5.3.1, and converting between mole fraction and concentration units. This yields

$$\frac{g_m^*}{k_BT} \approx \ln\left(\frac{C_{cmc}[mM]}{1000\mathcal{M}_{H_2O}[M]}\right), \tag{5.3.2}$$

where C_{cmc} is the experimental cmc and $\mathcal{M}_{H_2O}$ is the molarity of water at the temperature of interest (see Table F.2 in Appendix F for values). In writing Eq. 5.3.2,

the solution has been assumed to be dilute. The error in the model relative to the experiment can now be conveniently determined, as follows:

$$\begin{aligned}\Delta g_m^* &= \frac{g_{m,\text{model}}^*}{k_BT} - \frac{g_{m,\text{expt}}^*}{k_BT} \\ &= \ln\left(\frac{C_{cmc}^{\text{model}}[mM]}{1000\mathcal{M}_{H_2O}[M]}\right) - \ln\left(\frac{C_{cmc}^{\text{expt}}[mM]}{1000\mathcal{M}_{H_2O}[M]}\right) \\ &= \ln\left(\frac{C_{cmc}^{\text{model}}}{C_{cmc}^{\text{expt}}}\right), \end{aligned} \tag{5.3.3}$$

Values for $g_{m,\text{model}}^*$ and errors for each model are presented in Table 5.7.

The predicted cmc's in Table 5.6 are all based on results obtained for $25°C$, since the solubility models and density fitting used in modeling the surfactants were all based on experimental data at this temperature (see Appendix F). There are a few examples of longer chain sulfonates and sulfates in Table 5.6 which have reported cmc's at temperatures other that $25°C$. This is usually a result of the Krafft temperature being greater than $25°C$, indicating that the solubility limit of the surfactant is less than the cmc at that temperature (i.e., no aggregation in the bulk solution can occur).[12] However, these predictions can still be compared to this experimental data, if one recognizes that ionic surfactants (including the sulfonates and sulfates) do not typically exhibit a large dependence of the cmc on temperature. In Table 5.8 I collect some of the experimental data reported by Moroi and coworkers for sulfonates[240] and sulfates.[241] The final column contains values for $\Delta g_{m,\text{max}}^*$, which is the maximum difference in the effective modified free energy of micellization, corresponding to the two points with maximum difference in reported cmc. I compute this similarly to Eq. 5.3.3, but allowing for the small change in $\mathcal{M}_{H_2O}$ with temperature. That is,

$$\Delta g_{m,\text{max}}^* = \ln\left(\frac{C_{cmc}\left(T_1\right)/\mathcal{M}_{H_2O}(T_1)}{C_{cmc}(T_2)/\mathcal{M}_{H_2O}(T_2)}\right), \tag{5.3.4}$$

where T_1 is the temperature corresponding to the largest cmc in Table 5.8 for the surfactant of interest, and T_2 is the temperature corresponding to the smallest cmc. In examining the range of surfactants reported, it is clear that the g_m^* varies by at most only 0.22 k_BT units across an interval of $30°C$! This is approximately the er-

ror that can be expected by fitting a solubility model to linear and branched alkane solubility data, for example, as seen by the errors in Table F.9 in Appendix F. Since I am interested in predicting a cmc that is representative of each surfactant, to enable comparison between surfactant structures, and I am not focused on predicting the temperature-dependent behavior of these surfactants, in Table 5.6, I proceed in comparing my predictions for the cmc's at $25°C$ (which is not affected by the Krafft point, since this phase behavior is not modeled) to the experimental values, regardless of their measurement temperature. Note that, for nonionic surfactants such as the alkyl ethoxylates, there is a larger effect of temperature, due to dehydration of the ethoxylate head as temperature is increased.[242] Based on cmc data collected in Ref. [171], which demonstrates a decrease in cmc with an increase in temperature due to this dehydration effect, the variation in g_m^* is about $-0.77\ k_BT$ over a temperature range of $25°C$ for the surfactant $C_{12}E_6$ (with a cmc of 0.135 mM at $5°C$ and a cmc of 0.067 mM at $30°C$). However, since the experimental data for cmc's for nonionic surfactants in Section 5.4 were obtained at $25°C$, there are no concerns about a direct comparison to the predictions in those cases.

Table 5.6 demonstrates several important trends in the experimental cmc values, which I will first state as observations, and then describe below using the results of the MT theory. (1) For each type of surfactant in which the length of the alkyl chain is varied, the cmc decreases with increasing chain length. (2) For the ethoxysulfates (C_iE_jSUL), an increase in the number of ethoxylate units, j leads to a decrease in the cmc (in opposition to the trend observed for alkyl ethoxylates (C_iE_j) in Section 5.4.1). Finally, for fixed chain length, (3) the sulfonates (C_iSFN) exhibit a higher cmc than the sulfates (C_iSUL), (4) the sulfonates exhibit a higher cmc than the benzene sulfonates (C_iBZSFN), and (5) the sulfates exhibit a higher cmc than the ethoxysulfates.

Table 5.6: CMC Results for Linear Anionic Surfactants at $T = 25°C$

		CMC (mM)	MT Pred. CMC (mM)[a]				CSMT Pred. CMC (mM)[b]			
Surfactant	Ion	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_8SFN	Na+	155[243]	**190**	380	37	73	-	-	-	-
C_{10}SFN	Na+	41[243]	**34**	94	9.3	14	22.6±0.6 [16]	**70±2** [16]	7.1±0.7 [14]	9.6±0.5 [16]
C_{12}SFN	Na+	10[243] [35°C]	**6.6**	26	2.4	2.8	4.6±0.1 [9]	**19.2±0.4** [10]	1.73±0.03 [9]	1.96±0.04 [10]
C_{14}SFN	Na+	2.9[243] [45°C]	**1.4**	7.6	0.65	0.62	1.2±0.1 [5]	**6.5±0.3** [5]	0.504±0.004 [3]	0.460±0.003 [3]
C_{16}SFN	Na+	0.9[243] [52°C]	0.31	**2.2**	0.18	0.14	0.251±0.009 [5]	**1.96±0.07** [5]	0.146±0.003 [4]	0.109±0.002 [4]
C_8SUL	Na+	130[244]	**180**	370	34	67	-	-	-	-
C_{10}SUL	Na+	33[244]	**32**	93	8.6	13	**24.7±0.7** [12]	79±2 [10]	7.0±0.2 [12]	10.1±0.3 [12]
C_{12}SUL	Na+	8.2[244]	**6.3**	25	2.3	2.7	**4.9±0.2** [20]	20.8±0.8 [20]	1.86±0.08 [20]	2.10±0.08 [19]
C_{14}SUL	Na+	2.05[244]	**1.3**	7.4	0.61	0.58	**1.20±0.04** [9]	7±0.3 [10]	0.58±0.03 [8]	0.52±0.02 [7]
C_{16}SUL	Na+	0.45[244] [30°C]	**0.29**	2.1	0.17	0.13	**0.29±0.01** [9]	2.4±0.1 [10]	0.18±0.01 [9]	0.128±0.007 [8]
C_8BZSFN	Na+	11.1[245]	45	170	**31.7**	120	26.5±0.1 [4]	110.6±0.5 [4]	**15.51±0.05** [5]	6.114±0.006 [5]
C_{10}BZSFN	Na+	3.14[245] [50°C]	8.9	44	**8.0**	39	5.2±0.7 [10]	29±4 [10]	**3.8±0.6** [10]	1.5±0.1 [9]
C_{12}BZSFN	Na+	1.2[245] [60°C]	**1.9**	13	2.1	13	**1.1±0.2** [15]	8±2 [15]	**1.1±0.2** [14]	0.33±0.04 [12]
$C_{12}E_1$SUL	Na+	4.5[246]	**2.4**	11	0.94	1.0	**2.2±0.7** [5]	11±2 [5]	1.2±0.5 [5]	1.1±0.4 [5]
$C_{12}E_2$SUL	Na+	2.8;[246] 3.0[247]	**1.7**	8.6	0.62	0.67	**1.5±0.4** [5]	8±2 [5]	0.94±0.08 [5]	0.77±0.05 [5]
$C_{12}E_3$SUL	Na+	2.8[247]	**1.2**	6.6	0.38	0.41	0.7±0.2 [4]	**4±1** [4]	0.33±0.06 [4]	0.24±0.04 [2]

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] MT predictions do not have error bars, since the theory is deterministic based on user input.

[b] The error bars in the CSMT predictions represent the standard deviation in the predicted cmc arising from differences between the independent simulations run. The number of simulations is denoted in square brackets. Note that the four solubility models are applied to the same set of simulations, but difficulties with numerical convergence in certain cases may lead to some points being discarded.

Table 5.7: Free Energy of Micellization Comparison for Linear Anionic Surfactants at $T = 25°C$

		g_m^* (k_BT)[a]	MT Error (k_BT)[b]				CSMT Error (k_BT)[b]			
Surfactant	Ion	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_8SFN	Na+	-5.88	**0.20**	0.90	-1.43	-0.75	-	-	-	-
C_{10}SFN	Na+	-7.21	**-0.19**	0.83	-1.48	-1.07	-0.60	**0.53**	-1.75	-1.45
C_{12}SFN	Na+	-8.62[c]	**-0.42**	0.96	-1.43	-1.27	-0.78	**0.65**	-1.75	-1.63
C_{14}SFN	Na+	-9.85[c]	**-0.73**	0.96	-1.50	-1.54	-0.88	**0.81**	-1.75	-1.84
C_{16}SFN	Na+	-11.02[c]	-1.07	**0.89**	-1.61	-1.86	-1.28	**0.78**	-1.82	-2.11
C_8SUL	Na+	-6.05	**0.33**	1.05	-1.34	-0.66	-	-	-	-
C_{10}SUL	Na+	-7.42	**-0.03**	1.04	-1.34	-0.93	**-0.29**	0.87	-1.55	-1.18
C_{12}SUL	Na+	-8.82	**-0.26**	1.11	-1.27	-1.11	**-0.51**	0.93	-1.48	-1.36
C_{14}SUL	Na+	-10.20	**-0.46**	1.28	-1.21	-1.26	**-0.54**	1.23	-1.26	-1.37
C_{16}SUL	Na+	-11.72[c]	**-0.44**	1.54	-0.97	-1.24	**-0.44**	1.67	-0.92	-1.26
C_8BZSFN	Na+	-8.51	1.40	2.73	**1.05**	2.38	0.87	2.30	**0.33**	-0.60
C_{10}BZSFN	Na+	-9.77[c]	1.04	2.64	**0.94**	2.52	0.50	2.22	**0.19**	-0.74
C_{12}BZSFN	Na+	-10.73[c]	**0.46**	2.38	0.56	2.38	**-0.09**	1.90	**-0.09**	-1.29
$C_{12}E_1$SUL	Na+	-9.42	**-0.63**	0.89	-1.57	-1.50	**-0.72**	0.89	-1.32	-1.41
$C_{12}E_2$SUL	Na+	-9.86	**-0.53**	1.09	-1.54	-1.47	**-0.66**	1.01	-1.13	-1.33
$C_{12}E_3$SUL	Na+	-9.89	**-0.85**	0.86	-2.00	-1.92	-1.39	**0.36**	-2.14	-2.46

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] Effective value calculated from the experimental data in Table 5.6 using Eq. 5.3.2.

[b] Errors are computed using Eq. 5.3.3.

[c] These data are evaluated at temperatures other than $25°C$. See Table 5.6 for details.

Table 5.8: Experimental CMC Values as a Function of Temperature

Surfactant	Temperatures / CMC's (mM)					$\Delta g^*_{m,\max}$ (k_BT)
	$22°C$	$25°C$	$35°C$	$45°C$	$55°C$	
$NaC_{10}SFN$	39.4	38.8	40.1	40.6	41.6	0.08
	$35°C$	$40°C$	$45°C$	$50°C$	$55°C$	
$NaC_{12}SFN$	10.6	10.7	11.2	11.5	12.0	0.13
	$42.0°C$	$44.6°C$	$48.1°C$	$51.5°C$	$54.9°C$	
$NaC_{14}SFN$	2.9	3.0	3.1	3.3	3.5	0.19
	$25°C$	$30°C$	$40°C$	$50°C$	$55°C$	
NaC_8SUL	13.02	13.18	13.59	14.21	14.66	0.11
$NaC_{10}SUL$	3.30	3.29	3.37	3.55	3.69	0.10
$NaC_{12}SUL$	0.816	0.824	0.856	0.918	0.961	0.15
$NaC_{14}SUL$	0.205	0.208	0.222	0.243	0.259	0.22

Each of these five observations is reproduced by the four solubility models considered, with the sole exception of the KEKSS model and observation (4). As seen in Table F.8 of Appendix F, the transfer contributions of methylene and methyl groups in alkylarenes are reduced relative to those for the same groups in paraffins. While this may yield better results for solubility predictions for low molecular weight compounds such as toluene, ethyl benzene, and *n*-propyl benzene, it is clearly not appropriate for molecular-thermodynamic modeling of the micellization behavior of surfactants based on these compounds: instead of a drop of a factor of around 10 in the cmc of benzene sulfonates relative to the sulfonates, the KEKSS model predicts an increase by a factor of approximately 2 to 4.

The model exhibiting the best, most consistent agreement in MT predictions is the MHA model. The MHA2 model substantially overpredicts the cmc in many cases, while the KWB tends to underpredict. Recall that all four models were developed to replicate solubility data for linear and branched oils. In Table 5.9, I provide a comparison of the prediction of the four models for the transfer free energy of dodecane and a C_{12} tail to help explain the discrepancy between accurate prediction of oil properties and accurate prediction of tail properties. Note that there is no experimental verification possible for the transfer of a C_{12} tail, which is not a whole molecule. One can see that the replacement of a methyl group with a methylene group in moving from dodecane to a C_{12} tail has a clear model-dependent effect, with

Table 5.9: Transfer Free Energies (k_BT) for Dodecane and a C_{12} Tail

Model	Dodecane	C_{12} Tail
Experiment	-21.662	
MHA	-22.014	-19.989
MHA2	-20.450	-17.395
KWB	-20.994	-21.539
KEKSS	-21.830	-21.436

KWB exhibiting a more negative g_{tr} value for the C_{12} tail, and the other three models exhibiting a less negative value. Note that MHA2 does not fit dodecane as well as the other models – this is due to fitting of MHA2 to both linear and branched oils (MHA was fit to linear oils first, then to branched to obtain parameters for CH and C groups, which are not relevant here) and the inclusion of tetradecane and hexadecane in the fitting (which exhibit anomalous solubilities in comparison with the shorter linear alkanes, as discussed in Section F.5.5 of Appendix F, which have larger solubilities than expected – that is, less negative g_{tr} values). It is important to note that, since the true transfer free energy for a C_{12} tail is unknown, only a large study such as this one can elucidate which model is preferable. This is also true of alkyl benzene tails: in Table F.10 in Appendix F, the KWB model predicts the aqueous solubilities of various phenyldecanes and phenyldodecanes with the best accuracy. In Table 5.6, it is, in fact, KWB that predicts the best result for C_8BZSFN and C_{10}BZSFN, but the MHA model has comparable values, and is slightly closer in the prediction of C_{12}BZSFN. This is encouraging, because it seems to indicate that one model – the MHA model – shows promise for modeling the full set of anionic surfactants. Note that, with few exceptions, the MHA model predicts a value for g_m^* to within 1 k_BT of the experimental values, in contrast to the other models, which exhibit more marked deviation. This seems to indicate that the right balance between the amount of transfer free energy to ascribe to methyl groups versus methylene groups has been achieved (under the assumption that the other free-energy contributions do not provide a systematic under- or over-prediction).

Using the MT results for the MHA model, it is possible to understand the five observations related to the experimental measurements, as discussed above.

Observation (1), that cmc's decrease with an increase in chain length is a well-celebrated result of the hydrophobic effect.[34] Each additional methylene group in the surfactant tail increases the perturbation of the hydrogen-bonding network of water, increasing the driving force for removal of the tail from water through aggregation (or removal to surfaces).

Observation (2), that an increase in the ethoxylate spacer in the ethoxysulfates decreases the cmc, is a function of: (a) the constancy of the modeled head area for short ethoxylate units (due to relative inflexibility in comparison to long ethoxylate chains), and (b) the reduction in the charging free-energy penalty due to the increase in the distance-to-charge from the micelle core-water interface to the average charge location afforded by the ethoxylate spacer itself. Point (a) means that the steric free-energy penalty is not increasing in going from E_1 to E_3, while point (b) means that less of the negative transfer free energy will be offset (hence, the lower cmc predicted). The specific head properties can be reviewed in Table 5.3 in Section 5.2.1 above.

Observation (3), that the sulfonates exhibit a higher cmc than the sulfates for fixed chain length, has the same explanation as Observation (2). That is, the sulfonate group possesses a charge closer to the micelle core-water interface, which increases the electrostatic penalty. This is a small effect (about 0.2 k_BT, as calculated from the experimental cmc's and application of Eq. 5.3.3), but detectable with the MT model.

Observation (4), that the sulfonates have a higher cmc than the benzene sulfonates, is a consequence of the head/tail identification obtained from the computer simulations, which assigns three of the benzene carbon groups to the tail (in addition to the alpha methylene). For the MHA solubility model, this makes the transfer free energy more negative by $-3.119\ k_BT$. All the other free-energy contributions are also affected. In particular, the interfacial tension of alkyl benzenes is less than that for the alkanes by about 10 dyn/cm (see Ref. [248] for a compilation of values), meaning less of an interfacial free-energy penalty. However, the packing penalty increases substantially due to the confinement of the ring to the interface, leading to a 1.4 k_BT offset.

Observation (5), that the sulfates have a higher cmc than the ethoxysulfates is mainly explained by the difference in head/tail identification: the sulfates (C_iSUL) have a C_{i-1} tail, while the ethoxysulfates (C_iE_jSUL) have a C_i tail brought about by the increased distance between the charge group and the alpha carbon due to the ethoxylate spacer. There is also, as with Observation (2), an effect due to having the charge positioned further from the micelle core-water interface (where the charge is smeared across a larger surface area per surfactant during the electrostatic charging process).

Finally, the CSMT results in Table 5.6 uniformly point to a reduction in the predicted cmc relative the MT predictions using the same solubility model. Recall that the CSMT approach integrates computer simulation results into the MT framework in two ways: (i) through a means for performing the surfactant head/tail identification, and (ii) through the calculation of an updated transfer free energy, $g_{tr,CSMT}$, as described in Section 3.2.2 of Chapter 3. Since the first point is often incorporated into the MT predictions as well (as with several of the entries in Table 5.1). Therefore, the second point is the main area in which the MT and CSMT predictions will differ. Equation 3.2.12 in Chapter 3 explores the difference between the MT g_{tr} ($g_{tr,MT}$) and $g_{tr,CSMT}$. There are two contributions, one due to the partial dehydration of hydrophobic head groups (such as the alpha methylene excluded from consideration in the MT modeling in many ionic cases), and the other due to the subtle change in the structure of water about the various groups in the micelle core relative to the monomer state. The first contribution is typically negative (unless the group contribution of a group to the transfer free energy is positive – an example is an aromatic C group in the MHA model, with a transfer free energy contribution of $+0.914\ k_BT$). It is not clear that the second contribution (as formulated in Eq. 3.2.12) is guaranteed to have a particular sign, but it is clear from Table 5.7 that, at the least, it does not reverse the sign of the first contribution, such that $g_{tr,CSMT} - g_{tr,MT}$ is consistently negative, and the cmc predicted by CSMT is less than that for MT. As a result, models which tend to underpredict (such as the KWB and KEKSS models) will exhibit increased deviation from the experimental cmc. The MHA2 model, which overpredict the cmc

in the MT modeling for all anionic surfactants studied, is naturally improved. For the MHA model, recommended for MT modeling, the predictions are not greatly affected, except in the case of the benzene sulfonates, where the predictions are actually improved. However, the average root mean square error per surfactant (calculated from the error values for the MHA MT and CSMT models in Table 5.7 for which both values were available) increases slightly from 0.71 k_BT for the MT framework to 0.76 k_BT for the CSMT framework, indicating a (very) slight reduction in prediction performance. One further point which recommends the MT model in this case, however, is that Observation (1) is correctly predicted, whereas the CSMT inverts the relationship between the sulfonate and sulfate surfactants. Since there are various sources for error in the simulations (force field parameters, eccentricity in the micelle when making spherical rdf calculations), it is perhaps not surprising that an effect on the order of 0.2 k_BT may not be accurately captured. Since the MT model is not exposed to these additional source of errors, this is one recommendation for it. However, note the value of the computer simulations in ascertaining the head/tail identification for the benzene sulfonates: the MT model is limited by the rules of the thumb necessary to perform modeling. Accordingly, it is important to consider some measure of input from the computer simulations, even if the full CSMT framework is not always necessary to achieve quality predictions.

5.3.2 Cationic Surfactants

Tables 5.10 and 5.11 summarize the cmc results and deviations in predicted modified free energies of micellization for the cationic surfactants depicted in Figure 5-2. Both MT and CSMT predictions are shown. Note that not all systems have been simulated, since short chain cationic surfactant micelles (C_8 and C_{10}) have a tendency to eject monomers during the time scale of the simulation (tens of nanoseconds), complicating analysis. This ejection is likely due to their larger, bulkier head groups, which prevents the formation of a compact core, making it easier for any one chain to diffuse out of the micelle. I have chosen to analyze only whole micelles in this Thesis, to avoid complications introduced by the presence of monomers in the simulation cell

when calculating radial distribution functions for head/tail identification and when computing hydrating contacts.

There are several observations to be made based on the experimental results. (1) Each type of surfactant exhibits a decreasing cmc with increasing chain length. (2) For a fixed chain length, the alkyl trimethylammoniums (C_iTMA) have a larger cmc than the 1-alkyl pyridiniums (C_iPYR), which have a larger cmc than the 1-methyl,4-alkyl pyridiniums (C_iMPYR). (3) In comparing the alkyl trimethylammoniums to the alkyl sulfonates and alkyl sulfates in Table 5.6, the trimethylammoniums have a consistently higher cmc across all chain lengths.

Observation (1) is explained, as in Section 5.3.1, by the hydrophobic effect.[34] Each successive methylene group in a chain imparts a negative contribution to the overall transfer free energy of the surfactant tail, resulting in an increased driving force for micellization. This observation is reproduced in all four solubility models, although the MHA model provides the most accurate predictions (similarly to the anionic surfactants).

Observation (2) is not completely reproduced by the models. In fact, all four solubility models predict an *increase* in the cmc when selecting a 1-alkyl pyridium surfactant over a trimethylammonium surfactant of equal chain length. The MHA model correctly predicts the relationship between the trimethylammoniums and 1-methyl, 4-alkyl pyridiniums, as well as the 1-alkyl pyridinium and 1-methyl, 4-alkyl pyridiniums (which the MHA2 model also correctly predicts). However, the remaining solubility models do not.

To try to understand this situation, it is useful to examine the head/tail identifications used in the MT modeling, as summarized in Table 5.1. The traditional rule of thumb is used for the C_iTMA and C_iPYR surfactants (a C_{i-1} tail), while the tail for the C_iMPYR surfactants is given as B_0C_i (which indicates the full alkyl chain, plus consideration of the aromatic ring in the head region when packing). Since the KWB and KEKSS models are affected by nature of the tail (whether alkyl or alkylaryl), the B_0C_i identification triggers a change in the transfer free energy assigned to the chain, which is the reason neither the KWB nor KEKSS models correctly identify the

relationship between C_iPYR and C_iMPYR cmcs for fixed chain length. The correct identification of this relationship by the MHA and MHA2 models can be attributed to the extra methylene group considered to be in the tail of the C_iMPYR surfactants, which decreases the cmc. Additionally, the charge of a C_iMPYR surfactant is further from the micelle core-water interface than the charge of a C_iPYR surfactant. This also supports the trend. In examining the C_iTMA and C_iPYR surfactants, which have the same MT tail identification and very similar distance-to-charge (but different head lengths), one clear difference is the head area used in modeling these two surfactants: 28.4 Å^2 for C_iTMA versus 35.0 Å^2 for C_iPYR. To test the effect of head area on the results, I made additional MT predictions for the C_iPYR surfactants using the head area of C_iTMA and the MHA model, finding that the predicted cmc's for C_8 to C_{16} are 310, 53, 10, 2.1, and 0.44 mM, respectively. By comparing these values to the cmc values for C_iTMA in Table 5.10, one can see that the trend is still not captured. Since this is a comparison between aromatic and non-aromatic surfactants, it is possible that there is an unexplained effect, but this could also be simply due to errors in the models. Note that the absolute predictions are quite good for both classes of surfactants. The observed trend corresponds to a difference of about 0.4 k_BT (as seen by comparing g_m^* values in Table 5.11). This difference is within the errors relative to the predictions themselves.

Observation (3) is correctly reproduced by all four solubility models using the MT approach. It is not consistently reproduced using the CSMT approach (e.g., the MHA model predicts a higher cmc for the sulfonates), likely due to the consideration of the methyl groups in the head as having their usual alkyl group contribution to the transfer free energy (modulated by the fractional degrees of hydration of these groups, which are not in the tail). Further study is needed to evaluate what the correct group contributions of these methyls should be in a CSMT setting. The usual contribution is found by fitting to oils larger than pentane, but it is not clear that use of this value is appropriate for the case of a single methyl attached to the charged nitrogen, since the electrostatic potential about this methyl is affected by the local charge, as can be seen by ab initio calculations in a software package such as Gaussian 03.[234]

Table 5.10: CMC Results for Linear Cationic Surfactants at $T = 25°C$

		CMC (mM)	MT Pred. CMC (mM)[a]				CSMT Pred. CMC (mM)[b]			
Surfactant	Ion	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_8TMA	Br-	261[249]	**287**	610	55	110	-	-	-	-
C_{10}TMA	Br-	67.4[250]	**49**	150	13	20	-	-	-	-
C_{12}TMA	Br-	15.4[250]	**9.5**	38	3.4	4.0	3.9±0.3 [16]	**11.5±0.9** [17]	1.87±0.07 [12]	2.1±0.1 [11]
C_{14}TMA	Br-	3.78[250]	**2.0**	11	0.89	0.85	0.692±0.009 [5]	**3.28±0.03** [5]	0.376±0.004 [5]	0.320±0.004 [5]
C_{16}TMA	Br-	0.952[250]	**0.41**	3.1	0.24	0.18	0.0891±0.0009 [3]	**0.83±0.02** [5]	0.070±0.002 [4]	0.040±0.001 [4]
C_8PYR	Br-	193[171]	450	961	84	**170**	-	-	-	-
C_{10}PYR	Br-	44[251]	76	231	20	**30**	-	-	-	-
C_{12}PYR	Br-	10[251]	**14**	59	5.1	6.1	**7.2±0.4** [3]	26.8±0.8 [5]	2.53±0.09 [5]	4.0±0.1 [5]
C_{14}PYR	Br-	2.7[251]	**2.8**	16	1.2	1.2	**1.329±0.006** [4]	6.8±0.2 [5]	0.54±0.03 [5]	0.69±0.01 [5]
C_{16}PYR	Br-	0.64[251]	**0.58**	4.3	0.33	0.25	**0.244±0.006** [5]	1.84±0.02 [4]	0.111±0.003 [4]	0.111±0.003 [4]
C_8MPYR	I-	42.7[61] [30°C]	**270**	730	440	1700	-	-	-	-
C_{10}MPYR	I-	10.65[61] [30°C]	**47**	170	92	460	**13.1±0.1** [5][c]	49.6±0.5 [5][c]	24.1±0.2 [5][c]	5.48±0.04 [5][c]
C_{12}MPYR	I-	2.5[61] [30°C]	**9.1**	45	21	130	**2.46±0.04** [7][c]	12.5±0.2 [7][c]	5.44±0.08 [7][c]	1.09±0.02 [4][c]

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] MT predictions do not have error bars, since the theory is deterministic based on user input.

[b] The error bars in the CSMT predictions represent the standard deviation in the predicted cmc arising from differences between the independent simulations run. The number of simulations is denoted in square brackets. Note that the four solubility models are applied to the same set of simulations, but difficulties with numerical convergence in certain cases may lead to some points being discarded.

[c] Simulations run at $30°C$.

Table 5.11: Free Energy of Micellization Comparison for Linear Cationic Surfactants at $T = 25°C$

		g_m^* (k_BT)[a]	MT Error (k_BT)[b]				CSMT Error (k_BT)[b]			
Surfactant	Ion	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_8TMA	Br-	-5.35	**0.09**	0.85	-1.56	-0.86	-	-	-	-
C_{10}TMA	Br-	-6.71	**-0.32**	0.80	-1.64	-1.21	-	-	-	-
C_{12}TMA	Br-	-8.19	**-0.48**	0.90	-1.51	-1.35	-1.37	**-0.29**	-2.10	-1.99
C_{14}TMA	Br-	-9.59	**-0.64**	1.07	-1.45	-1.49	-1.70	**-0.14**	-2.30	-2.47
C_{16}TMA	Br-	-10.97	**-0.84**	1.18	-1.38	-1.67	-2.37	**-0.14**	-2.61	-3.17
C_8PYR	Br-	-5.66	0.85	1.61	-0.83	**-0.13**	-	-	-	-
C_{10}PYR	Br-	-7.14	0.55	1.66	-0.79	**-0.38**	-	-	-	-
C_{12}PYR	Br-	-8.62	**0.34**	1.77	-0.67	-0.49	**-0.33**	0.99	-1.37	-0.92
C_{14}PYR	Br-	-9.93	**0.04**	1.78	-0.81	-0.81	**-0.71**	0.92	-1.61	-1.36
C_{16}PYR	Br-	-11.37	-0.10	1.90	**-0.66**	-0.94	**-0.96**	1.05	-1.75	-1.75
C_8MPYR	I-	-7.17[c]	**1.84**	2.84	2.33	3.68	-	-	-	-
C_{10}MPYR	I-	-8.55[c]	**1.48**	2.77	2.16	3.77	**0.21**	1.54	0.82	-0.66
C_{12}MPYR	I-	-10.00[c]	**1.29**	2.89	2.13	3.95	**-0.02**	1.61	0.78	-0.83

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] Effective value calculated from the experimental data in Table 5.10 using Eq. 5.3.2.

[b] Errors are computed using Eq. 5.3.3.

[c] These data are evaluated at temperatures other than $25°C$. See Table 5.10 for details.

In examining all the predictions, it appears that the MT predictions of the MHA model offer the best agreement with experiment for the majority of the structures. The CSMT predictions of the MHA model provide a substantial improvement in the accuracy of the 1-methyl, 4-alkyl pyridiniums, but the predictions for the trimethylammoniums are worsened. In comparing the KWB and KEKSS results for the 1-alkyl pyridinium and 1-methyl, 4-alkyl pyridiniums, it is clear that the solubility models which exhibit a discontinuous change in the group contributions based on the total chemical nature of the surfactant tail hinder the prediction of qualitative trends between disparate chemical structures.

5.3.3 Zwitterionic Surfactants

Tables 5.12 and 5.13 summarize the cmc results and deviations in predicted modified free energies of micellization for the sulfobetaine surfactants considered in Figure 5-2. Notice that the CSMT predicted cmc's are listed in micromolar units and deviate significantly from the predicted experimental values. For this surfactant, the alkyl spacer and methyls attached to the ammonium groups were considered to have the usual transfer free-energy contributions of alkyl groups. Fractional degrees of hydration for these groups are between 0.60 and 0.88, but because there are several such groups, the overall contribution to $g_{tr,CSMT}$ is significant enough to greatly reduce the cmc. As was mentioned in Section 5.3.2 with the trimethylammonium case, it appears likely that these alkyl groups should not be treated as conventional alkanes for the purpose of computing $g_{tr,CSMT}$.

In this set of structures, it is apparent that there is a tendency to underpredict the experimental cmc even using the MT framework. In fact, the MHA2 model provides the best predictions, but this is fortuitous: since the MHA2 model consistently overpredicts all the anionic and cationic cmc's (see Tables 5.7 and 5.11, respectively), the systematic underprediction simply brings the MHA2 model closer to experiment.

One convincing explanation for the trend towards underprediction involves the head group parameters presented in Table 5.3. These were calculated for a fully-extended sulfobetaine group. However, during simulations of the sulfobetaine surfac-

tants, it becomes clear that the sulfobetaine head is most frequently in a state of association between the sulfonate group and the oppositely-charged ammonium groups in adjacent surfactant heads. Figure 5-5 shows a simulation snapshot of a micelle composed of C_{12}SBT(4) surfactants. Less than a quarter of the surfactant heads are extended. If one compares the MT predictions of C_{12}SBT(3) and C_{12}SBT(4), which differ only in terms of charge separation and head length, it is clear that decreasing the charge separation (and head length) increases the cmc, due to an increased electrostatic penalty. Accordingly, if the head groups are not fully-extended, then it is expected that the cmc should be larger than predicted. Additionally, even if there is cancellation of some charge in the theoretical charge layers due to positive and negative charge association, the association should also increase the expected head area representing steric interactions, since this head area is calculated assuming full extension.

These considerations may also help explain another observation with the experimental data: there is apparently an increase in cmc for C_{12}SBT(4) surfactant relative to C_{12}SBT(3). If the sulfobetaines are in a more collapsed state, then a larger spacer may lead to a larger effective head area (similar to the ethoxylates treated in Section 5.4.1). If the two types of sulfobetaines have a similar distribution of charges due to this collapse, then the electrostatic contribution will be similar, meaning that it would be the steric free-energy contribution that would explain the trend. Note that the MT predictions were made using conventional geometric arguments for a fully-extended chain, and the head areas are fixed to the same (cross-sectional) values. As a result, this observed trend is not captured. Interestingly, the trend of increasing cmc with an increase in the spacer length is not consistent across chain lengths: the values reported by Qu and coworkers[252] indicate that the cmc's of C_{12}SBT(4) and C_{18}SBT(4) are greater than their SBT(3) counterparts, while the cmc's of C_{14}SBT(4) and C_{16}SBT(4) are *less* than their SBT(3) counterparts. In carefully examining their surface tension versus surfactant concentration data, it appears that there is enough variability that this approximately 0.3 k_BT difference due to the spacer length may also be within experimental error.

Table 5.12: CMC Results for Linear Zwitterionic Surfactants at $T = 25°C$

	CMC (mM)	MT Pred. CMC (mM)[a]				CSMT Pred. CMC (μM)[b]			
Surfactant	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_{12}SBT(3)	1.61[252]	0.42	**4.13**	0.074	0.099	6 ± 2 [14]	**50±20** [14]	1.9±0.5 [14]	1.9±0.6 [14]
C_{14}SBT(3)	0.377[252]	0.029	**0.53**	0.0074	0.0068	0.34 ± 0.06 [5]	**6±1** [5]	0.19 ± 0.04 [5]	0.13 ± 0.02 [5]
C_{16}SBT(3)	0.029[252]	0.0020	**0.07**	0.00074	0.00047	0.019 ± 0.003 [5]	**0.62±0.08** [5]	0.016 ± 0.002 [5]	0.007 ± 0.001 [5]
C_{12}SBT(4)	2.2[252]	0.098	**0.97**	0.017	0.023	0.4 ± 0.1 [15]	**4±1** [15]	0.14 ± 0.05 [15]	0.13 ± 0.04 [15]

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] MT predictions do not have error bars, since the theory is deterministic based on user input.

[b] The error bars in the CSMT predictions represent the standard deviation in the predicted cmc arising from differences between the independent simulations run. The number of simulations is denoted in square brackets. Note that the four solubility models are applied to the same set of simulations, but difficulties with numerical convergence in certain cases may lead to some points being discarded.

Table 5.13: Free Energy of Micellization Comparison for Linear Zwitterionic Surfactants at $T = 25°C$

	g_m^* (k_BT)[a]	MT Error (k_BT)[b]				CSMT Error (k_BT)[b]			
Surfactant	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_{12}SBT(3)	-10.45	-1.34	**0.94**	-3.08	-2.79	-5.59	**-3.47**	-6.74	-6.74
C_{14}SBT(3)	-11.90	-2.56	**0.34**	-3.93	-4.02	-7.01	**-4.14**	-7.59	-7.97
C_{16}SBT(3)	-14.46	-2.67	**0.88**	-3.67	-4.12	-7.33	**-3.85**	-7.50	-8.33
C_{12}SBT(4)	-10.13	-3.11	**-0.82**	-4.86	-4.56	-8.61	**-6.31**	-9.66	-9.74

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] Effective value calculated from the experimental data in Table 5.12 using Eq. 5.3.2.

[b] Errors are computed using Eq. 5.3.3.

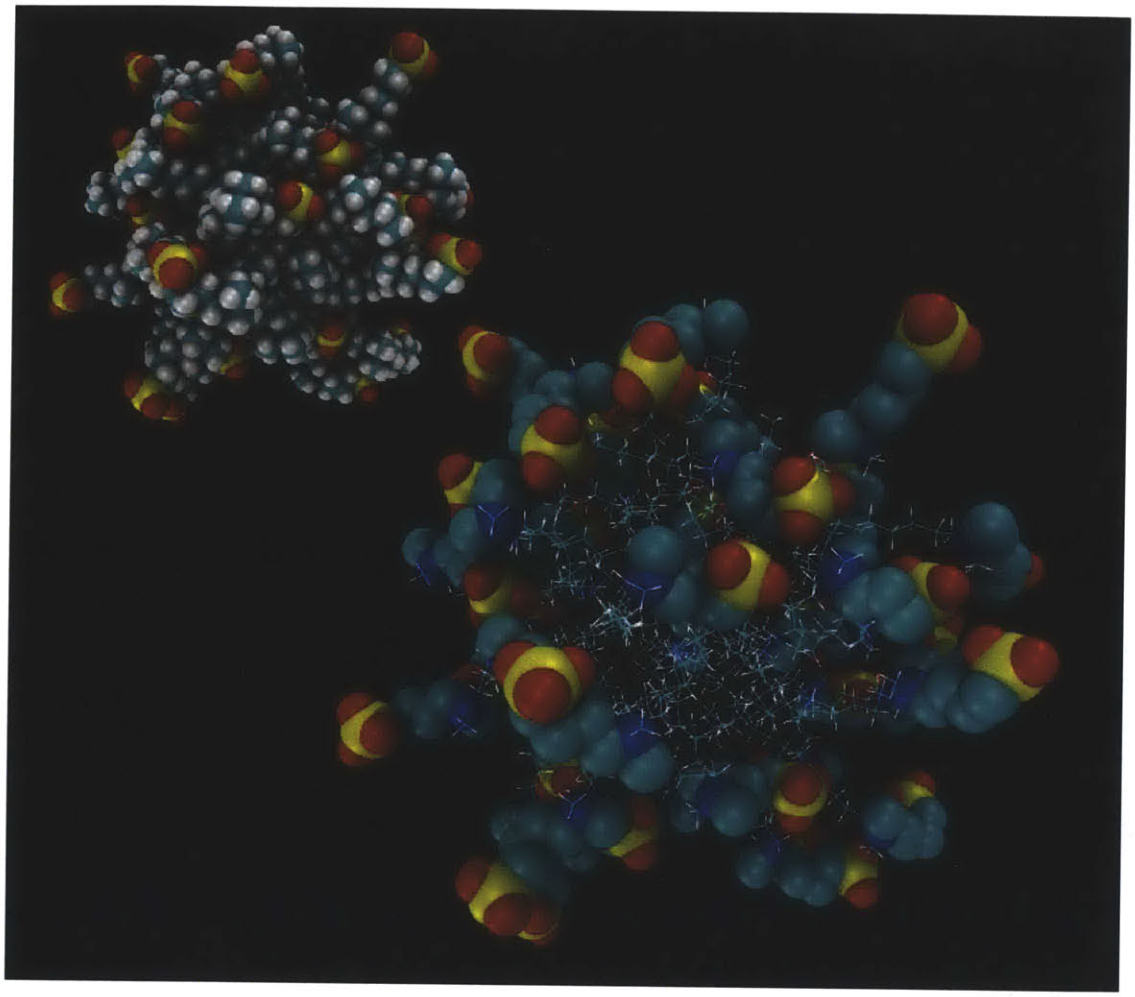


Figure 5-5: Simulation snapshot of a C_{12}SBT(4) sulfobetaine surfactant micelle composed of 36 monomers. In the upper left corner, all atoms are shown using a van der Waals representation. In the lower right, only the backbone atoms of the sulfobetaine are shown, to highlight the orientation of the sulfonate group relative to the alkyl ammonium group. Note that, while a few sulfobetaines are fully extended into solution (water not shown for clarity), the majority are condensed on the surface, with close proximity between the negatively-charged sulfonate and the positively-charge ammonium. (Color code: red – oxygen, yellow – sulfur, dark blue – nitrogen, teal – carbon, white – hydrogen.)

5.4 Modeling Results for Nonionic Surfactants

5.4.1 Ethoxylated Surfactants

Tables 5.14 and 5.15 summarize the cmc results and deviations in predicted modified free energies of micellization for the ethoxylate surfactants considered in Figure 5-4. In this series, there are only two points to consider: (i) the length of the surfactant chain, and (ii) the length of the ethoxylate chain. Here, I have selected alkyl chain lengths of C_8 to C_{12} and ethoxylate chain lengths of E_4 to E_{10}. As with the charged surfactants throughout Section 5.3, the experimentally-observed trend with increasing chain length is clear: additional methylene groups decrease (make more negative) the transfer free energy of the surfactant tail, leading to a decrease in the cmc. As for the effect of the ethoxylate chain length fixing the alkyl chain length, the trend is also consistent: an increase in the number ethoxylate units increases the cmc. Since these surfactants are nonionic, there is not an electrostatic component to the modeling. In the MT thought process, the only remaining head-related free-energy contribution is the steric free energy. As a result, the increase in ethoxylate head length can be viewed as leading to an increase in the head area of these surfactants,[81] as modeled in Eq. 5.2.1 in Section 5.2.2.

As with the charged systems, the MHA model provides the best agreement with experiment for most short-chain ethoxylate structures. For C_8E_8, $C_{10}E_{10}$, and $C_{12}E_{10}$, one can see an overprediction with this model. Although not shown, as the ethoxylate length continues to increase, the scaling relationship in Eq. 5.2.1 leads to a head area that crowds the interface, causing the steric free energy to diverge. However, surfactants such as $C_{12}E_{30}$ are physically realizable and exhibit cmcs.[171] The resolution of this issue is to recognize that these chains interpenetrate and are not well-represented as hard disks at long chain length (alternatively, hard disk areas can always be used as a fitting parameter, but will not follow the scaling law of Eq. 5.2.1). A theory for end-grafted polymers on surfaces exhibiting curvature, such as that due to Carignano and Szleifer,[253] and very much in the vein of the packing theory presented in Chapter 4 is thus recommended when considering surfactants with high ethoxylate content. I

have made progress in this area, but much remains to be done. This topic is addressed again briefly in Chapter 7.

In Table 5.14, as in Section 5.3, the CSMT approach leads to a decrease in the transfer free energy that causes substantial deviation in the prediction (but brings the MHA2 model into good agreement with experiment, since this model tends to over-predict in MT modeling). In this particular case, the tail identification for these C_iE_j surfactants is C_i. Further, no transfer contributions are assigned to the ether oxygens or ether methylene groups. In work not shown, but corroborated by the predictions of Stephenson and coworkers,[109] inclusion of transfer free energy contributions for these terms lead to even larger deviations. (Although I have investigated the KWB and KEKSS models for their use of a positive contribution for the ether oxygens, this positive contribution is more than offset by the intervening ether methylene groups' negative values. Accordingly, the ethoxylate head, when treated in this manner provides an additional negative contribution to the free energy.) Since the only groups with non-zero transfer free-energy contributions then are the same ones already fully represented in the MT modeling, the explanation for the deviation arises from the water structure term in the CSMT equation for the free energy of hydration (see Eq. 3.2.8 in Chapter 3) and the fact that certain tail identifications (see Table 5.4) involve the first ether oxygen, and, in certain cases, the first ether methylene. This leads to larger realizable micelle sizes and a slightly smaller head group (in these cases, the scaling law of Eq. 5.2.1 is applied with a fraction value for the number of ethoxylate units – with 1/3 applied for each CH_2 or O group in the head). Since the interfacial free-energy contribution per surfactant scales inversely to the radius,[35] the overall effect of these longer tails is to lead to a decrease in the cmc, even if there are no additional transfer contributions.

Table 5.14: CMC Results for Linear Nonionic Surfactants at $T = 25°C$

	CMC (mM)	MT Pred. CMC (mM)[a]				CSMT Pred. CMC (mM)[b]			
Surfactant	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_8E_4	8.0^{254}	**4.7**	19	0.48	1.1	1.10±0.03 [9]	**4.4±0.1** [9]	0.114±0.004 [9]	0.263±0.008 [9]
C_8E_6	9.9^{171}	**11**	45	1.2	2.7	3.2±0.9 [9]	**11±2** [9]	0.31±0.07 [9]	0.8±0.2 [9]
C_8E_8	10.4^{254}	34	137	3.6	**8.2**	**9±2** [8]	29±6 [8]	0.8±0.2 [8]	2.1±0.5 [8]
$C_{10}E_4$	0.5^{255}	**0.29**	2.1	0.044	0.069	0.055±0.002 [5]	**0.40±0.01** [5]	0.0081±0.0003 [5]	0.0131±0.0005 [5]
$C_{10}E_6$	0.9^{171}	**0.67**	4.9	0.099	0.16	0.12±0.02 [15]	**0.8±0.1** [15]	0.016±0.002 [15]	0.028±0.004 [15]
$C_{10}E_8$	1.1^{256}	**2.0**	15	0.30	0.47	0.29±0.07 [15]	**1.9±0.4** [15]	0.040±0.009 [15]	0.07±0.02 [15]
$C_{10}E_{10}$	1.24^{256}	5.9	43	0.88	**1.4**	**0.55±0.02** [5]	3.8±0.1 [5]	0.079±0.002 [5]	0.132±0.004 [5]
$C_{12}E_6$	0.064^{171}	**0.04**	0.52	0.0083	0.0092	0.0035±0.0005 [15]	**0.047±0.006** [15]	0.0007±0.0001 [15]	0.0008±0.0001 [15]
$C_{12}E_8$	0.0904^{171}	**0.11**	1.5	0.024	0.027	0.010±0.002 [15]	**0.12±0.03** [15]	0.0020±0.0005 [15]	0.0023±0.0006 [15]
$C_{12}E_{10}$	0.136^{257}	0.33	4.4	0.070	**0.077**	0.018±0.001 [5]	**0.24±0.01** [5]	0.0038±0.0002 [5]	0.0042±0.0002 [5]

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] MT predictions do not have error bars, since the theory is deterministic based on user input.

[b] The error bars in the CSMT predictions represent the standard deviation in the predicted cmc arising from differences between the independent simulations run. The number of simulations is denoted in square brackets. Note that the four solubility models are applied to the same set of simulations, but difficulties with numerical convergence in certain cases may lead to some points being discarded.

Table 5.15: Free Energy of Micellization Comparison for Linear Nonionic Surfactants at $T = 25°C$

	g_m^* (k_BT)[a]	MT Error (k_BT)[b]				CSMT Error (k_BT)[b]			
Surfactant	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_8E_4	-8.84	**-0.53**	0.86	-2.81	-1.98	-1.98	**-0.60**	-4.25	-3.42
C_8E_6	-8.63	**0.11**	1.51	-2.11	-1.30	-1.13	**0.11**	-3.46	-2.52
C_8E_8	-8.58	1.18	2.58	-1.06	**-0.24**	**-0.14**	1.03	-2.56	-1.60
$C_{10}E_4$	-11.61	**-0.54**	1.44	-2.43	-1.98	-2.21	**-0.22**	-4.12	-3.64
$C_{10}E_6$	-11.03	**-0.30**	1.69	-2.21	-1.73	-2.01	**-0.12**	-4.03	-3.47
$C_{10}E_8$	-10.83	**0.60**	2.61	-1.30	-0.85	-1.33	**0.55**	-3.31	-2.75
$C_{10}E_{10}$	-10.71	1.56	3.55	-0.34	**0.12**	**-0.81**	1.12	-5.79	-2.24
$C_{12}E_6$	-13.67	**-0.47**	2.09	-2.04	-1.94	-2.91	**-0.31**	-4.52	-4.38
$C_{12}E_8$	-13.32	**0.20**	2.81	-1.33	-1.21	-2.20	**0.28**	-3.81	-3.67
$C_{12}E_{10}$	-12.92	0.89	3.48	-0.66	**-0.57**	-2.02	**0.57**	-3.58	-3.48

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] Effective value calculated from the experimental data in Table 5.14 using Eq. 5.3.2.

[b] Errors are computed using Eq. 5.3.3.

5.4.2 Sugar-Based Surfactants

Tables 5.16 and 5.17 summarize the cmc results and deviations in predicted modified free energies of micellization for the sugar-based surfactants considered in Figure 5-4. As with the ethoxylated surfactants in Section 5.4.1, the sole difference in the MT modeling of these surfactants, when considering a fixed alkyl chain length, is selection of the head area. As such, there is no ready way to distinguish between stereochemical isomers, such as between the α and β anomeric forms of glucose and maltoside presented here, or between β-glucoside and β-galactoside, using the MT approach. However, the experimental cmc results clearly indicate differences between these surfactants. Nilsson and coworkers clarified the effect of the anomeric form of the alkyl glucosides on the phase diagram of these surfactants.[258] They concluded that the α anomeric form yielded a higher Krafft point due to increased crystal stability. For the maltosides, Dupuy and coworkers found that the anomeric forms had detectable effects on the packing of the monomers, influencing the asphericity of the micelles.[237] Other researchers have also studied the hydration states of these sugar-based surfactants as a function of stereochemistry. For example, Chong and coworkers concluded, based on molecular dynamics simulations, that the glucosides are more hydrated than the galactosides.[100]

As with the charged and ethoxylated surfactants already discussed in Sections 5.3 and 5.4.1, the MHA solubility model is capable of predicting well the micellization behavior of octyl galactoside (C_8GAL) and various glucosides (C_iGLU) and maltosides (C_iMAL). As mentioned, α and β anomeric forms of these surfactants have the same cmc prediction, and β-C_8GAL has the same cmc as β-C_8GLU. What is interesting is that the α and β forms *are* distinguishable using the CSMT approach, as seen by the cmc increase in the β anomeric form relative to the α anomeric form for both C_8GLU and C_{12}MAL. However, since comparison of β-C_8GAL and β-C_8GLU demonstrates that the trend in this particular stereochemistry is not captured, it appears that these systems require further study. Since the key input provided by the simulation for the MT modeling is an updated transfer free energy for the surfactant tail, it makes

sense that the stereochemistry involving the surfactant tail attachment to the head (the α and β anomer forms) would be more well-represented in the modeling than a stereochemical change elsewhere in the sugar head.

Finally, an additional type of surfactant, the methyl glucamides (C_iMEGA), is presented as a further test of the MT theory (with the head area obtained from Ref. [109]). These surfactants have not yet been fully analyzed using the improved CSMT approach described in this thesis in Chapter 3, although C_{10}MEGA has been investigated by Stephenson and coworkers[109] using an older approach to good effect.

5.5 Conclusions

In this Chapter, I applied my version of a molecular-thermodynamic (MT) model for surfactant micellization to the study of 55 linear surfactants for which experimental critical micelle concentration (cmc) data was available, including 16 anionics (sulfonates, sulfates, benzene sulfonates, and ethoxysulfates), 13 cationics (trimethylammoniums, 1-alkyl pyridiniums, and 1-methyl, 4-alkyl pyridiniums), 4 zwitterionics (sulfobetaines with 3- and 4-carbon spacers), 10 ethoxylates of various alkyl and ethoxylate chain lengths, and 12 sugar-based surfactants (galactosides, glucosides, maltosides, and methylglucamides). Of these 55 linear surfactants, simulations were conducted for 43 of them to enable prediction using a combined computer-simulation/molecular-thermodynamic (CSMT) model. In both MT and CSMT modeling, four solubility models were examined: the MHA model, the MHA2 model, the KWB model, and the KEKSS model (see Section F.5 of Appendix F for a full discussion).

The results of the modeling indicate that the MHA model within the MT framework offers accurate predictions for the anionics, cationics, ethoxylates, and sugar-based surfactants considered. The zwitterionic type examined (sulfobetaine) demonstrated a consistent underprediction, which I explained as being related to the actual orientation of the sulfobetaine head group at the micelle core-water interface. Further work is needed to more accurately model this type.

Table 5.16: CMC Results for Linear Nonionic Surfactants at $T = 25°C$

	CMC (mM)	MT Pred. CMC (mM)[a]				CSMT Pred. CMC (mM)[b]			
Surfactant	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_8GAL(β)	16[259]	**9.4**	37	0.98	2.2	**12±3** [8]	43±7 [8]	1.2±0.2 [8]	2.9±0.7 [8]
C_8GLU(α)	12[259]	**9.4**	37	0.98	2.2	**8.9±0.2** [7]	30.5±0.5 [7]	0.85±0.01 [7]	2.13±0.04 [7]
C_8GLU(β)	18.2;[260] 20[259]	9.4	**37**	0.98	2.2	**10.5±1.8** [5]	36.9±4.7 [5]	1.0±0.1 [5]	2.5±0.4 [5]
C_{10}GLU(β)	0.8[259]	**0.56**	4.1	0.084	0.13	**0.5±0.1** [5]	3.5±0.4 [5]	0.07±0.01 [5]	0.13±0.02 [5]
C_{12}GLU(β)	0.15[259]	0.033	**0.44**	0.0070	0.0078	0.029±0.005 [5]	**0.36±0.04** [5]	0.0059±0.0008 [5]	0.007±0.001 [5]
C_8MAL(β)	19.1[260]	**31**	130	3.3	7.5	-	-	-	-
C_{10}MAL(β)	2.0[261]	**1.8**	13	0.27	0.43	-	-	-	-
C_{12}MAL(α)	0.15[237]	**0.10**	1.4	0.022	0.024	**0.05±0.003** [5]	0.58±0.03 [5]	0.0097±0.0005 [5]	0.0116±0.0006 [5]
C_{12}MAL(β)	0.13;[260] 0.20[237]	**0.10**	1.4	0.022	0.024	**0.08±0.01** [15]	0.94±0.08 [15]	0.015±0.002 [15]	0.018±0.003 [15]
C_8MEGA	51.3;[262] 70[263]	**88**	350	9.2	21	-	-	-	-
C_{10}MEGA	4.8[262]	**5.2**	38	0.78	1.2	-	-	-	-
C_{12}MEGA	0.248[264]	**0.29**	3.9	0.062	0.069	-	-	-	-

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] MT predictions do not have error bars, since the theory is deterministic based on user input.

[b] The error bars in the CSMT predictions represent the standard deviation in the predicted cmc arising from differences between the independent simulations run. The number of simulations is denoted in square brackets. Note that the four solubility models are applied to the same set of simulations, but difficulties with numerical convergence in certain cases may lead to some points being discarded.

Table 5.17: Free Energy of Micellization Comparison for Linear Nonionic Surfactants at $T = 25°C$

	g_m^* (k_BT)[a]	MT Error (k_BT)[b]				CSMT Error (k_BT)[b]			
Surfactant	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
$C_8GAL(\beta)$	-8.15	**-0.53**	0.84	-2.79	-1.98	**-0.29**	0.99	-2.59	-1.71
$C_8GLU(\alpha)$	-8.44	**-0.24**	1.13	-2.51	-1.70	**-0.30**	0.93	-2.65	-1.73
$C_8GLU(\beta)$	-7.97	-0.71	**0.66**	-2.97	-2.16	**-0.60**	0.66	-2.95	-2.03
$C_{10}GLU(\beta)$	-11.14	**-0.36**	1.63	-2.25	-1.82	**-0.47**	1.48	-2.44	-1.82
$C_{12}GLU(\beta)$	-12.82	-1.51	**1.08**	-3.06	-2.96	-1.64	**0.88**	-3.24	-3.06
$C_8MAL(\beta)$	-7.97	**0.48**	1.92	-1.76	-0.93	-	-	-	-
$C_{10}MAL(\beta)$	-10.23	**-0.11**	1.87	-2.00	-1.54	-	-	-	-
$C_{12}MAL(\alpha)$	-12.82	**-0.41**	2.23	-1.92	-1.83	**-1.10**	1.35	-2.74	-2.56
$C_{12}MAL(\beta)$	-12.72	**-0.50**	2.14	-2.01	-1.93	**-0.72**	1.74	-2.40	-2.22
C_8MEGA	-6.82	**0.37**	1.75	-1.89	-1.06	-	-	-	-
$C_{10}MEGA$	-9.35	**0.08**	2.07	-1.82	-1.39	-	-	-	-
$C_{12}MEGA$	-12.32	**0.16**	2.76	-1.39	-1.28	-	-	-	-

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] Effective value calculated from the experimental data in Table 5.16 using Eq. 5.3.2.

[b] Errors are computed using Eq. 5.3.3.

The CSMT approach did not generally outperform the MT approach, except in certain cases where the MT approach overpredicted the cmc, since the tendency of the CSMT approach is to decrease (make more negative) the transfer free energy. However, the computer simulations informed the head/tail identification used in MT modeling where rules of thumb were not readily available, such as with the surfactants possessing aromatic rings (including the benzene sulfonates and the pyridiniums). Additionally, as mentioned, the computer simulations afforded valuable insight into the behavior of the sulfobetaine head groups, providing a source of recommendation for further improvement to the model. Finally, in the case of the sugar-based surfactants, the CSMT approach was able to capture the correct trend between α and β anomers of glucoside and maltoside surfactants. This type of prediction of micellization behavior as a function of stereochemistry is not possible in the current MT model alone, indicating that the CSMT model may be invaluable for certain applications which involve these types of surfactants.

The systems in this Chapter were selected to provide a thorough test of the various models and methods proposed, with an emphasis on the effect of changes in head group on micellization properties. In Chapter 6, I proceed to investigate the effect of changing the tail group, through the analysis of three types of branched surfactants and various branching configurations within each type.

Chapter 6

Branched Surfactants

6.1 Overview

In this Chapter, I apply the molecular-thermodynamic (MT) theory developed in Chapter 2 and the combined computer-simulation/molecular-thermodynamic (CSMT) framework developed in Chapter 3 to the study of the micellization behavior of branched surfactants. My original contributions in this area include: (i) developing a single-chain, mean-field packing theory for branched surfactants, which utilizes a rotational isomeric state (RIS) model proposed by Mattice[217] for the evaluation of chain conformations, (ii) determining updated parameters, using specific gravity data,[265] for a group contribution method for determining molecular volume, which allows chains possessing nodal CH alkyl groups to be considered, (iii) proposing a means for estimating the interfacial tension of singly-branched surfactant tails against water, and (iv) developing an automation framework for preparing and analyzing computer simulations of branched surfactants, which enables efficient calculation of the quantities needed for CSMT modeling. (Contribution (i) is described in Chapter 4, contributions (ii) and (iii) are described in detail in Appendix F, and contribution (iv) is presented in Appendix C.) Additionally, as with the linear surfactant systems in Chapter 5, I have identified four solubility models for use in assessing the transfer free-energy contribution of the surfactant tail to the free energy of micellization. Two of them (based partly on the correlations of Abraham[239] and labeled MHA and MHA2) in-

volve original work in fitting experimental linear and branched alkane solubility data. All four solubility models are summarized in Section F.5 of Appendix F.

This is the first time that branched surfactants have been considered in an MT theory in a systematic fashion and on an equal footing with linear surfactants. That is, I have been able to establish, for each of the free-energy contributions to the free energy of micellization in the MT theory, an appropriate model for branched surfactants that is compatible with predictions of linear surfactant micellization behavior. A previous effort, described in Ref. [204], which I collaborated in, involved several simplifying assumptions regarding packing and the evaluation of molecular descriptors. These assumptions are no longer required in my improved theory for branched surfactant micellization.

In Chapter 5, the focus of modeling was primarily on changes in the surfactant head group, and the role of the head group in property prediction. In this chapter, I restrict my study to three types of surfactants only, and focus instead on the role of the surfactant tail. These types were selected based on the availability of quality experimental data with which to compare, and involve ionic surfactants. In Section 6.2 I briefly describe the specific systems that I have selected for analysis – two anionic surfactant types and one cationic surfactant type. In Section 6.3, as in Chapter 5, I focus much attention on prediction of the critical micelle concentration (cmc) as a measure of the accuracy of my theoretical predictions. However, I also describe the effect of branching on optimal micelle shape and size and the degree of counterion binding to the charged micelles formed by the ionic surfactants considered. Finally, in Section 6.4, I summarize the key results presented in this Chapter.

6.2 Discussion of Systems

I have selected three types of branched surfactants for analysis: (i) alkyl (paraffin) sulfonates, exhibiting branching at the carbon alpha to the sulfonate head, (ii) alkyl benzene sulfonates, exhibiting branching at the carbon attached to the benzene ring at the 4-position, and (iii) 1-methyl, 4-alkyl pyridiniums, exhibiting branching at

various points along the primary alkyl chain. A summary of the chemical structures of these surfactants can be found in Figure 6-1. For all structures, I consider a single stereochemistry for simplicity of presentation, since there is no significant impact of changing this stereochemistry in these singly-branched cases, due to the symmetric nature of the functional group attached (in contrast to the situation with the anomeric forms of sugar-based surfactants, as described in Chapter 5). Note that, for surfactants possessing multiple branches in close proximity, the specific choice of stereochemistry becomes extremely important in terms of packing within the micelle core, as adjacent branches may sterically interact, depending on the connectivity.

Of the three surfactant types considered, the branched alkyl benzene sulfonates are encountered the most widely in industrial usage, and, in fact, represent one of the largest types of surfactants in use by net tonnage out of all surfactants, branched and linear.[12] It is important to note that these particular surfactants are quite commonly referred to as "linear alkyl benzene sulfonates",[27] since their structure can alternately be viewed as a benzene sulfonate functional group positioned at various points along a linear alkyl chain. From a surfactant micellization perspective, however, these structures should certainly be considered branched, as the specific orientation of a surfactant within a micelle (such that the head lies in the solvent region, and the tail lies in the micelle core) leads to a packing contribution to the free energy of micellization which differs from that for a linear tail. When the branching is at the alpha carbon, the situation can even be viewed as similar to that for a two-tailed surfactant, where the branching node is just inside the micelle core-water interface, rather than just outside. (Actual examples of two-tailed surfactants include the dialkyl ammonium bromides,[266] and, involving a head group spacer, the various dimeric or "gemini" surfactants.[17]) In fact, an experimental study of the alkyl sulfonates considered in Figure 6-1, with a similar pattern of branching at the alpha carbon as the alkyl benzene sulfonates, was motivated as a study of a series of "double-tailed" surfactants.[59] (The alkyl sulfonate surfactants, along with the 1-methyl, 4-alkyl pyridinium surfactants,[60,61] were synthesized with the specific aim of studying structure-property relationships experimentally.)

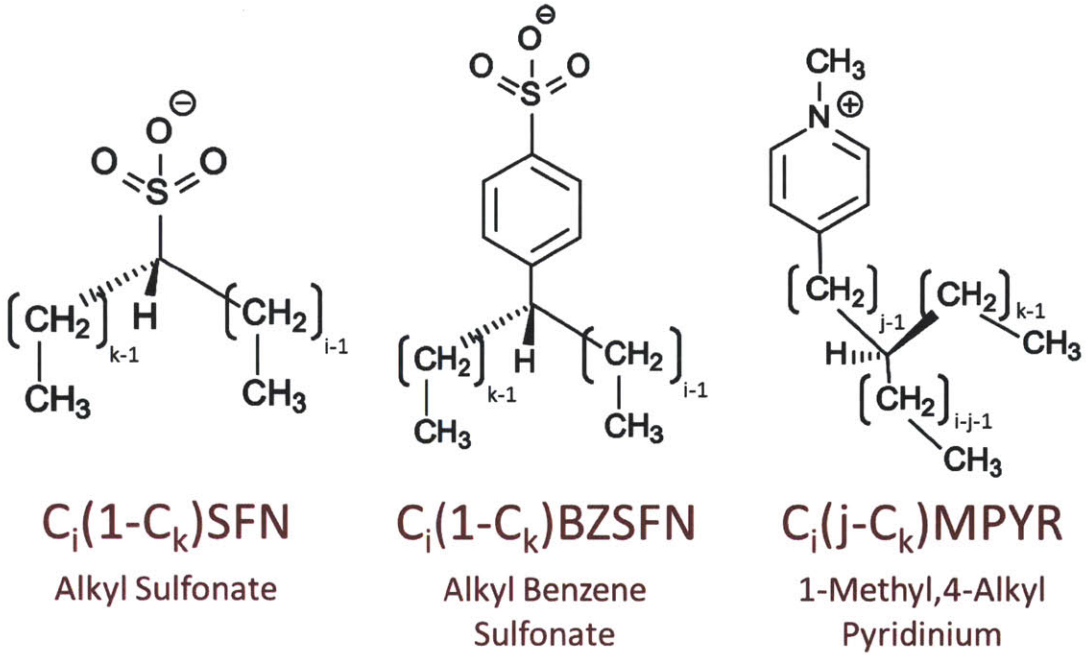


Figure 6-1: Chemical structures and labels for the branched surfactants investigated in this Thesis. The alkyl chains in these branched surfactants are specified generally by $C_i(j\text{-}C_k)$, where i is the number of carbons in the primary chain, j is the position of the nodal carbon in the primary chain (with $j = 1$ the carbon alpha to the main functional group), the minus indicates the stereochemistry of the secondary chain attachment as *gauche*$^-$ (i.e., offset from the main chain by -120 degrees, and k is the number of carbons in the secondary chain. (Note that these structures are intended to depict atom connectivity and stereochemistry only, and are not intended to convey accurate bond length, angle, and dihedral information.)

As with the linear surfactants in Chapter 5, both MT and CSMT models for branched surfactant micellization require identification of the surfactant head and tail, together with a set of molecular descriptors to describe the physicochemical properties of each. In contrast to the case with the linear surfactants, there are no well-established rules of thumb in performing this identification for branched surfactants, since they are less well-studied from a theoretical perspective. Srinivasan proposed that alkyl benzene surfactants be considered to have the benzene ring fully in the tail, viewing the benzene ring as conventionally hydrophobic.[119] However, it is not clear that the usual notion of hydrophobicity[34] can be ascribed to aromatic compounds, which exhibit a ring-stacking behavior quite different from aliphatic materials that is understood to contribute to solubility limits in aqueous solution.[267,268] Since Srinivasan used this tail identification solely for the purposes of conducting illustrative packing calculations, and did not perform micellization predictions to demonstrate the efficacy of this rule of thumb,[119] I decided to use the head/tail identification obtained using all-atomistic computer simulations of micelles together with the dividing surface approach described in Section 3.3 of Chapter 3 to develop the head/tail identifications for the MT predictions.

To my knowledge, aside from my collaboration with Lin and Blankschtein (described in Ref. [204]), only the alkyl benzene sulfonates have been systematically studied from a computer simulations perspective. For these, He and coworkers recently investigated the effect of benzene sulfonate position on the linear alkyl backbone on the surface behavior of these surfactants at the air-water interface (using all-atomistic simulations and the CHARMm force field).[269] These researchers also investigated the bulk phase behavior of these surfactants using a coarse-grained approach, identifying micellar, lamellar, hexagonal, or intermediate phases as a function of surfactant concentration and benzene sulfonate position.[176]

In order to maintain consistency with the study of linear surfactants presented in Chapter 5, I again selected force field parameters consistent with the OPLS-AA force field of Jorgensen and coworkers (described in Section D.2 of Appendix D).

For the alkyl sulfonate surfactants, the force field parameters of Canongia Lopes

and coworkers[194] were used, with the caveat that they did not consider branching at the alpha carbon, but the hydrogens in their model on that carbon were assigned a partial charge of zero, enabling any neutral group, such as the first methylene group of the side chain, to be attached at one of these hydrogen positions. The addition of this side chain introduces no new bond length, angle, or dihedral parameters beyond those which Canongia Lopes and coworkers already specified. Note that their force field parameters were designed to be compatible with OPLS-AA.[194]

For the alkyl benzene sulfonate surfactants, I supplied the missing parameters through ab initio calculations using Gaussian 03.[234] I provide the full details of this procedure as a case study for force-field development in Appendix E.2.

For the 1-methyl, 4-alkyl pyridinium surfactants, parameters for bonded interactions were obtained from the general work of Canongia Lopes and Pádua on alkyl pyridiniums.[193] I calculated atomic charges for the specific 1-methyl, 4-alkyl configuration using the CHELPG algorithm[233] in Gaussian 03,[234] as described in Section E.2.5 of Appendix E.

Once the force field parameters were specified, I performed monomer and micelle simulations using the automation framework I developed as described in Appendix C and the simulations protocols described in Section 3.4 of Chapter 3. Aggregation numbers were selected as multiples of 12, for the reasons described in Section 5.2 of Chapter 5.

Table 6.1 shows the tail identification results of the computer simulations of the surfactants in Figure 6-1, together with the number of simulations performed for each surfactant, the aggregation numbers considered, and the simulation temperature. In this Chapter, I have conducted the simulations at temperatures matching the experimental conditions, to gain the best possible understanding of the micelle structure and hydration.

There are several key observations in Table 6.1 which can be distilled into guidelines for conducting MT modeling of these surfactants. (Note that CSMT modeling is always performed with the head/tail identification obtained for the simulation under consideration – the CSMT results in Section 6.3 for the branched surfactants are aver-

Table 6.1: Branched Surfactant Tail Identifications from Computer Simulation

Surfactant	Ion	T (°C)[a]	Tail Identification[b]
C_7(1-C_5)SFN	Na+	40.0	S12: C_7(1-C_5) [6]; S18: C_7(1-C_5) [2]
C_8(1-C_4)SFN	Na+	40.0	S18: C_8(1-C_4) [6]; S24: C_8(1-C_4) [3]
C_9(1-C_3)SFN	Na+	40.0	S18: C_9(1-C_3) [6]; S24: C_9(1-C_3) [9]
C_{10}(1-C_2)SFN	Na+	40.0	S24: C_{10}(1-C_2) [5]; S36: C_{10}(1-C_2) [8]
C_{11}(1-C_1)SFN	Na+	40.0	S24: C_{11}(1-C_1) [5]; S36: C_{11}(1-C_1) [5]
C_{10}(1-C_2)SFN	Na+	45.0	S24: C_{10}(1-C_2) [8]; S36: C_{10}(1-C_2) [6]
C_{10}(1-C_4)SFN	Na+	45.0	S24: C_{10}(1-C_4) [10]; S36: C_{10}(1-C_4) [7]
C_{10}(1-C_6)SFN	Na+	45.0	S24: C_{10}(1-C_6) [10]; S36: C_{10}(1-C_6) [6]
C_{10}(1-C_8)SFN	Na+	45.0	S24: C_{10}(1-C_8) [10]; S36: C_{10}(1-C_8) [5]
C_{10}(1-C_9)SFN	Na+	45.0	S24: C_{10}(1-C_9) [9]; S36: C_{10}(1-C_9) [8]
C_5(1-C_3)BZSFN	Na+	25.0	S24: B_6C_5(1-C_3) [5]
C_6(1-C_2)BZSFN	Na+	25.0	S24: B_6C_6(1-C_2) [5]
C_7(1-C_1)BZSFN	Na+	25.0	S24: B_6C_7(1-C_1) [5]
C_6(1-C_4)BZSFN	Na+	25.0	S24: B_6C_6(1-C_4) [5]; S36: B_6C_6(1-C_4) [5]
C_7(1-C_3)BZSFN	Na+	25.0	S24: B_6C_7(1-C_3) [5]; S36: B_6C_7(1-C_3) [5]
C_8(1-C_2)BZSFN	Na+	25.0	S24: B_6C_8(1-C_2) [5]; S36: B_6C_8(1-C_2) [5]
C_9(1-C_1)BZSFN	Na+	25.0	S24: B_3C_9(1-C_1) [1], B_5C_9(1-C_1) [1], B_6C_9(1-C_1) [3] S36: B_5C_9(1-C_1) [1], B_6C_9(1-C_1) [4]
C_7(1-C_5)BZSFN	Na+	25.0	S24: B_6C_7(1-C_5) [5]; S36: B_6C_7(1-C_5) [5]
C_8(1-C_4)BZSFN	Na+	25.0	S24: B_6C_8(1-C_4) [5]; S36: B_6C_8(1-C_4) [5]
C_9(1-C_3)BZSFN	Na+	25.0	S24: B_6C_9(1-C_3) [5]; S36: B_6C_9(1-C_3) [5]
C_{10}(1-C_2)BZSFN	Na+	25.0	S24: B_5C_{10}(1-C_2) [2], B_6C_{10}(1-C_2) [3] S36: B_6C_{10}(1-C_2) [5]
C_{11}(1-C_1)BZSFN	Na+	25.0	S24: B_5C_{11}(1-C_1) [4], B_6C_{11}(1-C_1) [1] S36: B_5C_{11}(1-C_1) [3], B_6C_{11}(1-C_1) [2]
C_7(2-C_5)MPYR	I-	30.0	S24: B_3C_7(2-C_5) [5]
C_8(4-C_4)MPYR	I-	30.0	S24: B_0C_8(4-C_4) [9]
C_9(5-C_3)MPYR	I-	30.0	S36: B_0C_9(5-C_3) [6]
C_9(6-C_3)MPYR	I-	30.0	S36: B_0C_9(6-C_3) [5]
C_{10}(2-C_2)MPYR	I-	30.0	S36: B_0C_{10}(2-C_2) [6]
C_{10}(4-C_2)MPYR	I-	30.0	S36: B_0C_{10}(4-C_2) [6]
C_{10}(6-C_2)MPYR	I-	30.0	S36: B_0C_{10}(6-C_2) [5]
C_{10}(8-C_2)MPYR	I-	30.0	S36: B_0C_{10}(8-C_2) [5]

Abbreviation codes: SFN – sulfonate (SO_3^-), BZSFN – benzene sulfonate ($C_6H_4SO_3^-$), MPYR – 1-methyl pyridinium ($CH_3N^+C_5H_4$).

[a] Simulation temperature, corresponding to experimental conditions.

[b] See Section 3.3 of Chapter 3 for details of the head/tail identification procedure. Entries show the spherical aggregation number simulated as SN, where N is the aggregation number. Tail identifications are followed by the number of simulations possessing that identification. The formula B_k means k aromatic carbon groups are included in the tail in a symmetric fashion ($k = 0$ indicates the full aromatic ring should be considered in the head region for packing calculation purposes).

ages across all the micelle simulations performed.) The first observation is that there is a complete consistency in the tail identification for the branched sulfonates: all the alkyl groups are identified as being in the core, regardless of the length of either the primary or side chain, and regardless of aggregation number. The second observation is that the alkyl benzene sulfonates always exhibit a tail identification including the full alkyl tail and some portion of the benzene ring. For longer secondary tails and/or short primary tails, the full benzene ring is included. For the alkyl benzene surfactants with C_9(1-C_1), C_{10}(1-C_{12}), and C_{11}(1-C_1) alkyl chains, which have longer primary chains and very short secondary chains, there is more water penetration, possibly due to the short secondary chains acting in a more steric fashion. The effect is slight, with a few examples where the 1-carbon in the ring is hydrated enough to be considered head. The third observation is that all 1-alkyl, 4-methyl pyridinium surfactants except for that containing a C_7(2-C_5) chain exhibit a tail identification consisting of the entire alkyl region, but no part of the charged pyridinium ring. Since the nitrogen baring the majority of the charge is situated at the 1-position, and the alkyl at the 4-position, this distance is likely the reason that none of the alkyl groups are hydrated enough to be considered head. One can contrast this with the results for the linear 1-alkyl pyridinium surfactants, which had several alkyl groups identified as head, due to both the charged nitrogen and alkyl groups being in close proximity. (Note that it is the interaction of water with charged moieties which motivated the original rule of thumb that the carbon groups alpha to a charged moiety tend to be quite hydrated.[34]) The C_7(2-C_5)MPYR surfactant, which has the smallest tail length exhibits a tail identification in which half of the pyridinium ring is considered tail. This is likely the result of the smaller radial dimensions of these micelles, which increases the volume fraction of the pyridinium rings at the interface relative to the cases involving longer branched alkyl groups, which force the pyridinium ring to be positioned further from the micelle core center of mass. (Note that the linear versions of these surfactants, C_{10}MPYR and C_{12}MPYR have identifications in line with the longer branched chains – see Table 5.2 in Chapter 5.)

Based on the above three observations, and the consistent agreement in tail iden-

Table 6.2: Branched Surfactant Tail Identifications Used for MT Modeling

Surfactant	Tail Identification[a]
C_i(1-C_k)SFN	C_i(1-C_k)
C_i(1-C_k)BZSFN	B_6C_i(1-C_k)
C_i(j-C_k)MPYR	B_0C_i(j-C_k)

[a] The formula B_k means k aromatic carbon groups are included in the tail in a symmetric fashion ($k = 0$ indicates the full aromatic ring should be considered in the head region for packing calculation purposes).

tifications across almost all surfactants in each type, I have chosen tail identifications for MT modeling according to Table 6.2. Note that, in modeling the alkyl benzene sulfonates and 1-methyl, 4-alkyl pyridiniums, the location of the aromatic ring (benzene or pyridinium) is explicitly considered in conducting packing calculations. For the benzene sulfonates, the benzene is fully within the core, and for the 1-methyl, 4-alkyl pyridiniums, the pyridinium is fully within the head. In the latter case, the positioning of the pyridinium ring in the head does affect packing calculations, by restricting the external orientations achievable by the alkyl tail (orientations in which part of the pyridinium ring enter the core are forbidden under a strict head/tail identification, sharp micelle core-water interface approach of the sort described in Chapter 4).

Based on the tail identifications to be used in the CSMT modeling, summarized in Table 6.1, and the tail identifications to be used in the MT modeling, summarized in Table 6.2, there are only a few head types to consider. These are: (i) a sulfonate head (SFN), (ii) a sulfonate together with the 1-carbon of a benzene ring (SFN+B_1, where +B_1 indicates one aromatic carbon is included in the head), (iii) a sulfonate together with half of a benzene rings (SFN+B_3, where +B_3 indicates three aromatic carbons are included in the head), (iv) a 1-methyl pyridinium head (MPYR), and (v) a 1-methyl pyridinium head with half of the pyridinium ring in the core (MPYR-B_3, where -B_3 indicates 3 aromatic groups are excluded from the head).

Table 6.3 summarizes the molecular descriptors associated with these head types.

Table 6.3: Head Group Properties for Branched Surfactants

Head	a_h (Å^2)[a]	z[a]	d_z (Å)[a]	l_h (Å)[a]
SFN	24.9	-1	0.362	1.762
SFN+B_1[b]	24.9	-1	2.162	3.562
SFN+B_3[b]	24.9	-1	2.861	4.261
MPYR	35.0	+1	2.712	5.738
MPYR-B_3[b]	35.0	+1	4.222	7.248

Abbreviation codes: SFN – sulfonate (SO_3^-), MPYR – 1-methyl pyridinium ($CH_3N^+C_5H_4$).
[a] a_h: head cross-sectional area; z: valence; d_z: distance from micelle core-water interface to average charge location; l_h: average head length.
[b] $\pm B_k$ indicates the number of aromatic carbons to include (+) or exclude (-) from the head.

These include: (a) the cross-sectional head area, a_h, calculated based on the cross-sectional area of the main functional group under consideration (SFN or MPYR), (b) the valence of the head, z, (c) the distance from the micelle core-water interface to the average location of charge, d_z, and (d) the head length, l_h, evaluated from the micelle core-water interface to the maximum extent of the surfactant head. The head length is used in determining the location of the Stern surface for electrostatics calculations, as depicted in Figure 2-2 in Chapter 2.

6.3 Modeling Results

6.3.1 Branched Alkyl Sulfonate Surfactants

As discussed in Section 5.3.1 of Chapter 5, an excellent test for the accuracy of any model for the free energy of micellization of a surfactant involves comparison of experimental and predicted critical micelle concentration (cmc) values, since the cmc depends exponentially on this free energy. That is,

$$X_{cmc} \approx \exp\left(+\frac{g_m^*}{k_B T}\right), \tag{6.3.1}$$

where X_{cmc} is the cmc in mole fraction units, k_BT is the thermal energy (k_B is the Boltzmann constant, and T is the absolute temperature), and g_m^* is the modified free energy of micellization determined via the optimization process described in Section 2.4.1 of Chapter 2. As described in Chapter 5, if one is provided with a cmc, it can be useful to invert Eq. 6.3.1 and convert X_{cmc} into molar concentration units, as follows:

$$\frac{g_m^*}{k_BT} \approx \ln\left(\frac{C_{cmc}[mM]}{1000\mathcal{M}_{H_2O}[M]}\right), \tag{6.3.2}$$

in order to calculate an effective minimized modified free energy of micellization corresponding to the cmc. I use the term "effective" to reflect the fact that an actual experimental solution contains a population distribution of micelles[35] that is only approximately unimodal like the theory. In Eq. 6.3.2, C_{cmc} is the cmc in millimolar concentration units, and $\mathcal{M}_{H_2O}$ is the molarity of water at the temperature at which the cmc is measured.

In Chapter 5, I also described a useful measure of error when comparing predicted ("model") and experimental ("expt") cmc's:

$$\Delta g_m^* \equiv \frac{g_{m,\text{model}}^*}{k_BT} - \frac{g_{m,\text{expt}}^*}{k_BT} = \ln\left(\frac{C_{cmc}^{\text{model}}}{C_{cmc}^{\text{expt}}}\right), \tag{6.3.3}$$

where Δg_m^* is evaluated assuming that the predicted cmc holds at the experimental temperature, T. Note that the predictions in this Thesis are evaluated at a model temperature of $T_{\text{model}} = 25°C$, but the experimental temperature T is sometimes higher than this, whether due to issues with the Krafft point of a surfactant preventing measurement of a cmc at T_{model}, or interest on the part of the researchers in other properties, such as surface behavior, at the elevated temperatures. In Table 5.8 in Chapter 5, I demonstrated that linear anionic surfactants do not exhibit a large temperature dependence in the cmc (and an even smaller dependence of g_m^*, according to Eq. 6.3.1). A similar statement holds for the branched surfactants investigated in this Chapter. For example, for C_{10}(1-C_2)BZSFN, the cmc at $T = 25°C$ is 2.22 mM, while at $T = 50°C$ it is 2.69 mM.[270] If one assumes that the $25°C$ result holds at $50°C$, the error, determined using Eq. 6.3.3, is $\Delta g_m^* \approx 0.2\ k_BT$, which is quite small.

In Table 6.4, I have summarized the cmc results that I have obtained using the MT and CSMT approaches for the branched alkyl sulfonates. As in Chapter 5, I investigate four solubility models, with details of these provided in Appendix F. There are two sets of surfactants considered here: (i) a set of alkyl sulfonates containing 12 carbon groups in the alkyl chain, with C_{12}SFN serving as the linear comparison, and (ii) a set of alkyl sulfonates with 10 carbon groups in the primary chain, a variable secondary chain length, and C_{10}SFN serving as the linear comparison. Since the MT model is sensitive to the head/tail identification, I have included cmc values for the linear surfactants evaluated with both the usual rule-of-thumb tail identification of C_{i-1} for a C_iSFN surfactant, and a tail identification of C_i, in analogy to the branched surfactant tail identification of the full alkyl chain being considered tail (see Table 6.2). One can see, in comparing the experimental and predicted MT cmc values that Set (i) generally requires the C_i identification to obtain the correct qualitative trend in experimental cmc's. Note that the CSMT approach does not require any special handling to predict the trend, since it is less sensitive to the particular head/tail identification (specifically, if the alpha methylene is in the head group, it still may contribute to the CSMT transfer free energy according to Eq. 3.2.10 in Chapter 3).

There are two key observations to make regarding the trends in Sets (i) and (ii).

For Set (i), there is an increase in the cmc as the branched tail becomes more symmetric between the number of carbons in the primary and secondary chains. Since the interfacial tension of these surfactants is assumed to be equivalent, according to the arguments of Section F.6 of Appendix F (where I showed that calculating interfacial tension based on the total number of carbons in the tail, for singly-branched tails, is a reasonable approximation), and the head identifications are identical (indicating molecular descriptors related to steric and electrostatic free-energy contributions to the free energy of micellization are the same), the observed trend is due to the change in micelle core curvature with decreasing maximum tail length. That is, as the set proceeds from the linear case (C_{12}SFN) to the shortest tail (C_7(1-C_5)SFN), the maximum micelle radius decreases (due to chain packing considerations and the energetic penalty of forming a void in the micelle interior[271]).

Table 6.4: CMC Results for Branched Sulfonate Surfactants at $T = 25°C$

		CMC (mM)	MT Pred. CMC (mM)[a]				CSMT Pred. CMC (mM)[b]			
Surfactant	Ion	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_7(1-C_5)SFN	Na+	36.1[171,272] [40°C]	**27**	120	12	13	**45.1±0.7** [8]	218±4 [8]	20.5±0.3 [8]	20.9±0.3 [8]
C_8(1-C_4)SFN	Na+	28.3[171,272] [40°C]	**17**	72	8.1	8.2	**26.4±0.3** [9]	118±2 [9]	12.4±0.1 [8]	12.6±0.1 [8]
C_9(1-C_3)SFN	Na+	23.2[171,272] [40°C]	11	**44**	5.2	5.3	**16.0±0.3** [12]	68±1 [14]	7.7±0.1 [14]	7.8±0.1 [15]
C_{10}(1-C_2)SFN	Na+	18.6[171,272] [40°C]	8.0	**31**	3.9	4.0	**11.1±0.1** [12]	44.8±0.6 [13]	5.41±0.06 [13]	5.49±0.06 [12]
C_{11}(1-C_1)SFN	Na+	14.9[171,272] [40°C]	5.5	**21**	2.6	2.7	**7.8±0.1** [10]	30.9±0.5 [10]	3.84±0.05 [8]	3.90±0.04 [8]
C_{12}SFN[c]	Na+	10.4[171,272] [40°C]	**6.6**[d] 3.5[e]	26[d] **14**[e]	2.4[d] 1.5[e]	2.8[d] 1.5[e]	4.24±0.02 [9]	**17.69±0.08** [9]	1.610±0.007 [9]	1.824±0.008 [9]
C_{10}SFN[c]	Na+	39.8[59] [45°C]	**34**[d] 17[e]	94[d] **51**[e]	9.3[d] 5.5[e]	14[d] 7.3[e]	20.7±0.1 [6]	**63.9±0.3** [6]	5.99±0.03 [6]	8.54±0.04 [6]
C_{10}(1-C_2)SFN	Na+	18.2[59] [45°C]	8.0	**31**	3.9	4.0	**10.9±0.1** [14]	43.8±0.5 [14]	5.30±0.06 [14]	5.38±0.06 [13]
C_{10}(1-C_4)SFN	Na+	6.76[59] [45°C]	**3.2**	17	1.9	1.6	**4.69±0.08** [15]	25.5±0.5 [17]	2.77±0.05 [13]	2.33±0.04 [13]
C_{10}(1-C_6)SFN	Na+	1.66[59] [45°C]	**1.2**	8.4	0.85	0.59	**1.76±0.06** [12]	13.1±0.5 [15]	1.26±0.04 [13]	0.89±0.03 [12]
C_{10}(1-C_8)SFN	Na+	0.24[59] [45°C]	0.41	3.9	0.35	**0.20**	0.62±0.02 [13]	6.0±0.2 [12]	0.53±0.01 [11]	**0.30±0.01** [11]
C_{10}(1-C_9)SFN	Na+	0.089[59] [45°C]	0.21	2.3	0.20	**0.10**	0.32±0.01 [14]	3.6±0.2 [15]	0.30±0.01 [11]	**0.156±0.007** [11]

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] MT predictions do not have error bars, since the theory is deterministic based on user input.

[b] The error bars in the CSMT predictions represent the standard error in the mean predicted cmc arising from differences between the independent simulations run. The number of simulations is denoted in square brackets. Note that the four solubility models are applied to the same set of simulations, but difficulties with numerical convergence in certain cases may lead to some points being discarded.

[c] Linear molecules included for comparison with the branched results.

[d] C_{i-1} tail identification.

[e] C_i tail identification.

Since the area per surfactant increases as the micelle radius decreases, contributions to the free energy of micellization which are proportional to the area per surfactant (such as the interfacial free-energy contribution) will increase in magnitude, while properties approximately inversely proportional (such as the steric and electrostatic free energies) will decrease in magnitude.[35] The transfer free energy, from an MT perspective, is constant, so the observed trend (that the cmc increases, indicating a more-positive offset to the negative transfer free energy) points to a dominant contribution from the terms proportional to the area per surfactant.

For Set (ii), there is a decrease in the cmc with increasing secondary chain length. This can be understood in terms of the hydrophobic effect:[34] an increase in the number of methylenes present in the alkyl tail corresponds to an increase in the magnitude of the transfer free energy, providing a stronger driving force for micellization, which in turn allows micelle formation at lower surfactant concentrations. All four solubility models replicate this observation, since the transfer free energy ascribed to a methylene group within their separate group contribution formulas is always negative. The effect of the methylene group addition can be clearly seen in Table 6.5, by examining the values for g_m^*. Note that the transfer free energy contribution of a methylene group is a fixed value in each solubility model. However, the change in g_m^* in going from C_{10}(1-C_2) to C_{10}(1-C_4) is $-0.99\ k_BT$ (or roughly $-0.5\ k_BT$ per added methylene group), while the change in g_m^* in going from C_{10}(1-C_6) to C_{10}(1-C_8) is $-1.93\ k_BT$ (or roughly $-0.97\ k_BT$ per added methylene group). That is, as the secondary chain increases in length, there is less of a positive offset to the transfer free energy contributed by the additional methylenes (i.e., the "rate of change" in g_m^* with respect to secondary chain carbon number is increased), indicating a non-linear response of the various free-energy contributions to the number of carbon groups in the chain. Note that the primary chain length of surfactants in Set (ii) is fixed based on the C_{10} chain length; however, the tail volume increases as the secondary chain increases in length. These surfactants are roughly spherical, as will be seen in Table 6.6 shortly. Hence, the aggregation number decreases with increasing chain length. As a result the area per surfactant tends to increase.

Table 6.5: Free Energy of Micellization Comparison for Branched Sulfonate Surfactants at $T = 25°C$

		g_m^* (k_BT)[a]	MT Error (k_BT)[b]				CSMT Error (k_BT)[b]			
Surfactant	Ion	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_7(1-C_5)SFN	Na+	-7.33 [40°C]	**-0.30**	1.20	-1.11	-1.03	**0.22**	1.79	-0.57	-0.55
C_8(1-C_4)SFN	Na+	-7.57 [40°C]	**-0.51**	0.93	-1.26	-1.24	**-0.07**	1.42	-0.83	-0.81
C_9(1-C_3)SFN	Na+	-7.77 [40°C]	-0.75	**0.64**	-1.50	-1.48	**-0.38**	1.07	-1.11	-1.09
C_{10}(1-C_2)SFN	Na+	-7.99 [40°C]	-0.85	**0.51**	-1.57	-1.54	**-0.52**	0.87	-1.24	-1.23
C_{11}(1-C_1)SFN	Na+	-8.22 [40°C]	-1.00	**0.34**	-1.75	-1.71	**-0.65**	0.72	-1.36	-1.35
C_{12}SFN[c]	Na+	-8.57 [40°C]	**-0.46**[d] -1.09[e]	0.91[d] **0.29**[e]	-1.47[d] -1.94[e]	-1.32[d] -1.94[e]	-0.90	**0.53**	-1.87	-1.75
C_{10}SFN[c]	Na+	-7.23 [45°C]	**-0.16**[d] -0.86[e]	0.85[d] **0.24**[e]	-1.46[d] -1.99[e]	-1.05[d] -1.70[e]	-0.66	**0.47**	-1.90	-1.55
C_{10}(1-C_2)SFN	Na+	-8.01 [45°C]	-0.83	**0.53**	-1.55	-1.52	**-0.52**	0.87	-1.24	-1.23
C_{10}(1-C_4)SFN	Na+	-9.00 [45°C]	**-0.75**	0.92	-1.28	-1.45	**-0.37**	1.32	-0.90	-1.07
C_{10}(1-C_6)SFN	Na+	-10.41 [45°C]	**-0.33**	1.61	-0.68	-1.04	**0.05**	2.06	-0.28	-0.63
C_{10}(1-C_8)SFN	Na+	-12.34 [45°C]	0.53	2.78	0.37	**-0.19**	0.94	3.21	0.79	**0.22**
C_{10}(1-C_9)SFN	Na+	-13.33 [45°C]	0.85	3.25	0.80	**0.11**	1.27	3.69	1.21	**0.55**

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] Effective value calculated from the experimental data in Table 6.4 using Eq. 5.3.2 from Chapter 5.

[b] Errors are computed using Eq. 5.3.3 from Chapter 5.

[c] Linear molecules included for comparison with the branched results.

[d] C_{i-1} tail identification.

[e] C_i tail identification.

This has a very subtle effect on most of the free-energy contributions to g_m^*, which is, in fact, replicated in the MT predictions. For example, with the MHA model, the change in g_m^* for C_{10}(1-C_2) to C_{10}(1-C_4) is predicted to be $-0.92\ k_BT$, while the predicted change in g_m^* for C_{10}(1-C_6) to C_{10}(1-C_8) is $-1.06\ k_BT$, which demonstrates the correct trend (although it is not quite as negative). By carefully examining the various contributions to the free energy using the MT model, I was able to identify that the packing free energy contribution and the entropic penalty due to counterion binding both exhibited a decrease in the change in their free energies. For the packing free energy, C_{10}(1-C_2) to C_{10}(1-C_4) yielded a difference of $+0.63\ k_BT$, while C_{10}(1-C_6) to C_{10}(1-C_8) yielded a difference of +0.06, for a net difference between these two differences of $-0.57\ k_BT$. This was the largest-magnitude "change in the change in free energy" of all the contributions. The "change in the change in free energy" for the entropic penalty due to counterion binding was $-0.11\ k_BT$. For the interfacial, steric, and charging contributions (the dominant positive contributions), the values were $+0.17\ k_BT$, $+0.13\ k_BT$, and $+0.16\ k_BT$, respectively. This indicates that these other terms are continuing to increase in their "rate of change" with addition of methylene groups. However, the reduction in packing and the entropic penalty of counterion binding is more important.

In Table 6.6, in which I collect optimal micelle properties corresponding to the cmc's listed in Table 6.4 (for the particular example of the MHA solubility model and the MT framework), the reason that the packing term is no longer increasing is apparent: the spherical endcaps of the optimized micelles approach a steady value for the radius. Hence, the packing value no longer greatly varies once C_{10}(1-C_6) and longer chains are considered. In terms of the degree of counterion binding, there is also a "slow-down" in the decrease in this value as the chain length increases, due to the nonlinear form of the entropy of mixing.

There are several interesting points to be made regarding the data presented in Table 6.6.

Table 6.6: Optimal Micelle Shape, Size, and Degree of Counterion Binding Results for Branched Sulfonate Surfactants at $T = 25°C$

Surfactant	Ion	Shape	N_{total}[a]	N_{sph}[a]	N_{cyl}[a]	$l_{c,sph}$ (Å)[a]	$l_{c,cyl}$ (Å)[a]	β[b]
C_7(1-C_5)SFN	Na+	Spherocylinder	16	11	5	9.8	9.1	0.45
C_8(1-C_4)SFN	Na+	Spherocylinder	20	16	4	11.2	9.9	0.48
C_9(1-C_3)SFN	Na+	Spherocylinder	24	20	4	12.0	10.8	0.50
C_{10}(1-C_2)SFN	Na+	Spherocylinder	34	31	3	14.0	11.8	0.55
C_{11}(1-C_1)SFN	Na+	Spherocylinder	45	42	3	15.4	12.4	0.59
C_{12}SFN[c]	Na+	Sphere[d]	46	46	0	15.4		0.54
		Sphere[e]	47	47	0	16.0		0.59
C_{10}SFN[c]	Na+	Sphere[d]	30	30	0	12.6		0.51 (0.88[59])
		Spherocylinder[e]	40	37	3	14.0	11.0	0.60 (0.88[59])
C_{10}(1-C_2)SFN	Na+	Spherocylinder	45	42	3	14.0	11.8	0.55 (0.86[59])
C_{10}(1-C_4)SFN	Na+	Spherocylinder	26	24	2	13.3	11.7	0.47 (0.83[59])
C_{10}(1-C_6)SFN	Na+	Spherocylinder	20	18	2	12.7	12.0	0.37 (0.77[59])
C_{10}(1-C_8)SFN	Na+	Spherocylinder	17	16	1	12.7	11.8	0.29 (0.67[59])
C_{10}(1-C_9)SFN	Na+	Spherocylinder	16	15	1	12.7	12.3	0.25 (0.59[59])

Predictions made using the MHA solubility model and the MT framework.

[a] Micelle size properties: N_{total} – total aggregation number, N_{sph} – aggregation number in sphere or spherical end caps, N_{cyl} – aggregation number in cylindrical body, $l_{c,sph}$ – radius of sphere or spherical end caps, $l_{c,cyl}$ – cross-sectional radius of cylindrical body.

[b] Degree of counterion binding (number of bound counterions per surfactant). Experimental values in parentheses when available (see text for discussion).

[c] Linear molecules included for comparison with the branched results.

[d] C_{i-1} tail identification.

[e] C_i tail identification.

First, although spherocylinders are predicted for all branched sulfonates, it is important to note that the aggregation number in the cylindrical body is minimal, indicating that the optimal free energy of micellization for spheres and infinite cylinders is essentially identical, according to the discussion of interpolated geometries and their population distribution in Section 2.4.2 of Chapter 2. This means that these micelles can be considered as slightly elongated spheres. It is likely that an ellipsoid model would be most appropriate for modeling of these surfactants, although the non-regular geometry of ellipsoids present their own unique modeling challenges.[117]

Second, a decrease in the predicted aggregation number is observed in Set (i) as the maximum chain length shortens, and in Set (ii) as the secondary chain length lengthens. In both cases, one can see that the core-minor radius of the spherical endcaps is decreasing, contributing to this effect.

Finally, there is a reduction in the degree of counterion binding that occurs concomitantly with the decrease in the aggregation number. This can be understood to be related the associated decrease in the charge density as aggregation number decreases.[273] For Set (ii), in Figure 6-2, I have provided a visual plot of the predicted degrees of counterion binding in comparison with the experimental values as a function of secondary chain length. Note that the general decreasing trend is fairly well-predicted, with a similar slope (the change from 2 carbons to 9 carbons in the experimental measurement is -0.27, while the change in the predicted measurement for the same interval is -0.30). The reason for the offset is explained by Gunnarsson and coworkers: three means of measuring the degree of counterion binding are possible, including: (a) the use of thermodynamic techniques, (b) the use of transport measurements (such as conductivity), and (c) the use of spectroscopic techniques.[273] These techniques are listed in order of largest to smallest, with the spectroscopic techniques yielding values for the degree of counterion binding that include only counterions which are tightly bound to the micelle. As Goldsipe and Blankschtein clarify, this is the value expected to be most in agreement with the MT theory.[90] However, the values obtained from experiment and presented in Figure 6-2 are computed using thermodynamic techniques.[59] As a result, the offset is expected. In the example of

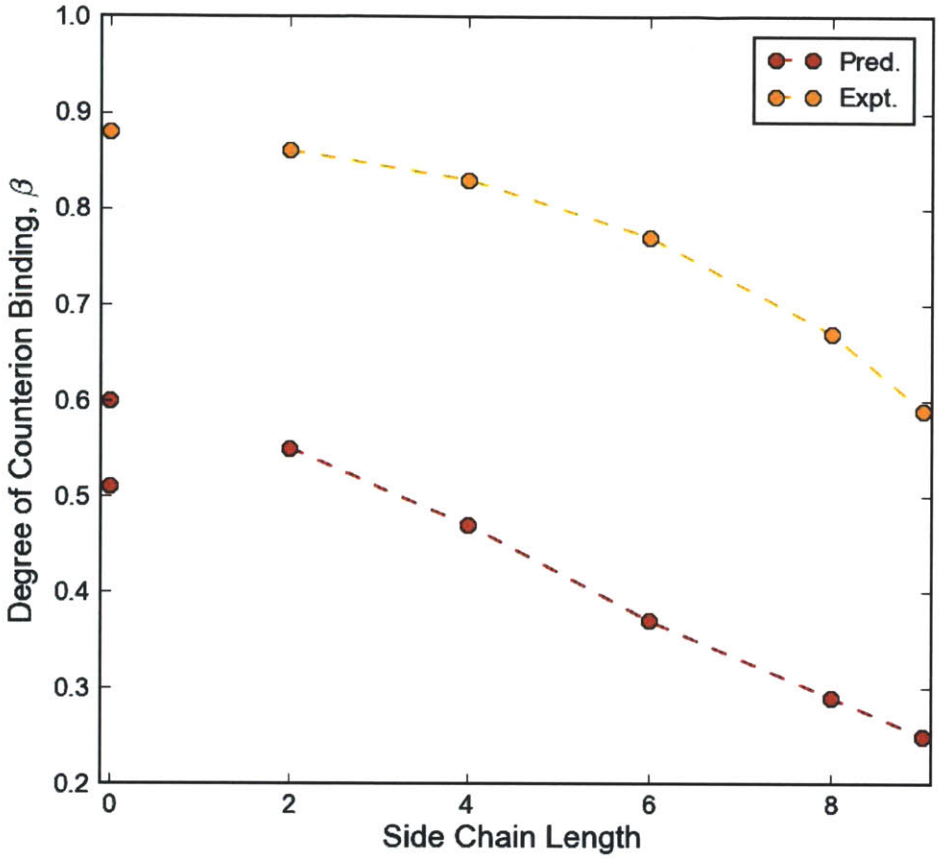


Figure 6-2: Comparison of predicted and experimental degrees of counterion binding for branched C_{10}(1-C_k) sulfonates as a function of side chain length, k. Note that $k = 0$ indicates a linear C_{10} sulfonate. Two predicted values are reported for this case, corresponding to two possible tail identifications. (See Table 6.6 for more information.)

cetyl trimethylammonium bromide, which Goldsipe and Blankschtein examined,[90] a similar comparison between spectroscopic-like values for β and thermodynamic values from experiment yields an offset comparable in magnitude to that seen here.

Having considered the trends in the data presented in this Section, and the insight gained from MT modeling (and by virtue of using a common molecular-thermodynamic framework, CSMT modeling), it is important to remark on the accuracy of the predictions in Table 6.4. As was the case with the linear surfactants in Chapter 5, the MHA model, based on the correlations developed by Abraham for

the Gibbs free energy of solution of alkanes,[239] and extended through my fitting of additional parameters for nodal carbons, performs quite well in both MT and CSMT settings, with deviations in the predicted free energies of micellization of usually less than 1 k_BT (corresponding to cmc predictions within a factor of 3 or less of the experimental values). In fact, the CSMT approach outperforms the MT approach in most cases. It is worth considering that the effect of an increase in system temperature for ionic surfactants is typically to increase the cmc. Since these predictions were made at $25°C$, and the experimental temperatures were either $40°C$ or $45°C$, this helps explain some of the underprediction observed. There are a few limited cases where overprediction is observed, including C_{10}(1-C_8) and C_{10}(1-C_9) benzene sulfonates. Here, the lack of quality data on interfacial tensions for longer chain alkanes could be an issue. Additionally, as the aggregation number decreases, there is a possibility that the various free energy models begin to break down due to the discreteness of the system. For example, the packing model in Chapter 4 assumes that chains interact through a mean field. I expect such an approach to be less accurate with small aggregation numbers.

The other solubility models, MHA2, KWB, and KEKSS, do not exhibit consistent-enough accuracy to be useful. The differences between the four models is solely due to the apportioning of the transfer free energy between groups. Because the surfactant tail is different from an actual oil (a C_{12} tail differs from dodecane in the replacement of a CH_3 group with a CH_2 group), it is only through testing of the sort presented in this Chapter and in Chapter 5 that a useful solubility model can be identified.

6.3.2 Branched Alkyl Benzene Sulfonate Surfactants

In Table 6.7 I have summarized the cmc results I obtained for three sets of branched alkyl benzene sulfonates. The three sets are grouped by the length of the linear alkyl chain to which the benzene sulfonate is attached, and I have selected an octyl, decyl, and dodecyl chain for study. The linear analogues, which have the benzene sulfonate attached at the alkyl chain 1-position, are represented as C_iBZSFN, while the branched alkyl benzene sulfonates are represented as C_i(1-C_k)BZSFN, where i

is the length of the primary (longest) chain, and k is the length of the secondary chain. Note that, for the branched surfactants in each set, $i + k$ equals 8 for the octyl set, 10 for the decyl set, and 12 for the dodecyl set. The corresponding linear surfactants are C_8BZSFN, C_{10}BZSFN, and C_{12}BZSFN. Since the tail identification obtained from computer simulations is overall quite consistently the full alkyl tail plus the full benzene ring for the branched surfactants, but the full alkyl tail and half the benzene ring for the linear surfactants, I have obtained MT predictions for the linear surfactants for both full and half-ring tails, in order to compare these identifications with the experimental trends. As in Section 6.3.1, it is important to note that the CSMT predictions are made using the precise tail identification obtained for each simulation, with the reported cmc an average across the simulations conducted (see Table 6.1 for the tail identifications obtained for each simulation).

There are two main trends to observe: (i) movement of the benzene sulfonate ring to the center of the alkyl chain uniformly increases the cmc (clearly seen with the decane and dodecane cases – there are too few experimental points for the octyl case to observe this, although the theory predicts it), and (ii) a change in the alkyl length from octyl to decyl and from decyl to dodecyl increases the cmc. Since Observation (ii) can be explained based on the hydrophobic effect,[34] I will focus on Observation (i).

The increase in cmc with shortening of the primary chain and lengthening of the secondary chain follows the same trend shown in Table 6.4 for the branched alkyl sulfonates. The reason is again intimately connected to the area per surfactant, which increases as the micelle dimensions decrease, and the trade-off between free-energy contributions proportional to this term (the interfacial free energy), and contributions roughly inversely proportional to this term (the steric and electrostatic free energies).[35]

Table 6.7: CMC Results for Branched Benzene Sulfonate Surfactants at $T = 25°C$

		CMC (mM)	MT Pred. CMC (mM)[a]				CSMT Pred. CMC (mM)[b]			
Surfactant	Ion	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_5(1-C_3)BZSFN	Na+	-	27	130	11	43	42.4±0.5 [5]	212±3 [5]	17.3±0.2 [5]	6.79±0.08 [4]
C_6(1-C_2)BZSFN	Na+	-	20	95	8.6	33	30.2±0.2 [5]	145.0±0.8 [5]	12.47±0.06 [5]	4.97±0.03 [5]
C_7(1-C_1)BZSFN	Na+	19[274] [55°C]	**16**	73	7.1	26	**24.1±0.2** [5]	110.9±0.7 [5]	10.21±0.05 [5]	4.11±0.02 [5]
C_8BZSFN[c]	Na+	11.1[245]	45[d] **12**[e]	170[d] 54[e]	**32**[d] 4.9[e]	120[d] 17[e]	26.5±0.1 [4]	110.6±0.5 [4]	**15.51±0.05** [5]	6.114±0.006 [5]
C_6(1-C_4)BZSFN	Na+	8.01[275]	**6.5**	40	3.4	16	**10.5±0.1** [10]	68.6±0.8 [10]	5.45±0.05 [8]	1.84±0.02 [9]
C_7(1-C_3)BZSFN	Na+	-	5.2	31	2.7	13	8.10±0.05 [10]	51.1±0.3 [10]	4.20±0.03 [10]	1.419±0.008 [10]
C_8(1-C_2)BZSFN	Na+	6.02[275]	**4.2**	24	2.2	10	**6.42±0.07** [10]	38.8±0.5 [10]	3.38±0.04 [10]	1.15±0.02 [10]
C_9(1-C_1)BZSFN	Na+	4.63[275]	**3.5**	20	1.9	8.5	5.6±0.9 [10]	30±6 [10]	**4±2** [10]	1.2±0.3 [10]
C_{10}BZSFN[c]	Na+	3.14[245] [50°C]	8.9[d] **2.8**[e]	44[d] 15[e]	**8.0**[d] 1.2[e]	39[d] 5.8[e]	5.2±0.7 [10]	29±4 [10]	**3.8±0.6** [10]	1.5±0.1 [9]
C_7(1-C_5)BZSFN	Na+	2.38[171,276]	**1.6**	13	1.0	6.3	**2.74±0.03** [10]	23.2±0.3 [10]	1.75±0.02 [9]	0.503±0.005 [9]
C_8(1-C_4)BZSFN	Na+	-	1.3	10	0.86	5.1	2.15±0.03 [10]	17.7±0.3 [9]	1.38±0.02 [10]	0.399±0.005 [9]
C_9(1-C_3)BZSFN	Na+	1.59[56]	**1.1**	8.3	0.71	4.1	**1.77±0.02** [10]	14.0±0.2 [10]	1.14±0.02 [10]	0.332±0.004 [10]
C_{10}(1-C_2)BZSFN	Na+	1.46[56]	**0.94**	6.8	0.61	3.4	**1.41±0.02** [10]	10±2 [10]	1.1±0.4 [10]	0.29±0.07 [10]
C_{11}(1-C_1)BZSFN	Na+	1.19[56]	**0.78**	5.6	0.51	2.9	**1.17±0.05** [10]	7±2 [10]	1.3±0.4 [9]	0.31±0.07 [10]
C_{12}BZSFN[c]	Na+	1.2[245] [60°C]	**1.9**[d] 0.60[e]	13[d] 4.4[e]	2.1[d] 0.33[e]	13[d] **2.0**[e]	**1.1±0.2** [15]	8±2 [15]	**1.1±0.2** [14]	0.33±0.04 [12]

The MT and CSMT predictions which are closest to the experimental values are bolded.

Unavailable experimental cmc's are marked with the symbol "-".

[a] MT predictions do not have error bars, since the theory is deterministic based on user input.

[b] The error bars in the CSMT predictions represent the standard deviation in the predicted cmc arising from differences between the independent simulations run. The number of simulations is denoted in square brackets. Note that the four solubility models are applied to the same set of simulations, but difficulties with numerical convergence in certain cases may lead to some points being discarded.

[c] Linear molecules included for comparison with the branched results.

[d] Using a tail identification of the full alkyl tail and half of the benzene ring.

[e] Using a tail identification of the full alkyl tail and the full benzene ring.

Note that, experimentally, the linear octyl and decyl surfactants fit into this trend within their respective sets. Although the linear dodecyl surfactant (C_{12}BZSFN) has a slightly higher cmc than the C_{11}(1-C_1)BZSFN surfactant, the difference is less than 0.01 k_BT, and the temperature at which the C_{12}BZSFN cmc is measured is sufficiently high (at 60°C versus the 25°C conditions under which the branched surfactant cmc's were obtained), such that the general trend of increasing cmc with increasing temperature for ionic surfactants (see Table 5.8 in Chapter 5) is sufficient to explain the difference. Measurement at the higher temperature was required due to Krafft point considerations for the linear surfactants.[245]

The trend described by Observation (i) is replicated in the MT predictions only when using the full alkyl tail plus full benzene ring tail identification. This demonstrates the sensitivity of the MT model to the particular head/tail identification used. Recall that the computer simulations of the linear surfactants indicated that only half of the ring is part of the tail (see Table 5.2 in Chapter 5). It can be readily seen that this identification does not lead to reproduction of the trend. However, the CSMT approach is much less sensitive to the particular head/tail identification: the decyl and dodecyl sets are predicted quite accurately by the MHA solubility model using the CSMT approach, and the trend is maintained between the linear and branched surfactants in the two sets. For the octyl set there is a deviation from the expected trend, although there is insufficient experimental data available to confirm the experimental cmc values in this case.

The numerical deviations in the predicted modified free energies of micellization are presented in Table 6.8. Note that the MHA model outperforms the other three solubility models in both the MT and CSMT frameworks, but that the KWB model is also quite a good model here. This may be due to the good agreement obtained by the KWB model in predicting the solubility of alkyl benzene oils, as can be readily observed from the values reported in Table F.10 in Appendix F.

Table 6.8: Free Energy of Micellization Comparison for Branched Benzene Sulfonate Surfactants at $T = 25°C$

		g_m^* (k_BT)[a]	MT Error (k_BT)[b]				CSMT Error (k_BT)[b]			
Surfactant	Ion	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_7(1-C_1)BZSFN	Na+	-7.97	**-0.18**	1.33	-1.00	0.30	**0.23**	1.75	-0.63	-1.54
C_8BZSFN[c]	Na+	-8.49	1.40[d]; **0.08**[e]	2.73[d]; 1.58[e]	**1.06**[d]; -0.82[e]	2.38[d]; 0.43[e]	0.87	2.30	**0.33**	-0.60
C_6(1-C_4)BZSFN	Na+	-8.84	**-0.21**	1.61	-0.86	0.69	**0.27**	2.15	-0.39	-1.47
C_8(1-C_2)BZSFN	Na+	-9.13	**-0.36**	1.38	-1.01	0.51	**0.06**	1.86	-0.58	-1.66
C_9(1-C_1)BZSFN	Na+	-9.39	**-0.28**	1.46	-0.89	0.61	0.19	1.87	**-0.15**	-1.35
C_{10}BZSFN[c]	Na+	-9.77	1.03[d]; **-0.12**[e]	2.63[d]; 1.55[e]	**0.93**[d]; -0.97[e]	2.51[d]; 0.60[e]	0.50	2.21	**0.18**	-0.75
C_7(1-C_5)BZSFN	Na+	-10.05	**-0.40**	1.70	-0.87	0.97	**0.14**	2.28	-0.31	-1.55
C_9(1-C_3)BZSFN	Na+	-10.46	**-0.37**	1.65	-0.81	0.95	**0.11**	2.18	-0.33	-1.57
C_{10}(1-C_2)BZSFN	Na+	-10.54	**-0.44**	1.54	-0.87	0.85	**-0.03**	1.92	-0.28	-1.62
C_{11}(1-C_1)BZSFN	Na+	-10.75	**-0.42**	1.55	-0.85	0.89	**-0.02**	1.77	0.09	-1.35
C_{12}BZSFN[c]	Na+	-10.73	**0.45**[d]; -0.71[e]	2.37[d]; 1.29[e]	0.55[d]; -1.30[e]	2.37[d]; **0.50**[e]	**-0.10**	1.88	**-0.10**	-1.30

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] Effective value calculated from the experimental data in Table 6.4 using Eq. 5.3.2 from Chapter 5.

[b] Errors are computed using Eq. 5.3.3 from Chapter 5.

[c] Linear molecules included for comparison with the branched results.

[d] Using a tail identification of the full alkyl tail and half of the benzene ring.

[e] Using a tail identification of the full alkyl tail and the full benzene ring.

Finally, in Table 6.9, I present information on optimal micelle shape, size, and degree of counterion binding obtained using the MHA solubility model and the MT approach. For the linear surfactants, results for both tail identifications are included. What can be observed here is a tendency for a decrease in the aggregation number among the branched surfactants within a given alkyl set as the primary chain length is shortened and the secondary chain length is lengthened (leading to a more symmetric configuration). These surfactants are predicted to be spherocylinders, with the cylindrical body becoming shorter as the total alkyl chain length is increased. That is, in transitioning from octyl to decyl to dodecyl for fixed secondary chain length, the ratio of the aggregation number found in the cylindrical body to the total aggregation number decreases, indicating the micelles are becoming more spherical. Similarly, within a given alkyl set, for surfactants possessing longer primary chains, the micelle shape becomes more spherical. A natural comparison here is to the linear surfactant within each set, which exhibits formation of either a spherical micelle or a spherocylindrical micelle with a negligible cylindrical body. In other words, as the secondary chain is shortened, a branched surfactant behaves more like a linear surfactant.

Experimental aggregation number data for these surfactants is not comprehensive, but does illustrate that the observed trends in predictions (number-averaged aggregation numbers) are upheld in the experimental systems with reasonable quantitative agreement. Quantitative accuracy in predicting aggregation numbers is complicated by the dependency of the population distribution of spherocylinders on the difference in free energies of micellization obtained for the optimized spherical and infinite cylindrical micelles obtained during the free energy minimization process. See Section 2.4.2 for an in-depth discussion of population distribution calculations for spherocylinders. Spherocylinders can grow to be very large in systems such as the linear alkyl ethoxylates possessing short ethoxylate chain lengths, reaching the thousands and tens of thousands of molecules.[171] In light of the fact that spherocylinders can span such a large range of aggregation numbers, in contrast to spheres, which are limited by the dimensions of the surfactant tail, the predictions here, which differ by only 10-20

molecules in total, are quite good.

Degrees of counterion binding are also included in Table 6.9. Although there does not appear to be experimental data to compare to, the trend is a decrease in counterion binding as the aggregation number decreases. Since the radial dimensions of the spherical endcaps and cylindrical body are also decreasing, leading to an increase in area per surfactant, the charge density at the micelle core-water interface is lowered. As mentioned in Section 6.3.1, electrostatic considerations explain this accompanying increase in surfactant dissociation (i.e., reduction in counterion binding) under these conditions.[273]

6.3.3 Branched 1-Methyl, 4-Alkyl Pyridinium Surfactants

The final type of surfactants that I will consider in this Chapter includes the 1-methyl, 4-alkyl pyridiniums. In Table 6.10 I consider two sets of these surfactants. Both sets involve a constant number of 12 carbons in the alkyl chain attached to the pyridinium ring. In the first set, as the primary chain length is decreased, the secondary chain length is positioned at the nodal carbon that results in a symmetric tail. For example, for the C_9(6-C_3) surfactant, there are three carbons in the secondary chain, and three carbons along the primary chain following the nodal carbon. As the primary chain shortens, the secondary chain lengthens to maintain the constant 12 carbons. Accordingly, the nodal carbon approaches the pyridinium ring. In the second set, for a fixed decyl primary chain, a two-carbon side chain is moved to various positions along the primary chain. Note that these surfactants provide interesting examples of movement of the position of the nodal carbon. The branched alkyl sulfonates in Section 6.3.1 and branched alkyl benzene sulfonates in Section 6.3.2 both exhibited branching at the 1-position.

Table 6.9: Optimal Micelle Shape, Size, and Degree of Counterion Binding Results for Branched Benzene Sulfonate Surfactants at $T = 25°C$

Surfactant	Ion	Shape	N_{total}[a]	N_{sph}[a]	N_{cyl}[a]	$l_{c,sph}$ (Å)[a]	$l_{c,cyl}$ (Å)[a]	β[b]
C_5(1-C_3)BZSFN	Na+	Spherocylinder	33	18	15	11.5	10.1	0.56
C_6(1-C_2)BZSFN	Na+	Spherocylinder	47	23	24	12.6	10.9	0.59
C_7(1-C_1)BZSFN	Na+	Spherocylinder	44	33	11	14.1	11.7	0.60
C_8BZSFN[c]	Na+	Sphere[d]	32	32	0	13.2		0.49
		Spherocylinder[e]	41	37	4	14.6	12.2	0.59
C_6(1-C_4)BZSFN	Na+	Spherocylinder	32 (17[171f])	20	12	12.6	11.2	0.51
C_7(1-C_3)BZSFN	Na+	Spherocylinder	45	29	16	14.1	12.0	0.56
C_8(1-C_2)BZSFN	Na+	Spherocylinder	45 (53[171f])	36	9	15.3	12.6	0.57
C_9(1-C_1)BZSFN	Na+	Spherocylinder	48 (63[171f])	42	6	16.1	13.4	0.58
C_{10}BZSFN[c]	Na+	Sphere[d]	47	47	0	16.0		0.52
		Spherocylinder[e]	55	52	3	17.2	14.1	0.60
C_7(1-C_5)BZSFN	Na+	Spherocylinder	32 (21[171g])	25	7	14.1	12.3	0.48
C_8(1-C_4)BZSFN	Na+	Spherocylinder	36	28	9	14.6	12.8	0.50
C_9(1-C_3)BZSFN	Na+	Spherocylinder	44 (24[171g])	37	7	16.1	13.8	0.54
C_{10}(1-C_2)BZSFN	Na+	Spherocylinder	50 (57[171g])	46	4	17.2	14.4	0.56
C_{11}(1-C_1)BZSFN	Na+	Spherocylinder	62	58	4	18.7	15.2	0.60
C_{12}BZSFN[c]	Na+	Sphere[d]	57	57	0	17.9		0.53
		Sphere[e]	60	60	0	18.9		0.59

Predictions made using the MHA solubility model and the MT framework.

[a] Micelle size properties: N_{total} – total aggregation number, N_{sph} – aggregation number in sphere or spherical end caps, N_{cyl} – aggregation number in cylindrical body, $l_{c,sph}$ – radius of sphere or spherical end caps, $l_{c,cyl}$ – cross-sectional radius of cylindrical body. Experimental values in parentheses when available (see text for discussion).

[b] Degree of counterion binding (number of bound counterions per surfactant).

[c] Linear molecules included for comparison with the branched results.

[d] Using a tail identification of the full alkyl tail and half of the benzene ring.

[e] Using a tail identification of the full alkyl tail and the full benzene ring.

[f] Measured at $22°C$.

[g] Measured at $30°C$.

Although there is scant experimental data available to compare against, the observed trends in the MT and CSMT predictions are interesting. In the first set, in which the primary chain varies in length, there is a maximum in the cmc observed in the intermediate cases, which have the nodal carbon positioned at the center of the primary chain. In the second set, in which the chain lengths are fixed, but the position of the nodal carbon is systematically varied, there is also a maximum in the cmc, observed when the secondary chain is located in the middle of the chain. In the first set, there are several competing effects, due to the change in micellar dimensions (which are summarized in Table 6.12). However, in the second set, the effect is clearly due to packing, since the optimal shape for these surfactants was found to be a sphere, and the size of the sphere was found to be dictated by the maximum extension of the primary chain. Note that this maximum extension is a result of the energetic benefit of reducing the area per surfactant – the accompanying increase in charge density results in an increase in counterion binding,[273] reducing the charging free-energy penalty through partial charge neutralization, and the interfacial free-energy penalty is also reduced.

In the single-chain, mean-field packing framework described in Chapter 4, lateral pressures are applied in order to enforce constant density constraints throughout the micelle core. For a sphere, the distribution of chain volume in space as a function of radius should be quadratic, since the volume of spherical shells is $4\pi r^2 \Delta r$ for some fixed shell width Δr, where r is the radial coordinate. Additionally, since one end of the chain is tethered at the micelle core-water interface, where the head attachment occurs, there are additional connectivity constraints to consider. Now, if a chain has a natural distribution of volume under a uniform pressure field within the spherical confinement of the core in which too much volume is located in the center of the core, the applied lateral pressure must be increased in that region, raising the constraint-related packing free energy (see Section B.3 of Appendix B for a description of the decomposition of the packing free energy into confinement and constraint contributions), and, therefore, the overall packing free-energy penalty. This is precisely the situation experienced with the second set of 1-methyl, 4-alkyl

pyridiniums examined here. When the branching is very far down the chain, near the primary chain terminus, this volume can be readily accommodated within the core, since the chain is flexible and can readily rotate to place the associated groups on average closer to the interface of the core, where there is more volume available to fill. Likewise, when the branching is very close to the head, the added volume of the branch is already in the region of the micelle core possessing the most volume (according to $4\pi r^2 \Delta r$). However, when the branching is in an intermediate position, and the presence of the aromatic ring in the head group reduces the freedom of the chain to rotate, it is more difficult to arrange configurations which can accommodate for the volume of the branch. To put this in quantitative terms for the surfactants considered here, the packing free energy at maximum extension for the $C_{10}(j\text{-}C_2)$ surfactants ($12.7\mathring{A}$) is 2.98 k_BT for $C_{10}(2\text{-}C_2)$, 3.11 k_BT for $C_{10}(4\text{-}C_2)$, 3.07 k_BT for $C_{10}(6\text{-}C_2)$, and 2.96 k_BT for $C_{10}(8\text{-}C_2)$. An examination of the decomposition of the free energy terms (not shown) indicates that this is the dominant contribution leading to the observed cmc maximum predicted. From the limited experimental data in Table 6.10, we can see that there is indeed an increase in cmc in going from $C_{10}(8\text{-}C_2)$ to $C_{10}(6\text{-}C_2)$. Additionally, the shapes of the micelles were identified experimentally to be spheres.[61] However, there is unfortunately not enough data to confirm the existence of the maximum predicted.

The accuracy of the predictions can be examined in Table 6.11. As with the linear 1-methyl, 4-alkyl pyridinium surfactants, there tends to be an overprediction in the MT modeling which is as yet unexplained, and should provide an avenue for future study. For the systems which had experimental data, I conducted a set of trial calculations using packing data without the aromatic ring in the head group. This allows the chain to sample space more freely, reducing the packing free-energy penalty. Indeed, this does reduce the cmc's predicted using the MHA solubility model, to less than 1 k_BT, perhaps indicating that a strict head/tail identification (in which the aromatic group may not participate in the micelle core to any extent) is too severe an assumption. I will discuss this again in the context of proposing a neutral atom packing approach in Section 7.3.3 of Chapter 7. Note, however, the excellent

agreement of the MHA solubility model with experiment when used in the context of CSMT modeling. In fact, throughout this Chapter, we have seen that the CSMT modeling results using the MHA solubility model have been excellent, indicating a promising application for the CSMT framework in the study of branched surfactants.

Finally, in Table 6.12, I present a brief summary of the shape results for the first set of 1-methyl, 4-alkyl pyridinium surfactants considered. As already mentioned, the second set, having a fixed main chain length and an optimal spherical micelle configuration, exhibit properties almost identical to the $C_{10}(8\text{-}C_2)$ surfactant. Note that all these surfactants are predicted to be either spheres or spherocylinders with negligible aggregation number in the cylindrical body. The reduction in the radius of the spheres or spherical endcaps is directly related to the geometric structure of the tail, which becomes more squat as the secondary chain is increased and the primary chain is decreased in length. For $C_9(6\text{-}C_3)$ and $C_{10}(8\text{-}C_2)$, experimental degrees of counterion binding are available and exhibit a very slight decrease in value. The predicted values also exhibit a slight decrease. The experimental identification of shape for these two surfactants is also that they are spherical.[60] Nusselder and Engberts, who have carefully studied the 1-methyl, 4-alkyl pyridinium surfactants to assess the relation between structure and morphology also indicate that certain of the surfactants examined here can form vesicles[60] (including $C_{10}(6\text{-}C_2)$ and $C_7(2\text{-}C_5)$). This particular geometry is not yet implemented in the MT framework which I have developed, but the work of Yuet and Blankschtein indicates how such a geometry can be considered.[86]

Table 6.10: CMC Results for Branched 1-Methyl,4-Alkyl Pyridinium Surfactants at $T = 25°C$

		CMC (mM)	MT Pred. CMC (mM)[a]				CSMT Pred. CMC (mM)[b]			
Surfactant	Ion	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_7(2-C_5)MPYR	I-	-	12	72	15	89	6±1 [5]	35±7 [4]	9±2 [5]	1.56±0.02 [3]
C_8(4-C_4)MPYR	I-	-	29	124	70	440	7.6±0.2 [6]	40.3±0.7 [9]	21.3±0.3 [6]	2.97±0.06 [6]
C_9(6-C_3)MPYR	I-	4.64[61]	**21**[c]; **12**[d]	89	51	320	**4.96±0.02** [4]	27.1±0.5 [4]	13.3±0.3 [4]	2.07±0.04 [4]
C_{10}(8-C_2)MPYR	I-	3.74[61] [30°C]	**14**[c]; **8.6**[d]	71	38	260	**3.61±0.09** [4]	19.0±0.5 [4]	9.4±0.2 [5]	1.61±0.02 [4]
C_{10}(2-C_2)MPYR	I-	-	15	78	42	300	4.02±0.03 [6]	21.0±0.2 [5]	10.56±0.06 [6]	1.789±0.009 [6]
C_{10}(4-C_2)MPYR	I-	-	17	86	46	310	4.4±0.1 [6]	23.5±0.7 [6]	11.4±0.2 [5]	1.93±0.04 [6]
C_{10}(6-C_2)MPYR	I-	-	17	83	47	280	4.27±0.07 [5]	23.4±0.4 [4]	11.2±0.2 [5]	1.84±0.03 [5]
C_{10}(8-C_2)MPYR	I-	3.74[61] [30°C]	**14**[c]; **8.6**[d]	71	38	260	**3.61±0.09** [4]	19.0±0.5 [4]	9.4±0.2 [5]	1.61±0.02 [4]

The MT and CSMT predictions which are closest to the experimental values are bolded.

Unavailable experimental cmc's are marked with the symbol "-".

[a] MT predictions do not have error bars, since the theory is deterministic based on user input.

[b] The error bars in the CSMT predictions represent the standard deviation in the predicted cmc arising from differences between the independent simulations run. The number of simulations is denoted in square brackets. Note that the four solubility models are applied to the same set of simulations, but difficulties with numerical convergence in certain cases may lead to some points being discarded.

[c] Packing calculated with the benzene ring present in the head region.

[d] Packing calculated with the benzene ring absent in the head region.

Table 6.11: Free Energy of Micellization Comparison for Branched 1-Methyl,4-Alkyl Pyridinium Surfactants at $T = 25°C$

		g_m^* (k_BT)[a]	MT Error (k_BT)[b]				CSMT Error (k_BT)[b]			
Surfactant	Ion	Expt.	MHA	MHA2	KWB	KEKSS	MHA	MHA2	KWB	KEKSS
C_9(6-C_3)MPYR	I-	-9.39	**1.51**[c] **0.95**[d]	2.95	2.40	4.23	**0.07**	1.76	1.05	-0.81
C_{10}(8-C_2)MPYR	I-	-9.60 [30$°C$]	**1.32**[c] **0.83**[d]	2.94	2.32	4.24	**-0.04**	1.62	0.92	-0.84

The MT and CSMT predictions which are closest to the experimental values are bolded.

[a] MT predictions do not have error bars, since the theory is deterministic based on user input.

[b] The error bars in the CSMT predictions represent the standard deviation in the predicted cmc arising from differences between the independent simulations run. The number of simulations is denoted in square brackets. Note that the four solubility models are applied to the same set of simulations, but difficulties with numerical convergence in certain cases may lead to some points being discarded.

[c] Packing calculated with the benzene ring present in the head region.

[d] Packing calculated with the benzene ring absent in the head region.

Table 6.12: Optimal Micelle Shape, Size, and Degree of Counterion Binding Results for Branched 1-Methyl,4-Alkyl Pyridinium Surfactants at $T = 25°C$

Surfactant	Ion	Shape	N_{total}[a]	N_{sph}[a]	N_{cyl}[a]	$l_{c,sph}$ (Å)[a]	$l_{c,cyl}$ (Å)[a]	β[b]
C_7(2-C_5)MPYR	I-	Spherocylinder	19	17	2	11.9	10.4	0.42
C_8(4-C_4)MPYR	I-	Spherocylinder	17	16	1	11.1	9.9	0.35
C_9(6-C_3)MPYR	I-	Sphere	23	23	0	12.6		0.40 (0.79[61])
C_{10}(8-C_2)MPYR	I-	Sphere	31	31	0	13.9		0.45 (0.80[61])

Predictions made using the MHA solubility model and the MT framework.

[a] Micelle size properties: N_{total} – total aggregation number, N_{sph} – aggregation number in sphere or spherical end caps, N_{cyl} – aggregation number in cylindrical body, $l_{c,sph}$ – radius of sphere or spherical end caps, $l_{c,cyl}$ – cross-sectional radius of cylindrical body.

[b] Degree of counterion binding (number of bound counterions per surfactant). Experimental values in parentheses when available (see text for discussion).

6.4 Conclusions

In this Chapter, I implemented a molecular-thermodynamic (MT) model for surfactant micellization to study 28 singly-branched surfactants, including 9 branched alkyl sulfonates, 12 branched alkyl benzene sulfonates (also commonly referred to as "linear alkyl benzene sulfonates"), and 7 branched 1-methyl, 4-alkyl pyridiniums. I considered a variety of branch positions and chain lengths in order to explore the effect of the structure of a surfactant tail on the micellization properties of the surfactant as a whole. Additionally, I performed a large set of computer simulations of each surfactant in monomer and micelle environments to implement a combined computer-simulation/molecular-thermodynamic (CSMT) framework, which aims to more-accurately quantify the hydrophobic effect through a microscopic consideration of surfactant atomic group hydration in the two environments. In both MT and CSMT modeling, as in Chapter 5, four solubility models were examined (MHA, MHA2, KWB, and KEKSS models, according to Section F.5 of Appendix F). The head/tail identifications used for MT modeling were obtained from the CSMT computer simulations, according to the dividing surface approach which I developed in Section 3.3 of Chapter 3.

The modeling results, in agreement with the results of Chapter 5 for linear surfactants, indicate that the MHA solubility model, in the context of both MT and CSMT predictions, presents the best, most consistent model for branched surfactant micellization. In most of the cases considered here, the CSMT predictions were more accurate, indicating that the CSMT framework (which has the advantage of being less sensitive than the MT framework to the particular surfactant head/tail identification used) may be particularly useful in studying more complex surfactant structures. The KWB model also provided good predictions for the full set of alkyl benzene sulfonates, due to its accurate fitting to alkyl arene aqueous solubility data. However, this model was inconsistent in its accuracy when applied to the other types of surfactants considered, and is therefore not recommended for use. The MHA2 and KEKSS models are also not suitable.

The observed trends within the sets of branched surfactants considered provide interesting insights into the role of the surfactant tail in determining micellization behavior.

For all surfactant types, the lengthening of either the primary or secondary chains decreased the cmc, according to the hydrophobic effect.[34] However, the impact of each additional methylene added is not constant and depends on which chain the methylene is added to. For example, lengthening of the primary chain increases overall micelle dimensions, but lengthening of the secondary chain at times leads to a slight reduction in the optimal micelle dimensions, as seen in Table 6.6. Although the same number of carbons may be present in these two cases, the impact on aggregation number is quite different.

With both the alkyl sulfonates and the alkyl benzene sulfonates, a more symmetric distribution of carbons between the primary and secondary chains leads to an increase in cmc, due to a decrease in chain length and an increase in the micelle core-water interfacial area per surfactant relative to a more asymmetric distribution. Properties such as counterion binding are closely related to the area per surfactant, and the free-energy penalties which offset the negative transfer free energy of the surfactant tail are also sensitive to this quantity. These more symmetric tails also lead to lower aggregation numbers overall.

Finally, the 1-alkyl, 4-methyl pyridinium surfactants exhibited interesting behavior when the total number of carbons in the alkyl tail were kept fixed, but the distribution of carbons between primary and secondary chains and the positions of the branch were changed. As more of the chain volume is concentrated at the center of the chain, due to positioning of the nodal carbon, a slight increase in cmc is predicted, which can be attributed partially to packing considerations. Such subtle effects cannot be readily explored in linear surfactant systems, and may provide avenues for surfactant design to achieve a desired performance.

The surfactant systems in this Chapter provided an initial test of the new theory which I have developed for the study of branched surfactant micellization. Future directions for this research are presented in Chapter 7, which follows.

Chapter 7

Conclusions and Future Research Directions

7.1 Overview

In the preceding Chapters, I have laid out, in comprehensive detail, a theoretical and simulations-based framework for predicting surfactant micellization behavior as a function of surfactant structure and solution conditions. I have also been able to demonstrate the effective application of this framework to the study of over 80 surfactants of various classes: including linear and branched surfactants, alkyl and alkyl arene tails, and anionic, cationic, zwitterionic, and nonionic heads. In this Chapter, I step back and provide an overview of what I have accomplished, and where I think that this research is headed. In Section 7.2, I provide a detailed summary of the key aspects of each Chapter. I also discuss briefly the nature of the supporting material supplied in the Appendices of this Thesis. In Section 7.3, I suggest several directions for future research, motivated by my observations in conducting the research for this Thesis. Finally, in Section 7.4, I close the Thesis with a few concluding remarks.

7.2 Thesis Summary

In Chapter 2, I presented the full molecular-thermodynamic (MT) framework which I have used to make quantitative predictions of surfactant micellization behavior based on input of the surfactant chemical structure. In Section 2.2 of Chapter 2, I demonstrated how the equations of bulk phase thermodynamics can be applied to a solution containing aggregating species, through the derivation of an expression for the chemical potential of each aggregate or free molecule (generically referred to as solution components) starting with an expression for the overall solution Gibbs free energy, which includes formation and mixing free-energy contributions.[81] The component chemical potential was shown to be a function of a modified mole fraction, which itself was defined in a manner which reduces to the usual mole fraction in a fully-dispersed system. Using the principle of multiple chemical equilibrium,[130,131] I derived a formula for the modified mole fraction that, in the case of a micelle, is a function of the difference in reference-state chemical potentials between the micelle at infinite dilution and its constituent surfactants and bound counterions. I defined this difference as the free energy of micellization, g_{mic}, which, along with entropic penalties representing the effect of confining the surfactants and the bound counterions to remain in the micelle, yields the modified free energy of micellization, g_m. I described how micellization properties such as the critical micelle concentration (cmc) and the micelle shape, size, and composition population distribution are functions of g_m. In Section 2.3, I presented a phenomenological thought process which enables estimation of g_m from the chemical structure of the surfactants and ions in solution. This thought process enables calculation of the free energy along a consistent thermodynamic path which traverses the initial state of a micelle constituent at infinite dilution through to the final state of that constituent fully integrated into the micelle structure. This path involves a series of discrete steps related to the self-assembly of the micelle, and requires an identification of the surfactant head (hydrated) and tail (dehydrated) regions. I was able to resolve a previous inconsistency in the model regarding the nature of the free energies of each step (whether Gibbs or Helmholtz)

by introducing a pressure-volume work term for the creation of a void in the bulk solvent to accommodate the creation of the micelle core. However, I showed that this term is in practice negligible, indicating that previous results obtained using the MT theory are effectively unaffected. In Section 2.4, I described several additional aspects of the computational process involved in carrying out MT predictions. I introduced the computer program CS-PREDICT which I developed to implement the MT framework, including a description of the algorithms contained therein. I also demonstrated how one computes properties for finite spherocylindrical micelles from optimized finite sphere and infinite cylinder g_m values. Finally, for ionic surfactants, I demonstrated an appropriate treatment for the electrostatic diffuse layer of counterions surrounding the micelle.[90] This diffuse layer is considered when closing mass balances for all the charged species in solution.

In Chapter 3, I discussed the combined computer-simulation/molecular-thermodynamic (CSMT)[108–110] framework, which makes use of computer simulations to carry out more-detailed calculations of the hydrophobic driving force for micellization. In Section 3.2, I described all the details of the CSMT framework, including the notions of a free energy of hydration and dehydration, the calculation of a modified transfer free energy for a surfactant tail, and a comparison of this modified CSMT transfer free energy with the original MT free energy, demonstrating that differences between the two values are due to: (i) partial transfer free-energy contributions from hydrophobic groups in the surfactant head region, due to partial dehydration in moving from the monomer to the micelle environments, and (ii) effects due to the changing structure of water about hydrophobic entities of differing size and shape,[32] such as the hydrophobic tail of a surfactant in the monomer state versus the hydrophobic micelle core (a much larger entity possessing a different geometry). In Section 3.3, I presented my novel concept for a method of identifying surfactant head and tail using computer simulations of a micelle and a dividing surface analysis inspired by the Gibbs equimolar dividing surfaces used in bulk phase thermodynamics.[142] I made the case that a proper use of this head/tail identification scheme requires an iterative approach to arrive at the location of a dividing surface which separates tail groups in the micelle

core from bulk solvent, ions, and head groups in a self-consistent manner. Application of this iterative approach yields results which are in good agreement with previous rules of thumb for the surfactant head, but also provides new identifications for surfactants which have not previously been characterized. In Section 3.4, I enumerated all aspects of the simulation protocol that I have used in conducting molecular dynamics simulations for CSMT calculations. Two key molecular dynamics programs were used: (i) Program BUILDER, a program which I conceived and developed for the automation of surfactant monomer and micelle simulation initialization, for systems containing any number of surfactant species, and (ii) GROMACS, a freely-available, open-source, parallelized code for fast and efficient simulation of all-atomistic surfactant systems using computer cluster resources. In Section 3.5, I described the various analyses necessary to obtain the CSMT transfer free energy from representative simulations for use in the MT theory. Use of this free-energy quantity, rather than direct use of hydration and dehydration contributions to g_m, allows for the prediction of micellization properties after simulating a single monomer and a single micelle, thereby avoiding the need that would otherwise arise for simulations at each possible aggregation number (which would make the CSMT approach entirely impractical). The analyses required include: (a) determination of radial distribution functions of groups about the center of mass of the micelle core, for the purpose of dividing surface location and head/tail identification, (b) counting of the hydrating contacts in monomer and micelle environments, for the purpose of computing fractional degrees of hydration (used in computing the free energies of hydration and dehydration), and (c) the determination of solvent accessible surface areas about individual surfactant atomic groups and about the micelle core (used in computing the free energy of hydration). These calculations have been implemented as a series of scripts and programs which I authored, and which I refer to as Program ANALYZER. Several routines from the GROMACS library are utilized in Program ANALYZER.

In Chapter 4, I presented a single-chain, mean-field packing theory[219] for determining the packing free-energy contribution to g_m as a function of the surfactant tail, and micelle shape and size. In Section 4.2, I described how the packing free-energy

contribution is a positive penalty which arises from the physical confinement of a surfactant tail within the micelle core due to the hydrophobic driving force. This confinement necessarily results in a loss of entropy for the chain confined relative to its free, unconfined state. Additionally, the experimental observation that the micelle core is liquid-like[34,186,211,212] leads to a constaint of constant density (incompressibility) which provides another penalty. Using the Maximum Entropy Principle,[277] I showed how one can readily derive a probability distribution function for a central chain, under the influence of a pressure field which enforces the constant density constraint (operating as a Lagrange multiplier to the chain volume). This pressure field is a mean field which represents the effect of all the other chains upon the configurations of the central chain. Once the pressure field is solved for, I explained how the packing free energy itself is calculated. However, in order to calculate the pressure field, one must first specify a model for the degrees of freedom of the central chain itself. In Section 4.3, I described the Rotational Isomeric State (RIS) theory,[213] which replaces continuous dihedral profiles with a set of discrete states representing energetic minima, and considers bond lengths and angles to be fixed at equilibrium values. I then provided two examples of useful RIS models, one for linear alkyl chains,[216] and one for branched alkyl chains (limited to chains with branches spaced several groups apart).[217] My use of a branched alkyl RIS model represents the first time that branched surfactant tails have been modeled without additional simplifying assumptions, such as assuming that the dihedral states of the branch connection to the nodal carbon are equipotential.[119] Since the RIS model for branched chains is systematic in its approach, I have been able to extend this model to arbitrarily-spaced branched chains, using algorithmic generation of the packing models.

In Chapter 5, I brought together the theoretical and simulations-based approaches described in Chapters 2 to 4 for the purpose of making practical predictions of micellization properties. In Section 5.2, I introduced a total of 55 linear surfactants that I would consider, divided into charged (anionic, cationic, and zwitterionic) and nonionic (ethoxylated and sugar-based) surfactant classes. I discussed the head/tail identifications obtained from the numerous computer simulations that I performed,

as well as the various molecular descriptors and their values which I used to model these surfactants. In Sections 5.3 and 5.4, I discussed the key results of the modeling, focusing on comparison of predicted cmc's to experimental cmc's. This comparison provides an excellent quantitative test of any micellization theory, due to the exponential dependence of the cmc on g_m. Therefore, small errors in the prediction of g_m, due to errors in any one of the free-energy contributions, would yield large errors in the cmc. I compared MT and CSMT cmc predictions, and additionally explored the effect of the solubility model used (to obtain the transfer free energy) on the results. I found that one particular solubility model was most effective: a model based on the Gibbs free energy of solution correlations for linear alkanes due to Abraham,[239] which I augmented with fitting to experimental alkyl arene aqueous solubility data[278,279] in order to expand the parameter set to include aromatic groups (labeled the MHA solubility model). In general, the modeling results using the MT framework and the MHA solubility model yielded good agreement with experimental cmc's. One exception includes the sulfobetaine zwitterionic surfactants, which, in the computer simulations, displayed a good deal of head-head association between the oppositely-charged ammonium and sulfonate regions. Since this effect was not explicitly modeled theoretically, an underprediction in the cmc's was observed, which provides a motivation for the development of improvements in the treatment of surfactant heads, a direction for future research discussed in Section 7.3.2. Although the CSMT approach did not outperform the MT approach for the linear surfactants considered, it did yield the correct trend for the cmc's of α and β anomers of both octyl glucoside and dodecyl maltoside. The MT approach, lacking a sensitivity to stereochemistry, is unable to discern such subtle effects.

In Chapter 6, I applied the MT and CSMT frameworks to a variety of branched surfactants, limiting the surfactant head types considered while introducing more variety into the selection of the tails. In Section 6.2, I introduced the 28 branched surfactants that I would consider, divided into three types: the alkyl sulfonates (anionic), alkyl benzene sulfonates (anionic), and 1-methyl, 4-alkyl pyridiniums (cationic). I found that the head/tail identifications obtained from computer simulations using

my dividing surface approach are extremely consistent within each type. I then discussed the molecular descriptors and their values which are associated with these identifications. In Section 6.3, I presented cmc and optimal micelle shape, size, and degree of counterion binding predictions. These branched surfactants were found to form very squat spherocylinders, indicating that they perhaps prefer a prolate ellipsoid geometry (a non-regular geometry that is not considered explicitly in my framework). The cmc results obtained using the MHA solubility model were found to provide the best agreement with experiment, as in Chapter 5. Finally, I called attention to the improved predictions of the CSMT model, which consistently resulted in very good predictions of the cmc, especially in the case of the 1-methyl, 4-alkyl pyridiniums, where the MT approach yielded a more substantial overprediction.

The key outcome of my modeling efforts in Chapters 5 and 6 include the identification of the MHA solubility model as the superior solubility model for use in estimating the transfer free energy of surfactant tails (and, in the case of the CSMT modeling, the partial transfer of hydrophobic groups in the surfactant head). Although the MT approach outperformed the CSMT approach in the linear surfactant study, the results of both approaches were generally quite good. I identified a few exceptions where the CSMT approach deviated more drastically from experiment, but identified this as the result of uncertainty as to which transfer free-energy contributions to assign to certain groups in the head. In the original publications of this approach, this was also a concern, specifically in the case of alkyl ethoxylates.[109] However, in modeling the branched surfactants, which all possessed compact head groups, the result was quite clear: the CSMT approach provided improved modeling capabilities, for all surfactant types considered.

In support of this work, I have assembled additional supplementary material into Appendices to this Thesis. Next, I will briefly summarize their contents along with some commentary regarding where they fit into the frameworks that I have developed.

In Appendix A, I describe vector rotations obtained using either Euler angles or quaternions (preferred when specifying an axis and angle for rotation). I then describe a two-step method for internal coordinate generation when building molecules up from

a full specification of the degrees of freedom of a molecule using bond lengths, angles, and dihedrals. These methods are used in both my packing software and in Program BUILDER, and are provided here for reference purposes.

In Appendix B, I provide a description of Markov-Chain Monte Carlo techniques for chain packing. I show how the packing free energy can be decomposed into confinement and constraint contributions, each of which can be calculated using separate Monte Carlo simulations. Computation of the constraint contribution is the most challenging step, involving use of the Overlapping Distribution Method[203,280,281] following the solution of the pressure fields. This Monte Carlo approach to determining the packing free energy becomes necessary for long or highly-branched chains (with greater than 18 rotatable bonds), since the degrees of freedom for these systems, even under the RIS approximation, still exceed the memory capabilities of conventional desktops when performing the usual, full-enumeration approach for chain packing. The Monte Carlo approach has a very low memory overhead, and begins to solve faster than the full-enumeration approach at around 16 rotatable bonds.

In Appendix C, I describe the algorithms underlying Program BUILDER, which I developed to automate the generation of numerically-stable initial conditions for surfactant micelle simulations. This Program makes use of basic chemical graph theory to build a topology of the bonds, angles, and dihedrals, drawing from a library of force-field data which I assembled from various sources (specifically, the OPLS-AA databases of the GROMACS program,[188,190] the publications of Canongia Lopes and coworkers[191–194] and Anderson and Wilson,[235] and my own novel force-field development efforts, described in Appendix E). This Program also implements various molecular dynamics algorithms, including Velocity Verlet dynamics[282] and Langevin dynamics.[283] I also include a discussion of random number generation, used in the packing program for both full-enumeration and Monte Carlo calculations in addition to the Langevin dynamics in Program BUILDER.

In Appendix D, I derive all the force expressions used in the molecular dynamics algorithms employed in Program BUILDER. I have included these derivations and results in this Thesis to serve as reference material for future users and developers

of my Program. These force expressions, especially for dihedrals, are not frequently provided, since the potential expressions are simpler in functional form. Nevertheless, it is the force expressions that are used in the update equations of both the Velocity Verlet and Langevin dynamics approaches.

In Appendix E, I discuss a variety of ab initio techniques that can be used to develop force-field parameters for new surfactant structures that may exhibit an incomplete specification in the existing literature. I present a particular example of the alkyl benzene sulfonates, which, within the OPLS-AA force field, were not fully specified.

Finally, in Appendix F, I present a complete summary of the molecular properties which I made use of in carrying out my MT and CSMT modeling. These include: (i) atomic weights and van der Waals radii for size determination, (ii) liquid water density and molarity, used in converting between surfactant mole fractions and concentrations, (iii) group contributions to surfactant tail volume, which I parameterized based on a large set of experimental specific gravity data,[265] converted into molecular volumes, (iv) estimates of surfactant tail solubilities, obtained from four different models, with the efficacy of the various models evaluated in Chapters 5 and 6, (v) estimates of surfactant tail-water interfacial tension, obtained from experimental data for analogous oils, and (vi) surfactant tail packing free energies, fit to polynomials and assembled into a library for deployment in MT and CSMT calculations (since fresh calculation of packing free energies as-needed would be too computationally expensive for the demands of the end-user).

7.3 Future Research Directions

7.3.1 Library Development for Surfactant Modeling

One of the practical goals of my research has been to create a computational framework that a user can use to rapidly model the micellization of surfactant systems of interest, in order to make quantitative predictions, explain qualitative trends, and

to guide formulation and surfactant synthesis efforts. As is perhaps clear from this Thesis, however, such a framework for predicting surfactant micellization behavior can be quite involved, incorporating inputs from computer simulations, which require computer cluster resources, and a statistical-thermodynamic theory for chain packing, which involves calculations that can also require substantial computing resources. In order to make the software available to a user who may not possess these resources, it is critically important to be able to develop libraries of simulation and packing data for the most common surfactant structures that will be encountered in practice. Even for users with expertise in computer simulations and modeling, incorporation of data into a library provides a useful organizing principle, divides up the substantial effort involved in modeling into manageable, smaller tasks, and provides an opportunity for the sharing and merging of the data between collaborators.

Expanding Existing Packing Libraries

In Section F.7 of Appendix F, I have provided the details of my current library of linear and branched surfactants packing data. All possibilities for singly-branched surfactants with a main chain length of up to C_{10} have been included, and many examples involving aromatic rings are also provided. Nevertheless, there is much room for expansion into: (i) longer-chain surfactants, (ii) surfactants possessing multiple branches, which becomes a combinatorial problem, (iii) surfactants containing additional functional groups in the tail (ethers, for example), and (iv) surfactants with alternative chemistries, including fluorinated and siloxane-based surfactants.[12]

Additionally, I have not yet attempted to compress this data beyond the use of polynomials for each individual case. It seems likely that homologous series of surfactants will exhibit packing behavior which can be characterized by a single polynomial with the introduction of functional dependence on, say, branch node and primary and secondary chain length. Perhaps further investigation will yield a functional form that allows for accurate interpolation between structures, thereby reducing or eliminating the need for additional packing calculations when dealing with new surfactants. Such efforts will likely require a substantial statistical analysis of a rich variety of chain

types.

Simulations of Cylindrical and Bilayer Aggregates

From a simulations perspective, in this Thesis I have focused on simulations of finite spheres. Accordingly, my library of CSMT transfer free energies and accompanying fractional degrees of hydration, solvent accessible surface areas, and head/tail identifications is obtained from this shape, and then assumed to remain valid when predicting the properties of other micelle geometries. Stephenson and coworkers discuss the challenge in determining realistic boundary conditions for simulating infinite cylinders and bilayers using periodic boundaries.[109] However, another challenge arises in the analysis of these shapes, since several of the analysis routines in GROMACS will need to be modified and re-compiled in order to support handling of these infinite, periodic geometries. For example, it is necessary to center a simulation about the simulated micelle, so that it does not drift across a periodic boundary when calculating time-averaged radial distribution functions for groups from the micelle core center of mass (e.g., in head/tail identifications). Otherwise, if the micelle partially crosses the boundary, the center of mass will be erroneously calculated as being located somewhere in the bulk aqueous solution, due to the emergence of part of the micelle on the other side of the simulation cell. Centering for finite spheres about a single point is straightforward. For cylinders and bilayers, however, the centering must be accomplished along a line and plane, respectively. In addition, flexure of the cylinder and the bilayer could complicate this procedure. Nevertheless, modeling of these other micelle shapes is an important priority in establishing how well the results obtained from spheres can be extrapolated to other shapes.

Direct Use of Simulations to Obtain Micelle Properties

Another interesting use for the simulations data obtained includes performing static and dynamic analyses of the simulated micelles. Bogusz and coworkers demonstrated the rich information that can be extracted from molecular dynamics simulations in their studies of octyl glucoside micelles.[96,97] One can obtain overall micelle shape and

size information, including the amount of eccentricity observed in the approximately spherical micelles, as well as details of the finer structure of the micelle, including average tail group locations, end-to-end distances, rate constants for changes in dihedral state, and the sizes of head group clusters and exposed core regions.[96] Stephenson and coworkers have demonstrated how molecular descriptors, such as the head/tail bond shielded area used in the calculation of interfacial free-energy penalties, can be determined from these simulations.[109]

Multicomponent Surfactant Systems

A move from single-component surfactant systems to multicomponent systems will also introduce new challenges. From a packing perspective, I have presented the theory and developed the software to enable the packing of multiple chains of different types. Polynomials for these systems involve fitting not only to micelle shape and size, but also to composition. Accurate fits typically require cross-terms within the polynomial (i.e., terms involving powers of both the radial variable and the compositional variable in a binary system of fixed shape). As the number of chains increases to three and beyond, the number of parameters needed to accurately fit the packing data increases substantially (e.g., Goldsipe and Blankschtein used 80 parameters in their fitting for ternary systems,[148] but binary systems typically require only 12 parameters, and single-component systems only four to five).

For multicomponent computer simulations, I have purposefully developed Program BUILDER to accommodate multicomponent surfactant micelles. These micelles are constructed according to user specification of an aggregation number for each surfactant (fixing micelle composition). The surfactants are then distributed randomly within the micelle, to avoid any initial bias within the simulation. The MT and CSMT approaches both readily accommondate multicomponent predictions, and I have presented the underlying thermodynamics in Chapter 2 to reflect this. It is an outstanding question as to whether simulations of mixed micelles are always necessary to obtain a composition-weighted transfer free energy. It is possible that an interpolation between CSMT transfer free energies obtained from pure micelles could

yield an accurate approximation for a mixed micelle. If this proves true, then the library of pure micelle simulation data that I have compiled as part of my doctoral research can be immediately leveraged into mixed micelle calculations.

Force-Field Development for Surfactants

Finally, it is noteworthy that computer simulations of more complex surfactants will almost necessarily require some amount of force-field development, since even a common surfactant type such as the alkyl benzene sulfonates required some work in this area to arrive at a full specification which is compatible with the OPLS-AA force field. The strategies that I have laid out in Appendix E should be useful in this effort. I would recommend that eventually all common surfactant types be investigated in this manner. Some structures that I have not yet considered include the amine oxides, betaines, dimethyl benzyl ammoniums, and dimethyl ethanol ammoniums. Although not presented here, I have considered phosphate surfactants, mainly in the context of fluorinated applications, and force-field parameters are available for common ions such as carboxylates (soaps). In terms of surfactant tails, a recent paper has called into question the OPLS-AA force field specification for alkanes when applied to long chains.[284] In my own investigation, I have also found some problems when simulating certain long chain C_{14} and C_{16} surfactants, where the micelle core may exhibit crystalline behavior. I recommend inclusion of these new parameters in future molecular dynamics simulation modeling using the CSMT approach, after rigorous comparison to the results obtained in this Thesis for the surfactants possessing shorter chains.

7.3.2 Improving Head Group Modeling

Assessing Head Group Behavior Using Simulations

The underprediction of cmc's observed in modeling the sulfobetaines (see Table 5.12 in Chapter 5) was rationalized in the text as possibly being due to head group interactions between the positively-charged ammonium and negatively-charged sulfonate groups on adjacent surfactants (illustrated using an equilibrium snapshot of a sulfobe-

taine micelle in Figure 5-5 of Chapter 5). Recognizing that the main contributions of a charged surfactant head group involve only the electrostatic and steric contributions to the free energy of micellization, such interactions between head groups must be considered in the context of the models of these contributions. Specifically, interactions between charges, such that oppositely-signed charges lie along the same surface of charge, will reduce the charging free-energy penalty by decreasing the charge density of that surface (recall that charges are not treated as points, but rather spread across the surface, such that the presence of oppositely-signed charges will offset each other). However, this increased head group interaction also increases the effective head area for the surfactants. From consideration of the sulfobetaines, it is clear that improved molecular descriptors are needed, for which additional investigation of the simulation results will be necessary. This particular example clearly illustrates that simple geometric arguments will not always suffice.

Another area for additional investigation involves the sugar-based surfactants. According to Bogusz and coworkers, the octyl glucosides exhibit some head group clustering,[96] presumably due to hydrogen-bonding effects. The MT model does not yet account for this type of interaction between surfactant heads, which would exhibit a free-energy benefit to association that would offset the steric interactions and the entropic losses that are also associated with this behavior.

Ethoxylate Head Packing

In Section 5.4.1 of Chapter 5, I showed that, as the ethoxylate chain length of the alkyl ethoxylate surfactants increases to 10, there is an increased positive deviation in the theoretical estimates of the modified free energy of micellization, g_m. This, in turn, leads to cmc predictions that are increasingly larger than the experimental cmc's. This type of deviation indicates that too high of a penalty is associated with some of the free-energy contribution to g_m. Since alkyl ethoxylates are nonionic, and the cmc overprediction is consistently observed by fixing tail lengths and increasing the ethoxylate chain length, it is clear from examination of the functional form of g_m (see Section 2.3 of Chapter 2) that the term responsible for the overprediction is the

steric free energy. Specifically, the problem arises from the specific scaling law used to evaluate the head area of ethoxylates of short length (centered on E_6), when it is implemented for longer ethoxylate chains.

As ethoxylate chain lengths increase, they behave more like polymeric materials, and the corona region can be visualized as being similar to a grafted polymer brush.[126] In this case, a hard-disk model for the steric free energy is inappropriate, because interpenetration between chains reduces the actual steric effect.

To address this situation properly for alkyl ethoxylates with long ethoxylate chains, I propose to apply the concepts of single-chain, mean-field packing (see Chapter 4) to packing in the head region. Figure 7-1 addresses the free-energy terms that would replace the steric free energy, with a phenomenological thought process which involves starting with the ethoxylate head groups after they have been separated from the tail groups in the original MT approach. A concentration step is performed, resulting in a Gibbs free energy, $-g_{dilution}$, which can be estimated from experimental determination of dilution enthalpies.[126] (Note that the solution should be concentrated to reach an average volume fraction of water equal to that encountered in the polymer brush region, which is an output of the packing calculation that I describe next.) At this point, the ethoxylate head is moved to the micelle core-water interface, where a Monte Carlo packing approach similar to that described in Appendix B can be applied, involving confinement and constraint calculations. The primary difference here, relative to surfactant tail packing, is that water is present, introducing a need for consideration of the change in the free energy of mixing between the water and ethoxylate chains when the ethoxylate chains are dispersed versus when they are tethered. In the dispersed case, the volume fraction of both the ethoxylate and the water is homogeneous. When tethered, the brush-like nature of the ethoxylate exhibits a high volume fraction of ethoxylate near the micelle core-water interface, and a low volume fraction of ethoxylate out towards the maximum-extended length of the chain. The spatial variation in volume fraction obtained from packing is a function of: (i) the surface coverage of the ethoxylate groups, (ii) the pressure field used to maintain constant density, and (iii) the specific mixing thermodynamics model used.[149] As in

Chapter 4, these effects are embodied in a probability distribution function which governs the behavior of the central chain.

An approach to model the mixture thermodynamics of this inhomogeneous system is to assume that, locally, a homogeneous theory for the free energy of mixing, such as the Flory-Huggins theory,[77,285] can be used. In this case, space is divided into concentric layers, a representative volume fraction is obtained for each layer, and a local contribution to the free energy of mixing is calculated, with the magnitude of the enthalpic interactions characterized using either a χ parameter[77] or a pairwise regular solution theory interaction parameter, A.[139] Calculation of the full free energy of mixing is then performed by integrating across these thin layers. A more rigorous approach would be to use a mixing free energy formalism that contains a gradient term, such as the Flory-Huggins-de Gennes approach.[286–288] Such an approach accommodates a spatially-varying volume fraction more directly.

As a final comment on the approach depicted in Figure 7-1, one must be careful in assessing the confinement free energy when performing packing calculations for both the surfactant tail and the head. Since these two regions are, in fact, connected, including an $f_{confine}$ term calculated independently will overpredict the effect of the confinement. What should be done instead is to carry out a confinement calculation for the entire surfactant molecule, evaluating its allowed configurations as a whole. In packing a $C_{12}E_6$ ethoxylate, for example, packing the C_{12} tail separately from the E_6 head, and then adding the results together, yields an estimate for the net confinement penalty which is very close to $\ln(2)\ k_BT$ higher than that obtained for $C_{12}E_6$ as an entire molecule, which is the result expected when the number of rejected configurations has been double-counted.

7.3.3 Improving Tail Group Modeling

Surfactant-Tailored Solubility Model

The results in Chapters 5 and 6 have illustrated that not all models for the aqueous solubility of oils are appropriate for modeling the aqueous solubility of the surfactant

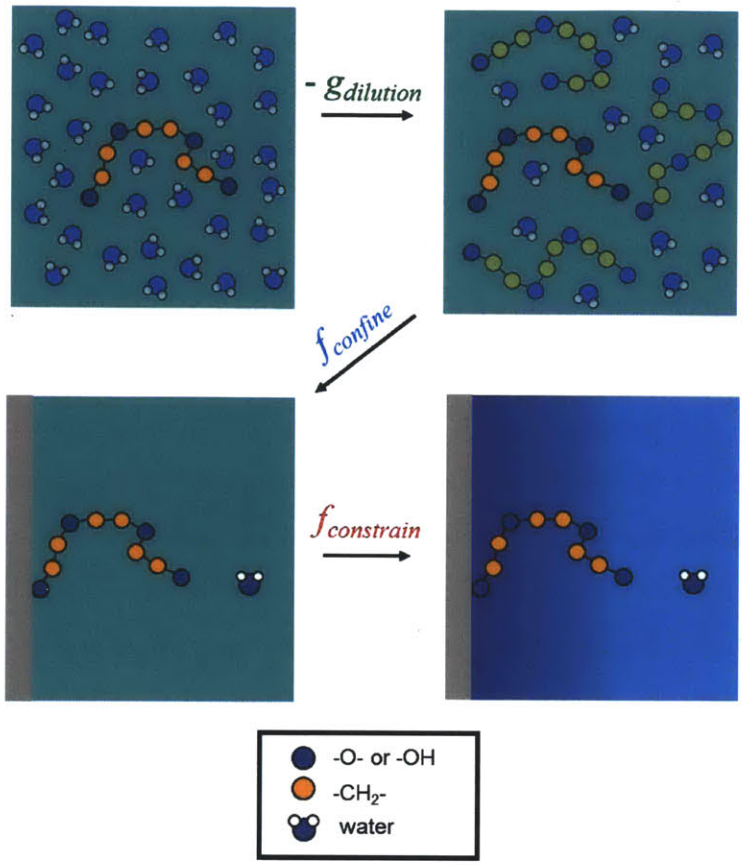


Figure 7-1: Schematic of the free-energy terms related to ethoxylate head packing in the corona region of an alkyl ethoxylate surfactant micelle. In the first step (upper left), the ethoxylate head group is present at infinite dilution. A concentration step at constant pressure and temperature (Gibbs free energy), represented as the reverse of a dilution step with associated experimentally-determined free energy $g_{dilution}$, results in a solution containing a volume fraction of water which is equal to the average value found in the micelle corona region (upper right). A central chain is then identified and pinned to the micelle core-water interface (depicted as a grey bar in the lower images), resulting in an entropic penalty due to confinement, $f_{confine}$ (a Helmholtz free energy) (lower left). The presence of all the other chains attached to the interface influences the central chain through a mean field (depicted at the lower right as a blue gradient background), obtained by enforcing a constant density constraint (assuming incompressibility) with solvent present as a mobile species. This results in a free-energy term $f_{constrain}$ (a Helmholtz free energy), which includes contributions from: (i) the pressure field, (ii) enthalpic interactions between the ethoxylate chain and the solvent, and (iii) the translational entropy of the solvent. Note that the constraint term is a function of the surface coverage, which in turn is determined based on the shape and size of the particular micelle under consideration. These three free-energy contributions are expected to replace the usual steric free energy, g_{st}.

tails. For linear surfactants, for example, the reason is due to the replacement of a terminal methyl in the oil case with a methylene group (to be attached to the head) in the surfactant tail case. Each solubility model presented in Section F.5 of Appendix F ascribes a different value for the transfer free energies of the methyl and methylene groups, such that all the models agree fairly well in predicting the solubility of reference oils, but exhibit differences of 1 k_BT or more in assessing the transfer free energy of a given surfactant chain. Although I was able to identify one particular solubility model as having a widespread applicability (the MHA solubility model), this situation illustrates that, perhaps, a solubility model tailored for surfactant modeling would be an effective alternative. In this case, one would select a set of surfactants with which to fit new parameter values for CH_3, CH_2, and other carbon-based groups to accurately reproduce cmc's, followed by analysis of a set of surfactants which are not included in the fitting. This type of approach may be necessary for evaluating proper parameters for groups such as a tetrafunctional C: a brief calculation of a transfer free energy for Triton X-100 (an alkyl phenol ethoxylate surfactant oxide possessing a 4-(1,1,3,3-tetramethylbutyl)-phenyl tail) appears not to be sufficiently negative to reproduce its experimental cmc using the MHA solubility model, perhaps due to the large positive value for C which I obtained through fitting to aqueous solubility data for a limited set of highly-branched oils.

Neutral Atom Approach

In Figure 7-2, I present an interesting surfactant structure for consideration: a di-alkyl dimethyl ammonium surfactant possessing a long primary chain and a shorter secondary chain emanating from the head group rather than from a branch node. This is truly a twin-tailed surfactant, since the branch point is in the head region. In practice, this type of surfactant exhibits a head/tail identification which includes some of the groups of the secondary chain in the tail. However, since the side chain is quite flexible, there are many configurations that will fail to satisfy this head/tail identification, resulting in a packing free energy which can be quite large as a result of these rejected conformations. However, in examining simulations of surfactants such

as these, it becomes apparent that the terminal groups of the secondary chain which are assigned to tail also spend some time extended out into the aqueous solution (disallowed in the packing model). As the side chain is lengthened, eventually there is enough hydrophobic material to fully integrate the second tail into the core, but, until that point, the packing model and computer simulations are not in agreement on the behavior of the surfactant chain, a result of the idealization of the theoretical micelle to include a sharp micelle core-water interface and a strict head/tail identification.

I have spent some time developing a packing framework to support a more flexible approach to packing. I refer to those groups which spend part of the time in each environment (head and tail) as having a "neutral" character. (That is, they are neither fully head nor fully tail, and are effectively neutral towards this type of assignation.) My idea for future work is to calculate tail properties as a function of conformation: if a particular arrangement of the surfactant results in the neutral groups depicted in Figure 7-2 entering the core region, then they are considered to contribute their full transfer free energy and volume to the properties of the tail. When these groups leave the core region, these properties are subtracted from the tail. According to the techniques discussed in Chapter 4, one can develop a probability distribution function by minimizing the total system free energy subject to constant density constraints. Originally, the free energy minimized was the packing free energy. In this new proposed framework, the free energy to be minimized should be the total modified free energy of micellization, since each free-energy contribution to this term has a dependence on either the chain conformation or the tail volume (through terms such as the area per surfactant molecule). However, rather than calculate each free energy term during the packing process, I have identified a technique that can be used to group together terms which are not expected to have a large dependence on tail conformation (such as the electrostatic charging contribution). The result is that a parameter can be introduced to represent these terms and their collective value, which is unknown at the time of preparation of the packing results. The key is then to develop the packing polynomials as done previously, but with an updated probability distribution function reflecting the neutral atom packing approach, and the addition of a dimension

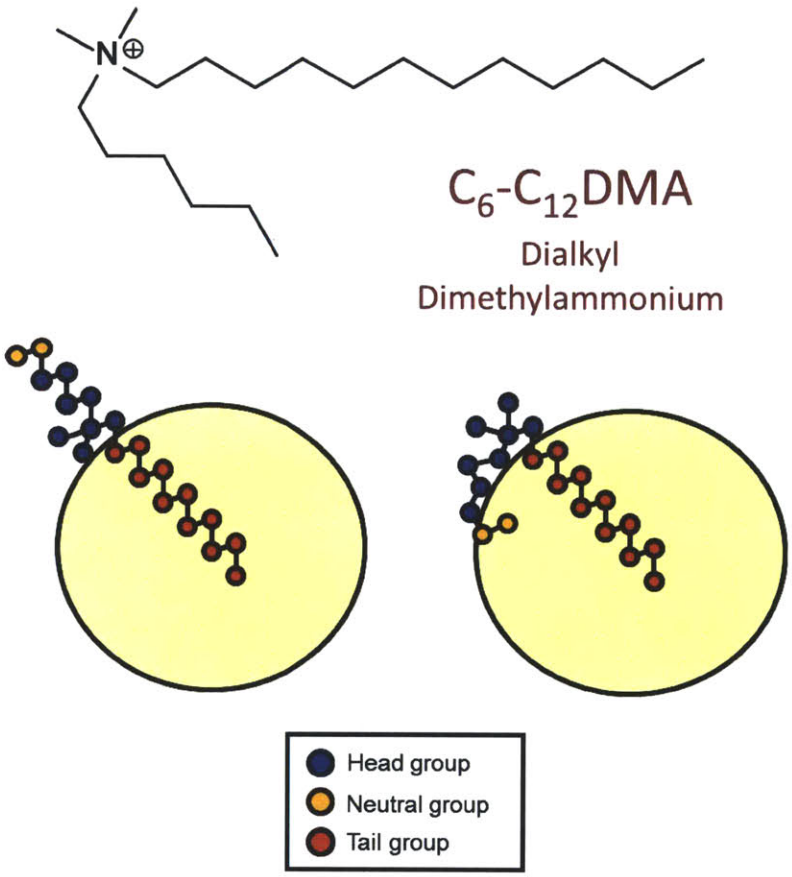


Figure 7-2: Conceptualization of the neutral atom packing approach. Pictured at top is a dialkyl dimethylammonium surfactant with C_6 and C_{12} tails. The primary tail is integrated into the micelle core, but the secondary tail possesses atomic groups which spend part of the time in the aqueous environment and part of the time loosely integrated into the core. Enforcing a strict head/tail identification in packing these surfactants can lead to high packing free energies. This can be corrected through the development of a packing approach which enables groups to enter and leave the core according to a modified probability distribution function obtained by minimizing all the free-energy contributions to the free energy of micellization with the packing free energy.

for fitting based on this grouped parameter. Then, in conducting MT modeling, at each iteration of the optimizer, this parameter can be directly calculated, and the estimate for the packing free energy read from the polynomial.

Other systems where this approach may prove valuable include those where the head/tail identification divides an aromatic ring. In such cases, the tail is hindered in its rotational freedom due to the requirement that some of the aromatic groups remain in the head region exclusively. The neutral atom approach would reduce the packing penalty in these cases, and bring the theoretical modeling more in line with the simulations, which indicate that groups near the interface are not limited from traversing the interface.

7.3.4 Modeling of Surfactant-Assisted Solubilization Phenomena

In aqueous systems containing surfactants and sparingly-soluble organic solutes, the presence of the solutes typically encourages the formation of surfactant micelles at concentrations which are smaller than the cmc of the surfactant in the absence of the organic solutes. This is due to the free-energetic benefit derived from encapsulating all, or part of, the organic solutes within the hydrophobic core of the surfactant micelles. This process, referred to as solubilization, with the solute referred to as a solubilizate, is driven by the very same hydrophobic effect that drives the micellization process in the absence of such solubilizates.[49] An often positive consequence of solubilization is that the total amount of solubilizate which can be added to the solution prior to precipitation is increased, because once the solubility limit of the solute has been reached in the bulk solution, a substantial additional amount of the solubilizate is present in encapsulated form inside the micelle cores.[289] Two quantities are typically used to quantify the extent of solubilization: (i) the maximum additive concentration (MAC), which is the total solute concentration that can be added before the solubility limit is reached (note that the MAC is greater than the solubility limit for the reason just described), and (ii) a mole-fractional distribution coefficient

(K), which is the ratio between the mole fraction of solute in the micelle and the mole fraction of solute in the bulk region of the solution.[290]

For solubilizates which are amphiphilic (possessing a head and a tail, and whose solubility limit is below a possible cmc), one can treat the system as a binary surfactant system, with the caveat that the conceptual surfactant corresponding to the actual solubilizate exhibits an MAC. This approach was utilized by Stephenson and coworkers to study the solubilization of ibuprofen in the anionic surfactant sodium dodecyl sulfate (SDS), the cationic surfactant dodecyl trimethylammonium bromide (CTAB), and the nonionic surfactant octaethylene glycol dodecyl ether ($C_{12}E_8$).[291] The MAC itself can be determined theoretically as a function of the surfactant concentration by increasing the total solubilizate solution concentration until the predicted monomer concentration reaches the solubility limit. The total solubilizate solution concentration obtained in this manner is the MAC. (Note that the solubility limit can also be more directly integrated into the equations, such that the MAC is an automatic output of the theory, as described by Srinivasan.[119]) Since the MT theory predicts the mole fractions of all species in both bulk and micelle environments, calculation of K is also straightforward using the theory.

For solubilizates which are hydrophobic (possessing no head), the situation is more complicated, because the hydrophobic solubilizates are mobile in the micelle core (lacking a head group to anchor the molecules at the micelle core-water interface). This has several interesting consequences that must be considered when conducting packing calculations of the type described in Chapter 4. First, the presence of mobile entities requires that a radially-dependent entropy of mixing term to be added to the theory (where radius refers to the usual radius for a sphere, to the cross-sectional radius for a cylinder, and to the distance from the mid-plane of a bilayer along the normal direction). Second, the maximum size of the micelle core is no longer limited by the maximum chain length of the surfactant tail, but rather by the amount of solubilizate present in the core, because these mobile solubilizates can fill any void at the center of the micelle which the tails cannot reach at larger radii, an effect referred to as swelling.[119] Finally, the presence of solubilizates that

exhibit an interfacial tension which is different than that of the surfactant tail can exhibit a preference, or an antipathy, for the surface. This is determined based on whether the solubilizate interfacial tension is lower, or higher, than the tail interfacial tension, respectively.[120] For example, in comparing benzene and a dodecyl tail, the benzene/water interfacial tension (34.1 dyn/cm at $25°C$)[292] is less than the dodecane/water interfacial tension (52.55 dyn/cm at $25°C$),[9] where the interfacial tension of the dodecyl tail is assumed to be the same as that of dodecane, as discussed in Section F.6 of Appendix F. Assuming that the interfacial tension at the interface can be computed based on a volume-average weighting in a thin layer near the interface, the reduction resulting from the presence of benzene near the interface will result in an excess (i.e., enhancement) of benzene in the interfacial region relative to that expected if there were no effect on the interfacial tension (e.g., if the solubilizate had instead been dodecane).[119]

The actual location of benzene (or of any other solubilizate) within a micelle core is referred to as the locus of solubilization.[120] Keeping in mind the procedure to determine a packing probability distribution function, discussed in Chapter 4, one can modify the packing framework to minimize not only the packing free energy, but also the interfacial free energy, along with the mixing free energy, simultaneously. The result is a packing framework which has as outputs: (i) the locus of solubilization, as a function of micelle shape, size, and composition, (ii) the packing free energy, (iii) the interfacial free energy, and (iv) the mixing free energy. Similar to the packing polynomials presented in Appendix F, polynomials should be generated for the interfacial free energy and the mixing free energy, both of which would now be directly coupled to the packing free energy. In fact, a single polynomial could be generated for the sum of the three free energies, although it is still instructive to consider the individual contributions of each term to the modified free energy of micellization in order to interpret the results obtained.

An important application to which the above considerations are relevant involves solubilizates of environmental relevance. These are the polycyclic aromatic hydrocarbons (PAH's), which are known to be toxic when present even in exceedingly low

molar concentrations. One avenue to pollution remediation involves treatment of contaminated soils with aqueous solutions containing surfactants which can solubilize these PAH's effectively.[293]

Starting from the computational framework that I developed in this Thesis to enable prediction of branched surfactant micellization behavior, it would be valuable to extend the solubilization work of Srinivasan to the case of branched surfactant systems. This would provide new strategies to tune surfactant structure – via branched tail selection – in order to maximize the uptake of solubilizates such as the PAH's into the micelle core.

7.4 Concluding Remarks

In this Thesis, my goal has been to provide a thorough development and exposition of a molecular-thermodynamic modeling framework for predicting surfactant micellization behavior in dilute aqueous solution. My intention throughout these pages has been to consistently call attention to the precise nature of the dependence of micellization properties on the surfactant chemical structure. A full understanding of this molecular aspect of the theory will be necessary in order to proceed with investigations of: (i) novel surfactants, (ii) mixtures of surfactants, (iii) solutions containing organic solubilizates, and even (iv) systems involving surface phenomena, since equilibrium surfactant behavior at surfaces is intimately related to, and at times limited by, the thermodynamics of bulk solution micellization behavior. By focusing on providing a systematic treatment of linear and branched surfactant micellization, I hope to have illustrated how increasingly-complex surfactants can be accommodated and modeled using this approach, opening the door to a more computationally-assisted, design-based philosophy for surfactant product formulation in the future. My investigation of molecular dynamics simulations of surfactant monomers and micelles, and the role that these simulations can play in a combined computer-simulation/molecular-thermodynamic approach will hopefully inspire additional efforts in this promising area. Additionally, I hope that my discussion of single-chain, mean-field theories has

communicated an excitement which will inspire others to continue to develop and improve these powerful methods for quantifying the effects of confinement and constraint in small systems. Finally, I believe that the knowledge gained from theoretical and simulations-based modeling is invaluable in understanding the fundamentals of self-assembly behavior. I anticipate with great interest the future developments in this challenging field of surfactant science.

Appendix A

Vector Rotations

A.1 Overview

In this thesis, vector rotations play a key role in two contexts: (i) chain packing, and (ii) surfactant and micelle creation in Program BUILDER. The goal of this Appendix is to clearly present the equations and relationships used in processing these rotations, as they are implemented in practice.

In the first context, conformations of a chain of atomic groups are most conveniently specified using internal coordinates (i.e., bond lengths, angles, and dihedrals), since potentials are typically formulated in terms of such coordinates and application of a chain torsional model, such as the Rotational Isomeric State approximation, involves specification of discrete values for the dihedral angles.[213] Accurate calculation of chain statistics within some external environment (e.g., a micelle core) requires not only a suitable representative set of conformations for a chain, but also a sampling of rigid body rotations for each specific conformation. These rigid body rotations require knowledge of the Cartesian coordinates of the chain within the coordinate system of the external environment. For each particular conformation, then, this implies that one must: (a) convert from the specified internal coordinates to Cartesian coordinates in some convenient molecule-level frame-of-reference, and then (b) sample the phase space of orientations of the molecule frame-of-reference relative to the frame-of-reference for the external environment. Task (a) can be accomplished using

a two-step placement algorithm of the sort outlined at the end of this Appendix, and task (b) can be accomplished using uniform sampling of Euler angles, as described in the next section.

In the second context, creation of surfactants and micelles in Program BUILDER requires specification of intramolecule connectivity (i.e., the molecule "topology") and an initial set of Cartesian coordinates for each atom in the simulation cell frame-of-reference. From these, a molecular dynamics simulation using a software package such as GROMACS may be conducted. Similar to the case of chain packing, development of the Cartesian coordinates of a particular surfactant molecule requires conversion from its internal coordinates; therefore, the same two-step placement algorithm can be used. Following this, a suitable simulations algorithm such as Langevin Dynamics (see Appendix C) can be used to generate a library of surfactant conformations. In practice, I store these conformations as "frames", or ordered lists of Cartesian coordinates, with one frame per conformation. Given a library of frames for a particular surfactant, a micelle can be created by first extracting at random a number of frames corresponding to the micelle aggregation number, then translating and rotating each conformation such that the surfactant head-tail axis is oriented towards the center-of-mass of the micelle (with the head oriented outward), and the individual surfactants are evenly dispersed on a surface corresponding to the intended micelle core-water interface boundary. Finally, further action is typically necessary to eliminate intermolecular atomic group overlaps (which otherwise lead to numerical instabilities based on the large interatomic forces generated) using molecule translations ("expanding" the micelle radially and laterally, depending on the geometry). The large structural gaps that result are then removed through a "compression" step, where a short molecular dynamics run combined with external forces are used to bring the surfactants back together in a controlled manner.

In the two contexts mentioned above, 3×3 rotation matrices are used to accomplish all vector rotations. The method for generation of these matrices, however, depends on the application. Here, I discuss the two techniques that I have used: Euler angles and quaternions. In practice, I have found that the Euler angle approach is useful in

packing short chains, where the set of chain conformations is generated in full prior to solution of the packing equations. In this case, it is quite straightforward to sample the Euler angles uniformly, thereby uniformly sampling the phase space of chain orientations. The quaternion approach, which allows specification of an arbitrary rotation axis and angle, is particularly useful in Monte Carlo chain packing from the standpoint of proposing orientational moves which satisfy detailed balance.[294] It is also useful for the purpose of aligning two vectors: for example, aligning the surfactant head-tail axis with the radial vector of a micelle. This involves using the dot product to determine the angle of rotation, and the cross product to determine the axis of rotation, between the source and destination vectors. Applying the rotation matrix generated by the corresponding quaternion to all atoms in the surfactant then yields the desired orientation.

A.2 Rotation Matrices from Euler Angles

In Fig. A-1, a vector to be rotated about the z-axis is first projected onto the x-y plane, where it has an initial offset of θ radians. It is then rotated in the direction of the y-axis (positive rotation) to a new offset of $\theta + \Delta\theta$ radians. If $r_x = r_p \cos\theta$ and $r_y = r_p \sin\theta$ are the initial x and y components of vector $\underline{r}$, where r_p is the length of $\underline{r}$ projected onto the x-y plane, the new components after rotation through $\Delta\theta$ radians are related to the old ones through the following relationships:

$$\begin{aligned} r'_x &= r_p \cos(\theta + \Delta\theta) \\ &= r_p [\cos\theta \cos\Delta\theta - \sin\theta \sin\Delta\theta] \\ &= r_x \cos\Delta\theta - r_y \sin\Delta\theta, \end{aligned} \tag{A.2.1}$$

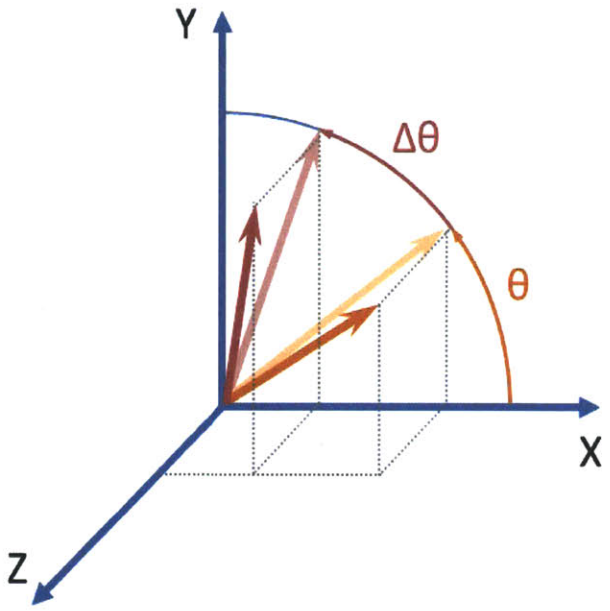


Figure A-1: Illustration of vector rotation about the z-axis from an initial position (denoted by the dark orange arrow) to a final position (denoted by the dark red arrow) through an angle $\Delta\theta$. (Vector projections onto the x-y plane are shown in lighter colors.) Here, the initial position is already offset θ radians from the x-axis.

and

$$\begin{aligned} r'_y &= r_p \sin(\theta + \Delta\theta) \\ &= r_p [\sin\theta \cos\Delta\theta + \cos\theta \sin\Delta\theta] \\ &= r_x \sin\Delta\theta + r_y \cos\Delta\theta, \end{aligned} \tag{A.2.2}$$

where the prime is used to denote the components of the vector after rotation, $\underline{r}'$.

Since the z-component is unaffected by this specific rotation, we have

$$r'_z = r_z. \tag{A.2.3}$$

Expressing the above transformation as $\underline{r}' = \mathbf{A}_z(\Delta\theta)\,\underline{r}$, where $\mathbf{A}_z$ is a matrix

representing rotation about the z-axis (a function of $\Delta\theta$), we can see from above that

$$\mathbf{A}_z(\Delta\theta) = \begin{bmatrix} \cos\Delta\theta & -\sin\Delta\theta & 0 \\ \sin\Delta\theta & \cos\Delta\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{A.2.4}$$

Rotations about the other two axes can be similarly developed as follows:[295]

$$\mathbf{A}_x(\Delta\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\Delta\theta & -\sin\Delta\theta \\ 0 & \sin\Delta\theta & \cos\Delta\theta \end{bmatrix}, \tag{A.2.5}$$

and

$$\mathbf{A}_y(\Delta\theta) = \begin{bmatrix} \cos\Delta\theta & 0 & \sin\Delta\theta \\ 0 & 1 & 0 \\ -\sin\Delta\theta & 0 & \cos\Delta\theta \end{bmatrix}. \tag{A.2.6}$$

In Fig. A-2, a typical application of these matrices to the problem of chain packing is illustrated. Here, we have already computed the Cartesian coordinates of the conformation of a linear chain (e.g., a linear alkane, where each CH_2 or CH_3 group is represented by a single point in space). We are interested in how this chain packs within a confinement (e.g., a sphere) when one end is tethered to the surface of the confinement, representing the real behavior of a surfactant tail within a micelle core, which is connected to a head group situated just outside the micelle core-water interface. Such packing calculations are detailed in Chapter 4, but, in short, we need to convert from the chain coordinate system to the micelle external coordinate system, and we additionally need to sample the space of all orientations of the chain robustly enough to generate consistent packing free energies for a particular chain length, architecture, and micelle geometry.[232]

My choice of molecule-level frame-of-reference when converting between the internal coordinates of a conformation and its Cartesian coordinates is such that the bond vector between the tethered atom group (a dummy group representing the first

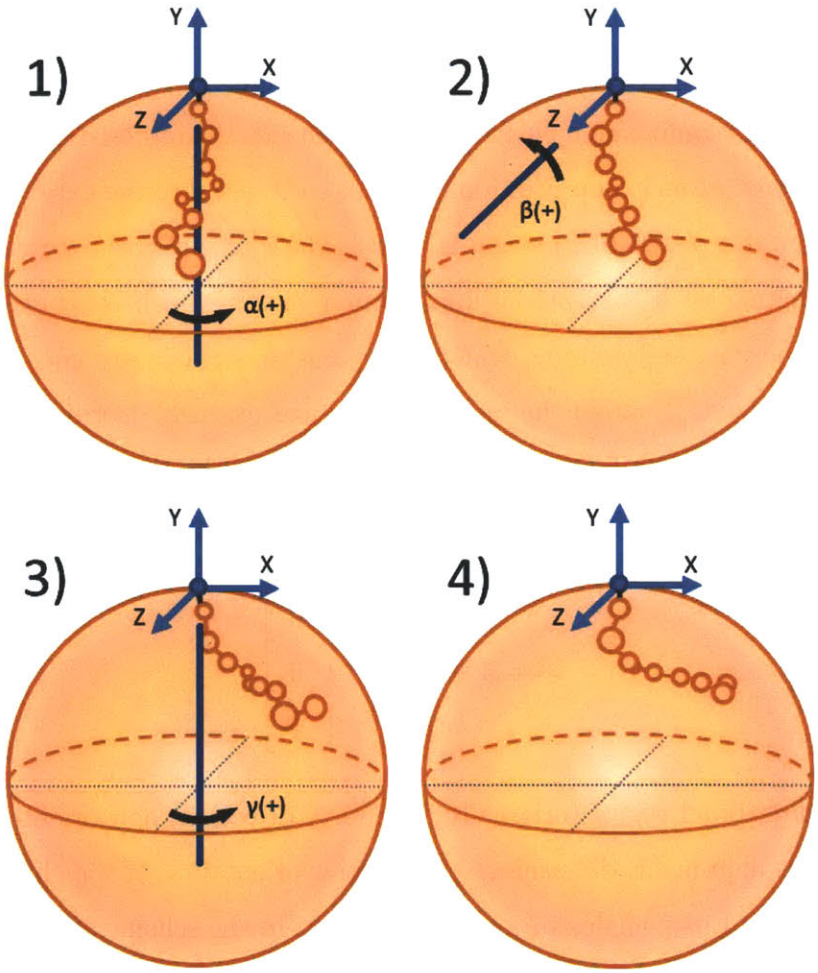


Figure A-2: Illustration of the effect of successive rotations of a chain of atom groups (circles connected by lines) through the three Euler angles α, β, and γ, where the chain is tethered at one end to a spherical surface representing an idealized micelle core-water interface. The coordinate system is positioned at the tethering point such that the y-axis is normal to the surface. The orientation of the x and z-axes is arbitrary, but kept consistent for all rotations. The direction of positive rotations is indicated with a black arrow, and positive angles have been selected for each rotation in this example. The rotation axis is emphasized with a purple bar. In 1), the initial chain orientation for a fixed internal chain conformation is shown. Rotation of the chain about the y-axis by an angle α yields the chain orientation in 2). Successive rotation about the z-axis by an angle β yields the chain orientation in 3). A final rotation about the y-axis by an angle γ yields the final chain orientation in 4). Note that the internal chain conformation is unaffected by these external rotations.

surfactant head atom) and the first atom directly connected to it lies along the x-axis direction, with the tethered atom group placed at the origin. (I will refer to this bond vector as the first bond vector of the chain). The origin of the external coordinate system of the micelle, for the purposes of single-chain, mean-field packing, is chosen to be located on the confinement surface directly beneath the tethered atom group (note that the tethered atom group position need not correspond to the external coordinate system origin, if a piston-like motion is used). The y-axis of this external coordinate system is aligned with the surface unit normal vector, which by convention points away from the enclosed volume, while the x-axis and z-axis are chosen arbitrarily, but then kept fixed. To align the molecule-level and external coordinate systems, we require that the first bond vector be reoriented to lie in the $-Y$ direction. Therefore, we first multiply all internal position vectors by a matrix which swaps axes,

$$\mathbf{A}_{swap} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{A.2.7}$$

leaving the first bond vector (originally in the $+X$ direction) now oriented as desired. Following this alignment, determination of a new orientation of the chain from specification of three Euler angles proceeds as follows. In the scheme shown in Fig. A-2, we first rotate the entire chain about the first bond vector (or, equivalently, rotate about the y-axis) through the angle α (Step 1). Note that this leaves the first bond vector unaffected. Next, we rotate the molecule about the z-axis through the angle β (Step 2), which serves the role of a polar angle for the first bond vector. Finally, we rotate again about the y-axis through the angle γ (Step 3), which serves the role of an azimuthal angle for the first bond vector. This y-z-y succession of rotations results in a final external conformation corresponding to the selected (α, β, γ) Euler angles (Step 4). It is clear that any relevant rigid body rotation is achievable in this scheme: the rotations through β and γ allow the first bond vector (when normalized) to sample all points on a unit sphere, and the first rotation through α covers all possible rotations of the molecule about the first bond vector axis.[296] (Note that

internal rotations about any other bond vector would generally result in the tethered atom leaving its anchored position.) At this stage, properties of the chain in the given conformation and orientation may be stored (e.g., the distribution of chain volume within the micelle core).

The rotations described above are represented mathematically by

$$\mathbf{A}_{rot}(\alpha,\beta,\gamma) = \mathbf{A}_y(\gamma)\,\mathbf{A}_z(\beta)\,\mathbf{A}_y(\alpha), \tag{A.2.8}$$

where the matrices on the right-hand side correspond to those presented in Eqs. A.2.4 and A.2.6, with the substitution of the specified Euler angle for $\Delta\theta$. Substituting, this matrix becomes

$$\begin{aligned}\mathbf{A}_{rot} &= \begin{bmatrix} \cos\gamma & 0 & \sin\gamma \\ 0 & 1 & 0 \\ -\sin\gamma & 0 & \cos\gamma \end{bmatrix}\begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & 1 & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{bmatrix} \\ &= \begin{bmatrix} \cos\gamma\cos\beta\cos\alpha - \sin\gamma\sin\alpha & -\cos\gamma\sin\beta & \cos\gamma\cos\beta\sin\alpha + \sin\gamma\cos\alpha \\ \sin\beta\cos\alpha & \cos\beta & \sin\beta\sin\alpha \\ -\sin\gamma\cos\beta\cos\alpha - \cos\gamma\sin\alpha & \sin\gamma\sin\beta & -\sin\gamma\cos\beta\sin\alpha + \cos\gamma\cos\alpha \end{bmatrix}.\end{aligned} \tag{A.2.9}$$

Our final procedure for generating a chain molecule orientation can now be summarized as follows:

$$\mathbf{A}_{ext} = \mathbf{A}_{rot}\mathbf{A}_{swap}, \tag{A.2.10}$$

where

$$\mathbf{A}_{ext} = \begin{bmatrix} \cos\gamma\sin\beta & \cos\gamma\cos\beta\cos\alpha - \sin\gamma\sin\alpha & \cos\gamma\cos\beta\sin\alpha + \sin\gamma\cos\alpha \\ -\cos\beta & \sin\beta\cos\alpha & \sin\beta\sin\alpha \\ -\sin\gamma\sin\beta & -\sin\gamma\cos\beta\cos\alpha - \cos\gamma\sin\alpha & -\sin\gamma\cos\beta\sin\alpha + \cos\gamma\cos\alpha \end{bmatrix}. \tag{A.2.11}$$

In chain packing calculations, we want to sample the phase space of orientations uniformly (i.e., without bias, since the probability of a given configuration will be modulated instead by the mean-field and conformation internal energy). This requires

Table A.1: Multiplication Table of Quaternion Basis Elements

	1	**i**	**j**	**k**
1	1	i	j	k
i	i	-1	k	-j
j	j	-k	-1	i
k	k	j	-i	-1

knowledge of the differential phase volume; that is, how the phase volume changes with the Euler angle coordinates. As mentioned above, the first bond vector samples the surface of a sphere, which has a differential surface area proportional to $\sin\beta d\beta d\gamma$, or, since $\sin\beta d\beta = -d\cos\beta$, proportional to $d(\cos\beta)\,d\gamma$. The rotation about the first bond vector by α provides a further differential contribution $d\alpha$, such that the phase volume is proportional to $d\alpha d(\cos\beta)\,d\gamma$. Uniform random sampling of α and γ from 0 to 2π and $\cos\beta$ from -1 to 1 will consequently result in the desired uniform sampling of phase space. For a brief discussion of uniform pseudorandom number generators, see Appendix C.

A.3 Rotation Matrices from Quaternions

Another useful method for describing rotations involves the quaternions, which are numbers possessing four basis elements, 1, i, j, and k, which obey $i^2 = j^2 = k^2 = ijk = -1$, from which the multiplication table in Table A.1 can be computed.[297] Note the anticommutativity of i, j, and k with each other.

A quaternion q can be represented by four real numbers (q_0, q_1, q_2, q_3), such that

$$q = q_0 + q_1 i + q_2 j + q_3 k, \tag{A.3.1}$$

where the basis element 1 is suppressed in writing. The norm of a quaternion is computed as follows:

$$|q| = q_0^2 + q_1^2 + q_2^2 + q_3^2. \tag{A.3.2}$$

Given a normalized rotation axis $\hat{n} = (n_x, n_y, n_z)$ and angle θ (given by the right-hand rule: if the eye looks in the direction of $\hat{n}$, a positive rotation is clockwise), the

quaternion that describes this rotation is given by[297]

$$q = \cos\frac{\theta}{2} + \sin\frac{\theta}{2}\left[n_x i + n_y j + n_z k\right], \tag{A.3.3}$$

where it is clear that $|q| = 1$ from application of Eq. A.3.2, as long as the rotation axis is normalized.

Given a second quaternion $p = (p_0, p_1, p_2, p_3)$, quaternion multiplication $q * p$ describes a rotation encoded in p followed by a rotation encoded in q. Writing p and q in the form of Eq. A.3.1 and utilizing the multiplication table in Table A.1 yields

$$\begin{aligned} q * p &= (q_0 + q_1 i + q_2 j + q_3 k)(p_0 + p_1 i + p_2 j + p_3 k) \\ &= (q_0 p_0 - q_1 p_1 - q_2 p_2 - q_3 p_3) + \\ &\quad + i(q_0 p_1 + q_1 p_0 + q_2 p_3 - q_3 p_2) + \\ &\quad + j(q_0 p_2 - q_1 p_3 + q_2 p_0 + q_3 p_1) + \\ &\quad + k(q_0 p_3 + q_1 p_2 - q_2 p_1 + q_3 p_0). \end{aligned} \tag{A.3.4}$$

Finally, a quaternion may be converted into a rotation matrix according to the following formula:[297]

$$\mathbf{A}_{rot}(q) = \begin{bmatrix} 1 - 2(q_2^2 + q_3^2) & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & 1 - 2(q_1^2 + q_3^2) & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & 1 - 2(q_1^2 + q_2^2) \end{bmatrix}. \tag{A.3.5}$$

A.4 Two-Step Internal Coordinate Generation

When developing an internal conformation of a molecule for chain packing, we start with a complete specification of internal coordinates (i.e., all bond lengths, angles, and dihedrals). We next specify a molecule-level Cartesian coordinate system, which is used to report all atom positions in vector component form. Assume that the first four atoms in the chain are denoted by A, B, C, and D, respectively, with positions $\underline{r}_A$, $\underline{r}_B$, $\underline{r}_C$, and $\underline{r}_D$, bond lengths l_{AB}, l_{BC}, l_{CD}, bond angles θ_{ABC}, θ_{BCD}, and bond dihedral ϕ_{ABCD}. One can specify a suitable coordinate system through selection of $\underline{r}_A =$

$(0,0,0)$, $\underline{r}_B = (l_{AB},0,0)$, and $\underline{r}_C = \underline{r}_B + (l_{BC}\cos\theta_{ABC}, l_{BC}\sin\theta_{ABC}, 0)$. Once these three positions have been specified, the position of atom D and all subsequent atoms in the chain can be computed following the two-step rotation procedure described below and illustrated in Fig. A-3 (where "two step" refers to the number of rotations required).[298]

We start by constructing the bond vectors

$$\underline{r}_{AB} = \underline{r}_B - \underline{r}_A \tag{A.4.1}$$

and

$$\underline{r}_{BC} = \underline{r}_C - \underline{r}_B. \tag{A.4.2}$$

Our goal is to calculate $\underline{r}_{CD}$ from l_{CD}, θ_{BCD}, ϕ_{ABCD}, and these two bond vectors. We begin by creating an initial bond vector $\underline{r}^0_{CD}$, which is set equal to the unit bond vector between atoms B and C as follows:

$$\underline{r}^0_{CD} = \widehat{\underline{r}}_{BC} = \frac{\underline{r}_{BC}}{|\underline{r}_{BC}|}. \tag{A.4.3}$$

We need the unit vector normal to the plane containing atoms A, B, and C to complete a rotation corresponding to θ_{BCD}. This unit normal is simply given by

$$\widehat{\underline{n}} = \frac{\underline{r}_{AB} \times \underline{r}_{BC}}{|\underline{r}_{AB} \times \underline{r}_{BC}|}. \tag{A.4.4}$$

Then, to satisfy θ_{BCD}, the first step of the two-step process is to generate a quaternion using Eq. A.3.3, the rotation axis in Eq. A.4.4, and the supplement of θ_{BCD}. Denoting Eq. A.3.3 as a function $Q\left(\widehat{\underline{n}}, \theta\right)$, we have

$$p = Q\left(\widehat{\underline{n}}, (\pi - \theta_{BCD})\right). \tag{A.4.5}$$

To satisfy ϕ_{ABCD}, the second step of the two-step process involves generating a second quaternion using $\widehat{\underline{r}}_{BC}$ as a rotation axis and the dihedral angle as the rotation

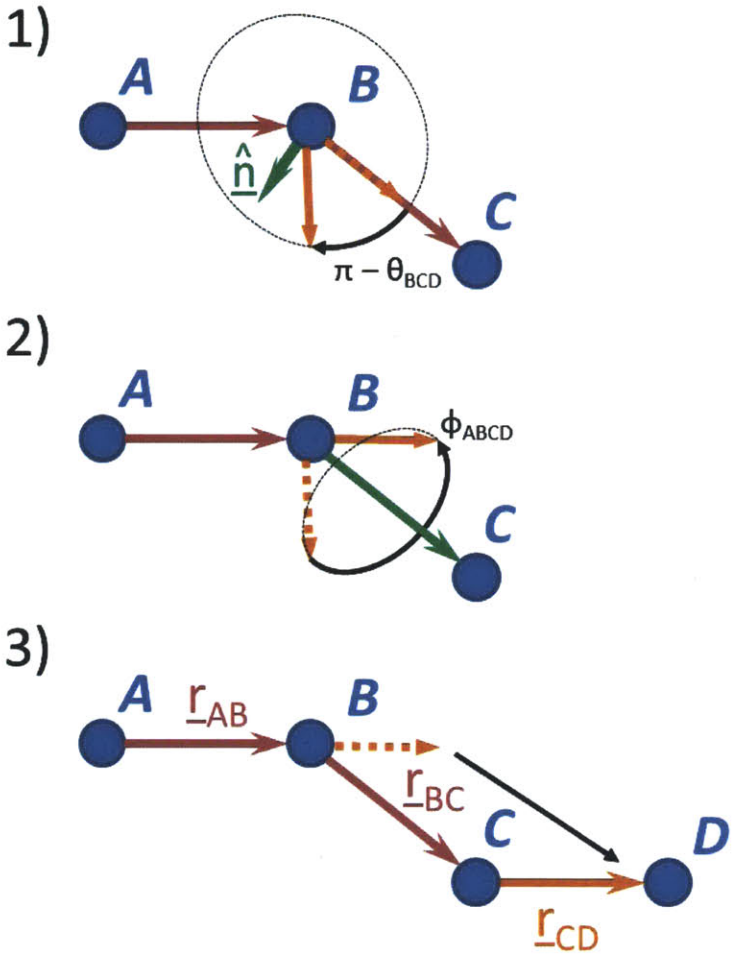


Figure A-3: Illustration of a two-step process (i.e., involving two rotations) for determining the position of atom D relative to its predecessors A, B, and C. The first rotation, illustrated in 1), is about the unit normal of the plane containing A, B, and C through an angle of $\pi - \theta_{BCD}$ radians. The second rotation, illustrated in 2), is about the $\underline{r}_{BC}$ bond vector (after normalization) through an angle of ϕ_{ABCD} radians. Following these two rotations, the resulting vector is scaled by l_{CD} and translated to C, as depicted in 3).

angle. Specifically,

$$q = Q\left(\hat{\underline{r}}_{BC}, \phi_{ABCD}\right). \tag{A.4.6}$$

Applied in order, these two rotations can be represented by the quaternion product $q * p$. It now remains to generate a rotation matrix for this product using Eq. A.3.5, apply it to $\underline{r}^0_{CD}$, and scale the result by l_{CD}, as follows:

$$\underline{r}_{CD} = l_{CD}\mathbf{A}_{rot}\left(q * p\right)\underline{r}^0_{CD}. \tag{A.4.7}$$

The position of atom D is then simply given by

$$\underline{r}_D = \underline{r}_C + \underline{r}_{CD}. \tag{A.4.8}$$

Appendix B

Monte Carlo Techniques for Chain Packing

B.1 Motivation

The use of Monte Carlo[299] techniques in this thesis is motivated by the need to conduct packing free-energy calculations of the type described in Chapter 4 for long or highly-branched chains. Here, I present a brief discussion of: (i) the problem of chain packing, (ii) the typical random sampling method used to solve the packing equations for short chains, and (iii) the limitations of that method, in order to motivate my use of more advanced Monte Carlo techniques.[203]

In Section 4.2.3 of Chapter 4, I presented the derivation of the chain probability distribution function (pdf) governing the conformational statistics of a surfactant tail packed in a micelle core under a single-chain mean field (SCMF) approximation.[82] In that system, conditions of constant temperature (T), constant total core volume ($\underline{V}_\xi$), and fixed total number of chains (n_ξ) are applied. (Note that I make use of the index ξ to indicate the specific micelle under consideration in a particular packing calculation – that is, I use ξ to index across the various possible micelle shapes, sizes, and compositions.)

The pdf result, which I obtained in Chapter 4 as Eq. 4.2.42, is equivalent to the pdf for a canonical ensemble[300] supplemented with a second Boltzmann factor due to

an additional imposed constraint of uniform density throughout the core, motivated by experimental observations of the fluidity of typical micelle cores.[34,82] The result is reproduced here for completeness:

$$p_i(\omega_i) = \frac{\exp(-\beta\epsilon_i(\omega_i))\exp\left(-\beta\sum_{l=1}^{L}\pi_l v_{i,l}(\omega_i)\right)}{Z_i}, \tag{B.1.1}$$

where ω_i is a particular chain conformation for species i relative to the coordinate system of the micelle core (i.e., ω_i is a vector representing a full specification of all atomic positions in the chain), p_i is the pdf for the chain, ϵ_i is the internal energy of the chain, $v_{i,l}$ is the volume of the chain in layer l of the micelle core (the volume of the core is divided into L discrete, concentric layers for modeling purposes), π_l is the Lagrange multiplier for the constant density constraint in layer l (interpreted as a lateral pressure[82]), $\beta = 1/k_BT$ is the thermodynamic beta (where k_B is the Boltzmann constant), and Z_i is the partition function, which ensures normalization of the chain pdf p_i. The terms p_i, ϵ_i, and $v_{i,l}$ are all functions of ω_i. The partition function Z_i is computed as follows:

$$Z_i = \sum_{\omega_i} \exp(-\beta\epsilon_i(\omega_i))\exp\left(-\beta\sum_{l=1}^{L}\pi_l v_{i,l}(\omega_i)\right), \tag{B.1.2}$$

where the summation is over all possible chain conformations and orientations within the micelle core.

Given a set of lateral pressure values $\{\pi_l\}$ which lead to satisfaction of the constant density constraints (see Eq. 4.2.28 in Chapter 4) and full knowledge of Z_i, one can exactly compute the expectation value for any function of chain conformation as follows:

$$\langle X(\omega_i)\rangle = \sum_{\omega_i} p_i(\omega_i) X(\omega_i), \tag{B.1.3}$$

where X is the desired function. For example, one may be interested in computing $\langle\epsilon_i(\omega_i)\rangle$, the average chain internal energy for species i in the core, or $\langle v_{i,l}(\omega_i)\rangle$, the average volume of species i found in layer l.

In practice, however, full knowledge of Z_i is neither available nor obtainable: the space of all possible chain conformations is prohibitively large. Instead, one must sample the space of conformations in some manner, and estimate $p_i(\omega_i)$ from this limited set of samples.

One manner of sampling is, of course, random sampling.[203] For example, the chain model that I employ is one in which bond lengths and bond angles are fixed, and bond dihedrals are described using the Rotational Isomeric State (RIS) theory[213] (see Section 4.3 of Chapter 4 for full details). In an approach which I will hereafter refer to as the "full-enumeration" approach, all possible internal conformations of a chain are determined through enumeration of the backbone dihedral RIS states. Within the RIS formalism, a typical treatment of an all-carbon dihedral in a saturated hydrocarbon (alkane) involves the identification of three local minimum-energy states which represent the dihedral angles most likely to be adopted by the chain. These three states replace the full, continuous torsional potential energy profile, discretizing internal rotations.[214,301] A linear or branched surfactant tail, possessing n_c total carbon atoms and n_m methyl groups, has $n_c - n_m - 1$ rotatable bonds of importance. (Note that bonds connecting to terminal methyl groups, while rotatable, exhibit three-fold symmetry with respect to energy levels, such that their rotation need not be considered;[216] the connection with the surfactant head is also assumed symmetric.) For a three-state RIS model, the number of possible conformations is thus $N_{conf} = 3^{n_c - n_m - 1}$. For example, a linear C12 tail has $n_c = 12$, $n_m = 1$, and $N_{conf} = 59049$. The sampling procedure is to generate each of these internal conformations once in turn (no preferential weighting is given); then, given a particular internal conformation, one samples some of the possible external orientations for the chain in the coordinate system of the micelle core (with the additional constraint that one end of the chain be held fixed at the micelle core boundary, where an attachment to the surfactant head in the real physical process of micelle core packing exists). This orientational sampling involves uniform random sampling of the three Euler angles according to the prescription given in Section A.2 of Appendix A. A typical number of external samples which produces sufficient accuracy is 500 samples per internal conformation

(see Chapter 4 for a comparison of various levels of sampling). In this manner, the state-space of chain conformations is uniformly sampled, and p_{ω_i} can be approximated by computing Z_i as in Eq. B.1.2, but with the summation running over the sample set.

Note that the set of internal conformations and external orientations can be stored in random-access memory (RAM) during non-linear solution of the density constraint equations for $\{\pi_l\}$. This has two advantages: (i) avoidance of any requirement to regenerate the sample set at each step of the non-linear solver, and (ii) allowance of the use of a numerical Jacobian by the non-linear solver. (My non-linear solver of choice for this purpose is the HYBRD solver in the MINPACK package.[156])

The drawback of the full-enumeration approach is that uniform sampling of the internal conformations is wasteful: many high-energy configurations are sampled which contribute negligibly to Z_i in Eq. B.1.2 due to high internal energies ($\epsilon_i(\omega_i)$). Similarly, uniform sampling of external conformations can also be wasteful: orientations in which a large amount of chain volume is found in layers with large lateral pressures will also lead to very small Boltzmann factors in Eq. B.1.1. These drawbacks can ultimately prevent calculation due to resource limitations as n_c increases. On a single 2.4GHz processor, when examining a linear C16 tail ($n_c = 16$, $n_m = 1$, and $N_{conf} = 4782969$) for the three regular geometries of sphere, cylinder, and bilayer, with 15 core minor radii per shape, the full-enumeration approach requires roughly 20 hours of computational time and 1.5GB of RAM. Each additional rotatable bond introduces roughly a factor of three increase in the time, with a smaller factor increase in the RAM due to compression of the data. Beyond perhaps 16-18 rotatable bonds, the time and memory requirements are too severe for conventional computing.

An alternative approach is to sample the space of conformations non-uniformly. For this purpose, the Monte Carlo technique of importance sampling – in particular, application of the Metropolis method[302] – is a useful strategy. In the discussion which follows, I will briefly present the concept of a Markov chain (not to be confused with the surfactant chains being packed), its use in determining ensemble averages, and the role of the Metropolis method in equilibrating a Markov chain. In contrast to

the full-enumeration approach, where conformations are stored in memory in advance of non-linear solution of the packing constraints, each trial set of $\{\pi_l\}$ requires fresh sampling of the surfactant chains (i.e., development of a new Markov chain), since the lateral pressure-volume Boltzmann weighting in Eq. B.1.1 directly impacts the Metropolis acceptance criterion. This also leads to a practical requirement for an analytical Jacobian, since numerical Jacobians are typically generated via finite-difference methods, which would require at least one additional Monte Carlo run per Jacobian matrix element (in a forward-differencing scheme[156]). Nevertheless, the benefits of this approach are substantial for long chains: one finds convergence of expectation values within a much shorter computational time frame, and an elimination of the need for large RAM storage space.

B.2 Markov-Chain Monte Carlo

B.2.1 Estimators

One of the primary goals of Monte Carlo approaches is to efficiently estimate expectation values of the type found in Eq. B.1.3, which, when combined with Eqs. B.1.1 and B.1.2, can be written as follows:

$$\langle X_\mu \rangle = \frac{\sum_\mu X_\mu \exp\left(-\beta U_\mu\right)}{\sum_\mu \exp\left(-\beta U_\mu\right)}, \tag{B.2.1}$$

where I am making a symbolic replacement of the suffix μ for a functional dependence on ω_i, to enable a direct comparison to the discussion of importance sampling by Newman and Barkema.[300] In the context of packing, $U_\mu = \epsilon_\mu + \sum_{l=1}^{L} \pi_l v_{\mu,l}$.

Equation B.2.1 is exact. However, when sampling the space of states μ, only a discrete number of states can practically be sampled. This leads to an estimator for $\langle X_\mu \rangle$ of the form[300]

$$X_M = \frac{\sum_{i=1}^{M} \frac{X_{\mu_i}}{w_{\mu_i}} \exp\left(-\beta U_{\mu_i}\right)}{\sum_{i=1}^{M} \frac{1}{w_{\mu_i}} \exp\left(-\beta U_{\mu_i}\right)}, \tag{B.2.2}$$

where M is the total number of samples collected, i indexes the samples, and w_{μ_i}

is the weighting given to μ_i in the sampling procedure. During uniform, random sampling, all w_{μ_i} are the same, leading to the special case

$$X_M = \frac{\sum_{i=1}^{M} X_{\mu_i} \exp\left(-\beta U_{\mu_i}\right)}{\sum_{i=1}^{M} \exp\left(-\beta U_{\mu_i}\right)}, \tag{B.2.3}$$

which bears a clear similarity to Eq. B.2.1. This is the situation encountered in the full-enumeration approach to chain packing, and shares all the corresponding drawbacks.

As long as $w_{\mu_i} > 0$ for all μ_i when $X_{\mu_i} \exp\left(-\beta U_{\mu_i}\right)$ is non-zero,[303] X_M will approach $\langle X_\mu \rangle$ as M approaches infinity.[300] The origin of Eq. B.2.2 is perhaps more clearly seen in the integral form of Eq. B.2.1:

$$\langle X_\mu \rangle = \frac{\int X_\mu \exp\left(-\beta U_\mu\right) d\mu}{\int \exp\left(-\beta U_\mu\right) d\mu}. \tag{B.2.4}$$

Multiplying and dividing within each integral in Eq. B.2.4 by w_μ (a weighting given to μ such that $w_\mu > 0$ for all μ when $\exp\left(-\beta U_\mu\right)$ is non-zero) yields

$$\langle X_\mu \rangle = \frac{\int \frac{X_\mu \exp(-\beta U_\mu)}{w_\mu} w_\mu d\mu}{\int \frac{\exp(-\beta U_\mu)}{w_\mu} w_{\mu_i} d\mu}. \tag{B.2.5}$$

If we define a new variable u, such that $du = w_\mu d\mu$, and sample u uniformly, rather than μ, then the result is Eq. B.2.2.[303] (Note that μ becomes a function of u for the purpose of computing X_μ, U_μ, and w_μ.[203])

There is another useful special case for Eq. B.2.2. If the weightings w_{μ_i} are selected such that $w_{\mu_i} \propto \exp\left(-\beta U_{\mu_i}\right)$, then

$$X_M = \frac{1}{M} \sum_{i=1}^{M} X_{\mu_i}, \tag{B.2.6}$$

which is simply the mean of the values of X obtained for the samples in the sample set. Choosing samples such that Eq. B.2.6 holds can be readily accomplished through the use of Markov chains.[300]

B.2.2 Markov Chains

A Markov chain can be viewed as a sequence of system states obtained through a discrete-time stochastic process, wherein a state at time $t+1$ (say, state ν), is obtained only from knowledge of the state at current time t (say, state μ), based on a set of transition probabilities $P(\mu \rightarrow \nu)$, such that $P(\mu \rightarrow \nu) \geq 0$ and $\sum_{\nu} P(\mu \rightarrow \nu) = 1$ (i.e., there is always a next state).[300,303,304] Each Markov chain is initialized at $t = 0$ to some valid initial state according to some initial probability distribution. Given proper selection of $P(\mu \rightarrow \nu)$, there is a notion of a Markov chain coming to equilibrium over time (a dynamic process[305]), such that states will begin appearing in sequence with probabilities proportional to the Boltzmann factors corresponding to the energies of those states. This corresponds to the case of $w_{\mu_i} \propto \exp\left(-\beta U_{\mu_i}\right)$ mentioned in the previous section, which thereby allows use of Eq. B.2.6.[300]

Proper selection of $P(\mu \rightarrow \nu)$ involves several considerations, as discussed by Newman and Barkema.[300]

First, all $P(\mu \rightarrow \nu)$ should be time-invariant, if an equilibrium following the Boltzmann distribution is to be obtained, since the Boltzmann weightings $\exp\left(-\beta U_{\mu_i}\right)$ are not functions of the Markov chain time coordinate. History beyond the current state is also not involved.

Second, any state must be accessible from any other state in finite time (i.e., a condition of ergodicity), since each state with finite energy has a non-zero Boltzmann weighting. Note that this does not disallow $P(\mu \rightarrow \nu) = 0$, but indicates that some route between μ and ν, via intermediate states, must exist.

Third, to maintain equilibrium once it is reached, the condition of balance must be satisfied. This condition can be expressed as a requirement that the probability of transitioning into a particular state μ during a given time step from some state ν is equal to the probability of transitioning out of that state μ:

$$\sum_{\nu} p_{\mu} P(\mu \rightarrow \nu) = \sum_{\nu} p_{\nu} P(\nu \rightarrow \mu), \tag{B.2.7}$$

where p_{μ} is the equilibrium (infinite-time) probability distribution function (pdf)

value for state μ, and p_ν is the equilibrium pdf value for state ν. Satisfying this condition at equilibrium implies that the master equation[304]

$$\frac{dp_\mu}{dt} = -\sum\nolimits_\nu p_\mu P(\mu \to \nu) + \sum\nolimits_\nu p_\nu P(\nu \to \mu) \tag{B.2.8}$$

is equal to zero for all states μ. This ensures that, once at equilibrium, equilibrium is maintained.

Fourth, to guarantee that equilibrium can be reached, a stricter, detailed balance condition must be satisfied. This condition is written as follows:[305]

$$p_\mu P(\mu \to \nu) = p_\nu P(\nu \to \mu). \tag{B.2.9}$$

As discussed in more detail by Newman and Barkema, this condition equates the rate of transitions from μ to ν with that from ν to μ, which can be shown to guarantee that an equilibrium is reachable where each time step maintains a fixed pdf. Otherwise, the concept of limit cycles would be a possibility, where a particular pdf is obtained every n time steps instead (a form of dynamic equilibrium).[300]

Finally, to ensure that the equilibrium reached follows the Boltzmann distribution (i.e., describes the canonical ensemble), one can use the Metropolis algorithm.[302]

B.2.3 Metropolis Algorithm

Rearranging Eq. B.2.9, one obtains

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{p_\nu}{p_\mu}. \tag{B.2.10}$$

To match the canonical ensemble pdf, we require for both $i = \mu$ and $i = \nu$ that

$$p_i = \frac{\exp(-\beta U_i)}{\sum_i \exp(-\beta U_i)}. \tag{B.2.11}$$

Substituting Eq. B.2.11 into Eq. B.2.10, and recognizing that $\sum_\mu \exp(-\beta U_\mu) =$

$\sum_\nu \exp(-\beta U_\nu)$, since μ and ν are both indexing the same set of states, we find that

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \exp(-\beta(U_\nu - U_\mu)). \tag{B.2.12}$$

Next, it is typical to decompose the transition probabilities P into products of a selection probability g and an acceptance probability A.[300] Specifically,

$$P(\mu \to \nu) = g(\mu \to \nu) A(\mu \to \nu), \tag{B.2.13}$$

and

$$P(\nu \to \mu) = g(\nu \to \mu) A(\nu \to \mu). \tag{B.2.14}$$

If one chooses the selection probabilities to be symmetric, such that

$$g(\mu \to \nu) = g(\nu \to \mu), \tag{B.2.15}$$

then substitution of Eqs. B.2.13 and B.2.14 into Eq. B.2.12 yields

$$\frac{A(\mu \to \nu)}{A(\nu \to \mu)} = \exp(-\beta(U_\nu - U_\mu)). \tag{B.2.16}$$

It remains to determine a functional form for $A(\mu \to \nu)$ which satisfies this expression. The Metropolis algorithm involves the following particular selection:

$$A(\mu \to \nu) = \begin{cases} \exp(-\beta(U_\nu - U_\mu)), \text{ if } U_\nu > U_\mu \\ 1, \text{ if } U_\nu \leq U_\mu \end{cases}. \tag{B.2.17}$$

To verify that Eq. B.2.16 is satisfied, assume that $U_\nu \leq U_\mu$. Then, $A(\mu \to \nu) = 1$ and $A(\nu \to \mu) = \exp(-\beta(U_\mu - U_\nu))$, such that Eq. B.2.16 holds. Likewise, if $U_\nu > U_\mu$, then $A(\mu \to \nu) = \exp(-\beta(U_\nu - U_\mu))$ and $A(\nu \to \mu) = 1$, such that, again, Eq. B.2.16 holds. Although alternatives to this scheme exist, the Metropolis algorithm is typically among the most efficient.[203]

Implementation of Eq. B.2.17 requires a high-quality uniform random number

generator (see the discussion at the end of Section C.6 of Appendix C). One generates a uniform random number ξ on $(0,1)$. If $\xi < A(\mu \rightarrow \nu)$, the new state is accepted. Otherwise, the old state is maintained for an additional time step.[203]

B.2.4 Proposing Moves for Chain Packing

Given a current state μ, the set of states ν for which $g(\mu \rightarrow \nu)$ is non-zero represents the space of moves that the designer of a Monte Carlo algorithm chooses in order to optimize convergence of the Markov chain to equilibrium and X_M to $\langle X_\mu \rangle$. If the moves proposed are too drastic, such that $U_\nu \gg U_\mu$, then the proposed states will be repeatedly rejected, increasing the time to convergence. Within the chain-packing system there are two main considerations: (i) the internal conformation of a chain, and (ii) the external orientation of the chain. I have selected three basic moves in my work: (a) a change in RIS state for a single dihedral angle along the chain (a "point mutation" where the dihedral for change is randomly selected, and the new RIS state is also randomly selected), (b) a twist of the chain about a central axis, through some random angle $\Delta\theta < \Delta\theta_{MAX}$, and (c) a swing of the chain in a random direction, through some random angle $\Delta\theta' < \Delta\theta'_{MAX}$. All three move types satisfy the reversibility embodied in Eq. B.2.15. I propose one move per Markov chain time step, according to some pre-determined fraction assigned to each move type and random draw of a uniform variate from $(0,1)$ to select the move type. In general, the goal is a target acceptance ratio near 50%.[203] For a more concrete discussion of this procedure in application, see Chapter 4.

B.3 Computing the Packing Free Energy

B.3.1 Decomposition of f_{pack}

One of the important goals of packing is to determine the free-energy penalty associated with confining surfactant tails to lie within a micelle core, and constraining them to remain fluid-like. In Chapter 4, the packing free energy was computed as a differ-

ence between the final packed state and the initial free, unconfined state, which led to a term containing a ratio of partition functions for the two states which converges as the number of external orientations is increased (recall that the internal conformations are fully-enumerated and cannot be increased in number). This procedure worked because all chain conformations were considered, including conformations which were valid only in the free state (e.g., where a group in the tail exceeded the micelle core interface). Such conformations contribute to the free-state partition function, but not to the packing partition function.

In the case of Markov-Chain Monte Carlo, the situation is different. We are only sampling conformations that have a non-zero probability in the packed state. As a result, we are not able to directly compute a free energy between the free and packed states. Instead, I introduce a decomposition of the packing free energy which is equivalent to the old approach, but which is more suitable for Monte Carlo modeling:

$$f_{pack} = f_{confine} + f_{constrain}. \tag{B.3.1}$$

where $f_{confine}$ is the free-energy penalty of confinement of a free chain to a bounded volume, and $\underline{F}_{constrain}$ is the free-energy penalty of constraining that chain to exhibit a mean distribution of volume which satisfies constant-density constraints.

B.3.2 Calculation of $f_{confine}$

The first term, $f_{confine}$, involves the entropic penalty that arises from restricting the number of accessible states of a chain by virtue of imposing a core boundary. In this case, no particular lateral pressure field is applied (one can envision a constant-pressure field instead, where there is no preference for any particular arrangement of chain volume within the core, but the chain must reside completely within the core). A hybrid method suffices in this situation: moves are proposed to change the internal conformation of the chain, then a large (constant) number of external orientations

are performed randomly. The free energy of confinement is then computed as follows:

$$f_{confine} = -k_B T \ln\left(\frac{Z_{confine}}{Z_{free}}\right), \tag{B.3.2}$$

where the ratio $Z_{confine}/Z_{free}$ is equal to the fraction of chains that fall within the core boundary when averaged across, such that $f_{confine} > 0$. Note that, for spheres, cylinders, and bilayers, this ratio is always less than 0.5, since, under the best circumstances, a flat interface removes at least half of the possible conformations of the free state, and additional conformations are removed from the remaining half due to chain flexibility (wherein some configurations can still exceed the interface in spite of the head-tail connection being appropriately oriented into the micelle core).

B.3.3 Calculation of $f_{constrain}$ – The Overlapping Distribution Method

Using a non-linear solver to propose $\{\pi_l\}$ values, and the procedures in Section B.2 to generate the appropriate expectation values needed to evaluate the constant density constraints, a final set of lateral pressures can be obtained which represents convergence to the fully-constrained state. Since the $f_{confine}$ term above links the free state to a confined, constant-pressure state, one needs a technique to bridge two systems with identical geometries but having differing distributions (characterized by differing $\{\pi_l\}$ values – for convenience we can consider the constant-pressure state to be a zero-pressure state without affecting results, since absolute free energies can be known only to within an additive constant).

An effective procedure for accomplishing this task is the Overlapping Distribution Method (ODM).[203,280,281] In particular, I follow the discussion of Frenkel and Smit[203] in presenting the key equations of the ODM developed by Bennett.[281] The discussion here will be brief – application to practical systems is reserved for Chapter 4.

The partition function of a canonical ensemble is given by

$$Z_\mu = \int \exp\left(-\beta U_\mu\right) d\mu, \tag{B.3.3}$$

where the energy U_μ is computed as $U_\mu = \epsilon_\mu + \sum_{l=1}^{L} \pi_l v_{\mu,l}$ in a packing context (see Eq. B.1.2 above).

If we have two pdfs labeled 0 and 1 (e.g., differing in $\{\pi_l\}$ values), the Helmholtz free energies are computed as

$$f_1 = -k_B T \ln Z_{1,\mu}, \tag{B.3.4}$$

and

$$f_0 = -k_B T \ln Z_{0,\mu}. \tag{B.3.5}$$

The difference in these free energies is simply:

$$\Delta f = f_1 - f_0 = -k_B T \ln \left(\frac{Z_{1,\mu}}{Z_{0,\mu}} \right), \tag{B.3.6}$$

which can be expanded using Eq. B.3.3, as follows:

$$\Delta f = -k_B T \ln \left(\frac{\int \exp\left(-\beta U_{1,\mu}\right) d\mu}{\int \exp\left(-\beta U_{0,\mu}\right) d\mu} \right). \tag{B.3.7}$$

We now define a difference in energies between the two pdfs for a particular chain state μ:

$$\Delta U_\mu \equiv U_{1,\mu} - U_{0,\mu}. \tag{B.3.8}$$

We can now ask what the probability of a particular ΔU_μ is when running a Monte Carlo simulation under the $\{\pi_l\}$ conditions of pdf 1:

$$P_1\left(\Delta U_\mu\right) = \frac{\int \exp\left(-\beta U_{1,\mu}\right) \delta\left(U_{1,\mu} - U_{0,\mu} - \Delta U_\mu\right) d\mu}{Z_{1,\mu}}. \tag{B.3.9}$$

The function in Eq. B.3.9 collects together all of the states where Eq. B.3.8 is exactly satisfied, via the Dirac delta distribution δ. Equation B.3.9 can be further manipulated using Eq. B.3.8, as follows:

$$
\begin{aligned}
P_1(\Delta U) &= \frac{\int \exp\left(-\beta\left(U_{0,\mu}+\Delta U_\mu\right)\right)\delta\left(U_{1,\mu}-U_{0,\mu}-\Delta U_\mu\right)d\mu}{Z_{1,\mu}} \\
&= \frac{Z_{0,\mu}}{Z_{1,\mu}}\exp\left(-\beta\Delta U_\mu\right)\frac{\int \exp\left(-\beta U_{0,\mu}\right)\delta\left(U_{1,\mu}-U_{0,\mu}-\Delta U_\mu\right)d\mu}{Z_{0,\mu}} \\
&= \frac{Z_{0,\mu}}{Z_{1,\mu}}\exp\left(-\beta\Delta U_\mu\right)P_0\left(\Delta U_\mu\right), \qquad \text{(B.3.10)}
\end{aligned}
$$

where the first step is made possible by the presence of the Dirac delta distribution, the second step is a trivial manipulation (multiplying and dividing by $Z_{0,\mu}$, and the third step involves an analogous expression to Eq. B.3.9). Next, returning to Eq. B.3.6, we can now make use of Eq. B.3.10:

$$
\begin{aligned}
\Delta f &= -k_B T \ln\left(\frac{Z_{1,\mu}}{Z_{0,\mu}}\right) \\
&= -k_B T \ln\left(\frac{\exp\left(-\beta\Delta U_\mu\right)P_0\left(\Delta U_\mu\right)}{P_1\left(\Delta U_\mu\right)}\right) \\
&= \Delta U_\mu - k_B T\left(\ln P_0\left(\Delta U_\mu\right) - \ln P_1\left(\Delta U_\mu\right)\right) \qquad \text{(B.3.11)}
\end{aligned}
$$

Rearranging Eq. B.3.11, we find that

$$
\ln P_1\left(\Delta U_\mu\right) = \beta\left(\Delta f - \Delta U_\mu\right) + \ln P_0\left(\Delta U_\mu\right). \qquad \text{(B.3.12)}
$$

We now define two convenient functions:

$$
\zeta_0\left(\Delta U_\mu\right) \equiv \ln P_0\left(\Delta U_\mu\right) - \frac{\beta\Delta U_\mu}{2}, \qquad \text{(B.3.13)}
$$

and

$$
\zeta_1\left(\Delta U_\mu\right) \equiv \ln P_1\left(\Delta U_\mu\right) + \frac{\beta\Delta U_\mu}{2}. \qquad \text{(B.3.14)}
$$

To compute $\zeta_0\left(\Delta U_\mu\right)$ in Eq. B.3.13, one runs a Monte Carlo simulation where the pdf is set to pdf 0 (say, the constant-pressure situation). For each state generated in the Markov chain at equilibrium, one computes ΔU_μ, and increments a counter in

a histogram of ΔU_μ values. Probabilities $P_0(\Delta U_\mu)$ are determined directly from this histogram. Similarly, $\zeta_1(\Delta U_\mu)$ is computed by running a Monte Carlo simulation where the pdf is set to pdf 1 (say, the constrained set $\{\pi_l\}$). Again, each state generated in the Markov chain at equilibrium is followed by computation of ΔU_μ and appropriate binning, from which $P_1(\Delta U_\mu)$ can be computed.

The definitions of the functions in Eqs. B.3.13 and B.3.14 were carefully chosen by Bennett.[281] Upon inserting them back into Eq. B.3.12, we find:

$$\zeta_1(\Delta U_\mu) - \frac{\beta \Delta U_\mu}{2} = \beta(\Delta f - \Delta U_\mu) + f_0(\Delta U_\mu) + \frac{\beta \Delta U_\mu}{2}, \tag{B.3.15}$$

which simplifies to

$$\beta \Delta f = \zeta_1(\Delta U_\mu) - \zeta_0(\Delta U_\mu). \tag{B.3.16}$$

Accordingly, the free-energy difference can be directly calculated as the difference between ζ_1 and ζ_0. In practice, since each histogram has many points at which this difference can be evaluated, I use the entire population to first determine statistical outliers due to poor sampling in any particular ΔU bin, remove these outliers if they fail a statistical outlier test, and then average across the remaining bins to arrive at a value for Δf. A variance can also be computed to estimate error. In many applications, following the removal of outliers, the standard deviation of the mean is around $0.2 k_B T$.

If the lateral pressures in the confined state are selected to be uniformly zero, the final relationship for $f_{constrain}$ becomes:

$$f_{constrain} = -k_B T \ln \frac{Z_{constrain}}{Z_{confine}} - \sum_l \pi_l \langle v_{\mu,l} \rangle \tag{B.3.17}$$

where the first term is computed using the ODM (Δf), and the second (lateral pressure-volume) term is computed for the constrained pdf. If the lateral pressures in the confined state were set to some constant, but non-zero value, an additional lateral pressure-volume term would be needed in Eq. B.3.17.

As a final comment, the above expressions for $f_{confine}$ and $f_{constrain}$ were presented

in single-component form. Extension to the multi-component case involves computing the lateral pressures as applied to the whole micelle, then composition-weighting the individual species $f_{confine}$ and $f_{constrain}$ terms. In other words, mixed-chain systems are assumed to exhibit non-ideality only through the lateral pressure term. See Eq. 4.2.55 in Chapter 4 for an example of composition-weighting to arrive at a full expression for the packing free energy.

Appendix C

Molecular Simulation Algorithms

C.1 Overview and Purpose

In my thesis research, I have made many substantial improvements and new contributions to the combined computer-simulation/molecular-thermodynamic (CS-MT) framework of Stephenson and coworkers[108-110] (see Chapter 3 for a detailed description of this framework). These include: (i) setting solvent accessible surface area probe radii and atomic van der Waals radii to literature-consistent values, (ii) generating a procedure for determining surfactant head and tail from simulation, (iii) introducing statistical replicates to all analyses, (iv) randomizing multi-component surfactant micelle generation for arbitrary aggregation numbers and compositions (the original scripts were deterministic and limited to a few fixed compositions), and (v) creating a full suite of programs for generating and analyzing simulation data (the original approach involved laborious use of spreadsheets, which was error-prone and not scalable for large studies). The current CS-MT workflow, which I developed, is presented in Fig. C-1 as a flowchart.

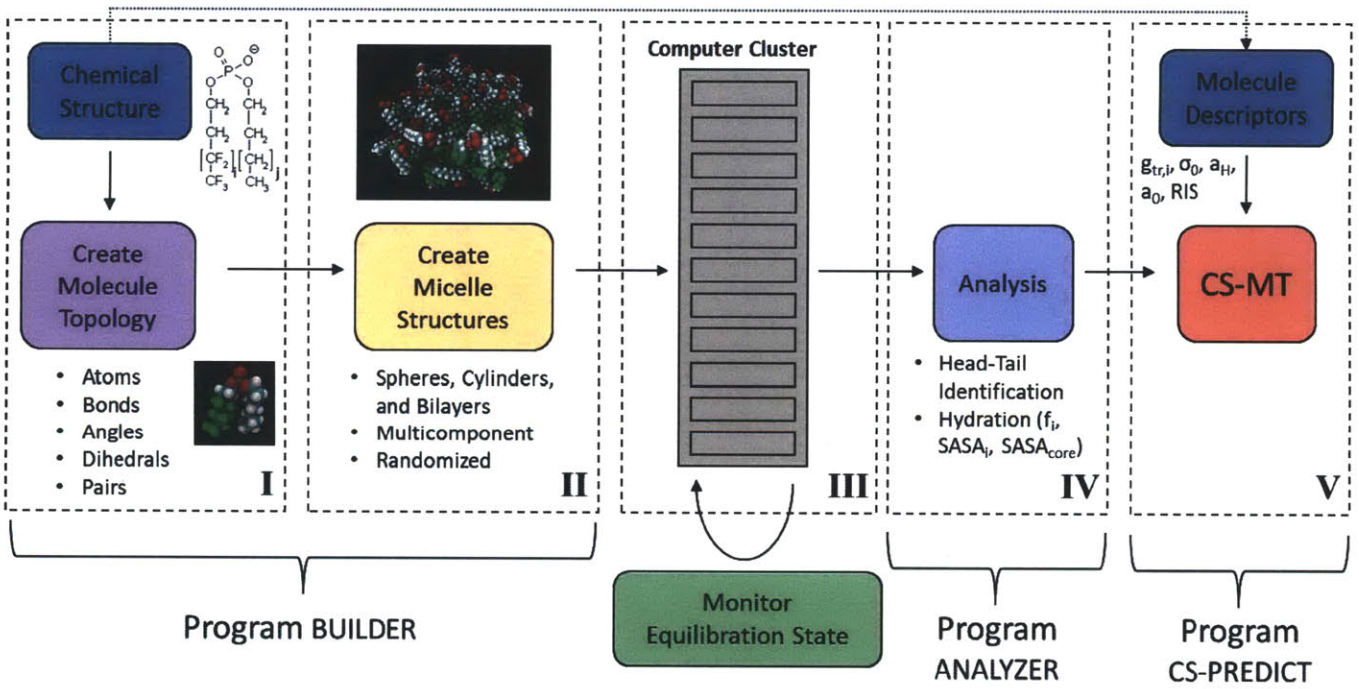


Figure C-1: The workflow for a CS-MT calculation (divided into five main Tasks), with associated programs. In Task I, the user specifies a chemical structure (such as the hybrid fluorocarbon/hydrocarbon phosphodiester shown here), from which a valid molecular topology file and library of configurations is generated. In Task II, single or multi-component micelles are created from these configuration libraries for a specified geometry, aggregation number, and composition, in a randomized fashion. In Task III, micelle (or monomer) simulation files are submitted to a computer cluster, where progress is actively monitored. In Task IV, analyses of the simulation results yield a head-tail identification, hydration data, and other properties (such as counterion binding). In Task V, the analysis results are combined with molecular descriptors based on the structure specified in Task I and provided as input to the CS-MT framework, from which micellization properties are computed. Program BUILDER is responsible for handling Tasks I and II. The analysis scripts and programs responsible for handling Task IV are collectively referred to as Program ANALYZER, and the program used for the final CS-MT calculations in Task V is Program CS-PREDICT (a significantly improved version of the Blankschtein Group Programs PREDICT and MIX2).

The three programs mentioned in Fig. C-1 – Program BUILDER, Program ANALYZER, and Program CS-PREDICT – all involved a substantial effort to create. In this Appendix, I will focus on the various algorithms that I selected and used in the first program mentioned, Program BUILDER. The types of analyses performed using Program ANALYZER are described in Chapters 5 and 6, while the theories implemented in Program CS-PREDICT are described in Chapters 2 and 3.

Program BUILDER was conceived and developed to address the need for rapidly creating micelle topologies in a more randomized way, particularly for multicomponent surfactant micelles. I also have been able to modify the program for use in force-field parameterization applications (specifically, bond torsion parameterization – see Section E.2.4 of Appendix E for more details).

When creating surfactant micelles for simulation, there are two main tasks to perform, represented as Tasks I and II in Fig. C-1, and handled in turn by Program BUILDER.

In Task I in Fig. C-1, the user specifies the head group of a surfactant of interest, its main chain length, and the details of any branching present (including the positioning of branches, their lengths, and the stereochemistry of each nodal carbon). Using chemical graph theory (see Section C.2 below) and a force field, one can create a fully-specified molecular "topology" for the surfactant, which is a listing of: (i) force-field atom types, masses and partial charges for each atom in the molecule, along with an assigned numbering for the atoms, (ii) all bonds, angles, and dihedrals present, encoded as lists of two, three, and four atom numbers, respectively, and (iii) all special non-bonded pairs (e.g., 1-4 pairs in OPLS-AA), encoded as a list of pairs of atom numbers.[190] For a molecular dynamics (MD) program such as GROMACS,[189] the force-field parameterization can be either included in the topology explicitly, or read from a force-field file using the atom types specified. I typically will include any special force-field overrides or uniquely generated parameterizations explicitly, rather than add such entries to the force field directly. In this manner, the topology files generated are transferable to other users of the GROMACS software (since the force-field files installed by default are thereby unaltered).

Once a topology has been generated, it becomes possible to create a Cartesian coordinate representation of the molecule by selecting equilibrium bond lengths and angles as initial states, and choosing representative dihedral angles (e.g., aromatic rings, such as benzene, will have an initial sequence of dihedrals of 0° and 180° that allow for in-plane ring closure, while linear alkanes can be initialized in a *trans* configuration, where all *C-C-C-C* dihedrals are set to 180°). A two-step placement method (described in Section A.4 of Appendix A) can then be applied to obtain the Cartesian coordinates of each atom from these internal coordinates.

Given the topology, force field, and initial coordinates, one can then conduct MD simulations of the molecule in vacuum or other environments (which may require additional topologies and coordinates to represent, say, solvent molecules in a liquid system). As suggested by Bogusz and coworkers,[96] one can also build up a library of surfactant configurations (using Langevin Dynamics to represent solvent collisions implicitly for a surfactant otherwise modeled in vacuum – see Section C.6 below).

In Task II in Fig. C-1, in which the actual micelle is to be created (requiring one topology file per unique species present and one set of coordinates specifying the position of all atoms in the system), user input of the desired micelle shape, aggregation number, and composition allows Program BUILDER to proceed with the following four-step method for micelle generation. (Note that no solvent need be present in executing these steps, and the solvent addition step is delegated to GROMACS prior to an MD run via the "genbox" command.[190])

In Step 1, for the aggregation number specified, a set of "attachment points" approximately equally-spaced on the surface of the micelle geometry of interest are generated, using a variation of a particle-based method described by Paul Bourke (University of Western Australia) for spheres (see http://paulbourke.net/geometry/spherepoints/ for more information). This involves populating the surface randomly with unit charges (one per desired attachment point), each connected to either the center of the sphere, center line of the cylinder, or center plane of the bilayer via damped springs, with rest length equal to the radial dimensions of the user-specified initial geometry. (Note that cylinders and bilayers require the additional feature of periodic bound-

ary conditions in one and two directions, respectively.) A steepest-descent energy minimization, followed by a short Langevin Dynamics run of this system leads to rearrangment of the unit charges spatially to maximize the distance between charges, thereby minimizing the electrostatic potential. The springs ensure constraint of the unit charges to the surface.

In Step 2, once the positions of the attachment points are computed, they are filled randomly with surfactant molecules by: (i) drawing surfactant configurations randomly from the configuration libraries, such that the total count of each surfactant is preserved in accordance with the target counts determined from the user-specified overall aggregation number and micelle composition, and (ii) pairing each drawn surfactant configuration with one unoccupied attachment point (selected randomly), followed by removing that attachment point from consideration for future draws. The actual placement of a selected surfactant molecule at an attachment point involves: (a) translating the coordinates of the drawn configuration, such that an atom identified in advance as the anchor aligns with the coordinates of the attachment point, and (b) rotating the surfactant such that its alkyl tail is pointed inwards, towards the center of the geometry. Restrictions on this random micelle generation process can be introduced as desired; for example, one may want to ensure that the two leaflets of a bilayer have equal composition.

In Step 3, an initial micelle configuration generated in the manner described in Steps 1 and 2 above will likely exhibit atomic "overlaps" between molecules – a situation which involves two non-bonded atoms on separate molecules being found close together in space, such that their van der Waals radii overlap significantly. Such atomic overlaps lead to the generation of large forces within an MD simulation, which causes numerical instabilities at the typical all-atom simulation timesteps of 1-2 fs. Therefore, the positions of the centers of mass of the surfactants must be scaled radially (and also laterally, in the case of cylinders and bilayers), until no overlaps are found to occur.

In Step 4, a final compression step is needed to address the fact that the micelle which results from the scaling of surfactant positions outward in Step 3 is usually too

diffuse to simulate (i.e., there is too much void space in the interior which would be filled by solvent – leading to the rapid disintegration of the micelle structure). This compression step involves conducting a special Velocity Verlet Dynamics simulation (see Section C.5 below), in which the centers of mass of each molecule are given a "push" towards the center of the micelle geometry. That is, a fictitious velocity component is added to the velocities of all the atoms in each molecule, where this component varies depending on a given molecule's position in the micelle – namely, its distance from the micelle center point (sphere), center line (cylinder), or center plane (bilayer). The parameters related to the strength of the push are specified by the user, and the compression step is considered complete when a test probe representing a solvent molecule fails an insertion test at some threshold level (also set by the user – typically a requirement of at least 90% insertion failure yields a suitably "dry" core for simulation). An example of this compression process for a di-C8,12-alkyldimethylammonium bromide surfactant micelle is shown in Fig. C-2.

Following this four-step procedure, one obtains a micelle which is ready for MD simulation using GROMACS. The only details that remain include adding explicit solvent to the simulation box containing the micelle, and performing a subsequent energy minimization prior to simulation (to reduce the likelihood of numerical instabilities). These aspects are addressed with suitable GROMACS function calls prior to a full MD run.

In the remainder of this Appendix, I describe the key algorithms used in Program BUILDER (mentioned or alluded to in the discussion above), in more detail. Program BUILDER was used in generating all the monomer and micelle initial configurations and files run as part of the molecular simulations aspect of my Thesis research. I developed Program BUILDER using Nokia's Qt/C++ platform (under LGPL license). Program BUILDER is currently proprietary software of the Blankschtein group.

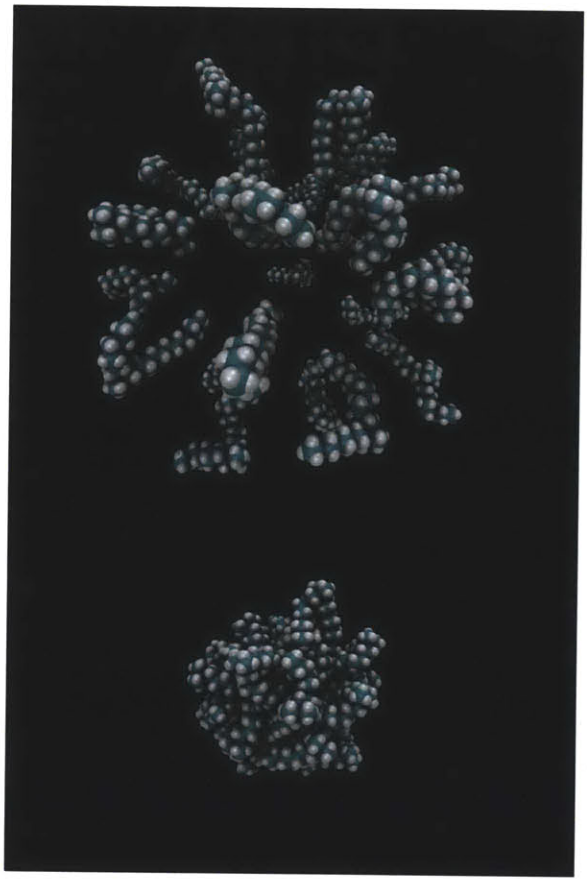


Figure C-2: Compression of a micelle from an expanded state (top) to a compact state (bottom) at the end of the micelle generation process in Program BUILDER. The compact state reduces the likelihood of solvent being inserted into the micelle when using the genbox command in GROMACS (which fills the simulation cell with explicit solvent). The micelle shown is composed of 24 di-C8,12-alkyldimethylammonium bromide surfactants (color code: white – hydrogen, teal – carbon).

C.2 Topology Generation

Chemical graph theory is the application of graph theory to the description of molecular structures.[306] A graph $G(V, E)$ consists of a set of vertices, V, together with a set of edges, E, connecting pairs of vertices, and therefore is well suited for the description of a molecule, where atoms (vertices) are connected to each other via bonds (edges). A walk between two vertices $x_0, x_k \in V$ consists of a sequence of vertices $\{x_0, x_1, ..., x_{k-1}, x_k\}$ that link the starting vertex in the sequence, x_0, to the ending vertex in the sequence, x_k, through edge-connected intermediate (or "inner") vertices.[307] The number of edges traversed in a walk is its length. Special types of walks include: (i) self-returning walks, where the starting and ending vertices are the same, (ii) self-avoiding walks (also called paths), where neither edges nor vertices may be revisited, and (iii) cycles, which are walks where $x_0 = x_k$ (i.e., self-returning) and $\{x_0, x_1, ..., x_{k-1}\}$ is a path (i.e., self-avoiding for all but the final vertex).[306,307] (Cycles are often encountered when dealing with molecules containing rings, aromatic or otherwise). The notion of self-avoiding walks, or paths, is used in defininining the distance between two vertices, which is taken to be the length of the shortest path between them.[307]

Figure C-3 shows two examples of molecular graphs and a related representation, the rooted tree, which I will describe shortly. The two molecules portrayed are ethane (carbons at vertices 4 and 5, hydrogens at vertices 1-3 and 6-8) and benzene (carbons at vertices 1-6, hydrogens at vertices 7-12), where benzene clearly exhibits a cycle (e.g., the sequence of vertices {1,2,3,4,5,6,1} is a cycle of length 6). In developing a molecular topology for simulating these two compounds, we must identify bonds, angles, and dihedrals in these molecules. It should be clear that bonds correspond to a walk of length 1, angles correspond to a walk of length 2, and dihedrals correspond to a walk of length 3. Any two vertices which have a distance of three or less, then, must have a bonded representation within the topology. For benzene, note that vertices such as 1 and 4 will actually exhibit two dihedrals ({1,2,3,4} and {1,6,5,4}). For rings with fewer members than 6, it is also possible for two vertices to share multiple types

of bonded interactions. For example, in tetrahydrofuran – a five-membered ring – any two atoms in the ring will be connected via a bond angle *and* a bond dihedral. In my work, all systems consist of non-cyclic or six-membered cycles only. Therefore, it is sufficient to determine the minimum distance between any two vertices, and compute all the unique paths for that distance, for the purposes of completing the topology.

Computing the distance between two vertices can be accomplished readily via the concept of an adjacency matrix.[306] This is an $N \times N$ matrix $\mathbf{A}(G)$, for a graph G whose set of vertices V comprises N elements. The elements of $\mathbf{A}(G)$ are a_{ij}, where $a_{ij} = 1$ if there is an edge in E connecting vertices $x_i, x_j \in V$, and $a_{ij} = 0$ otherwise. Using ethane in Fig. C-3 as an example, one finds:

$$\mathbf{A}(G) = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \left[\begin{array}{cccccccc} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right] \end{array} . \tag{C.2.1}$$

Note that multiplying $\mathbf{A}(G)$ by a column vector containing only one non-zero entry, say at position k, will return a column vector consisting of non-zero entries only in positions corresponding to vertices directly connected to vertex $x_k \in V$ via an edge in E. If one left-multiplies this resulting vector by $\mathbf{A}(G)$ again, one obtains a new result containing non-zero entries for vertices which are *two* edges away from the starting vertex (note that this will include the original vertex x_k!). Continuing this process, one finds that applying the matrix product $\mathbf{A}^n(G)$ to the original column vector (non-zero entry only for vertex k) will return a vector containing non-zero entries for all vertices accessible from vertex x_k by a walk of length n. In other words, if $a_{ij}^{(n)}$ is an element of $\mathbf{A}^n(G)$ and $a_{ij}^{(n)} \neq 0$, x_i and x_j are connected by a walk of

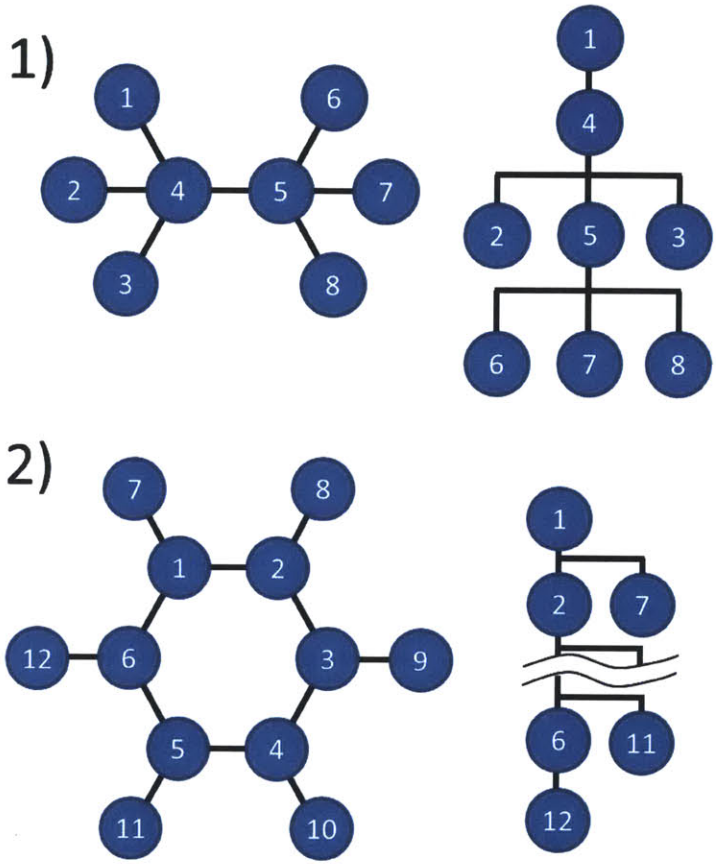


Figure C-3: Molecular graph (left) and rooted tree (right) for: 1) ethane, and 2) benzene. Each molecular graph consists of a collection of vertices (blue circles) and edges (black lines) connecting the vertices. Each rooted tree form consists of a linked list of nodes (blue circles) starting at the root (here, vertex 1). (The two wavy lines associated with the rooted tree at bottom right simply indicate that not all the vertices are being shown, to reduce the size of the diagram for display purposes.) The graph form is useful for finding and specifying all bonded interactions in a molecule (i.e., bonds, angles, and dihedrals), while the rooted-tree form is useful for systematically determining the Cartesian coordinates for all atoms in a molecule from their internal coordinates. As demonstrated in 2), molecular graphs may contain cycles, while trees may not. Therefore, when storing molecules in a tree representation, additional information must be stored regarding broken connectivity (e.g., that vertex 6 connects back to vertex 1 in benzene).

length n. If one performs successive multiplications of $\mathbf{A}(G)$ and records the first value of n at which $a_{ij}^{(n)} \neq 0$ for each pair x_i and x_j $(i \neq j)$, and finds for a particular pair that $n \leq 3$, then that pair has a bonded interaction. It remains to trace the explicit sequence of vertices that accomplish the walk – this is performed in Program BUILDER to obtain the list of bonded interactions.

Returning to Fig. C-3, the rooted trees depicted for each molecule are obtained by selecting a certain atom to be the root vertex.[307] Here, vertex 1 is chosen for this purpose for both molecules. Then, all vertices connected to vertex 1 via an edge are listed below vertex 1. Taking these vertices, one then examines all vertices connected to these by an edge, prohibiting back-tracking to vertices already recorded. The result is a loss of information regarding cycles within the molecules, but the rooted-tree representation is particularly effective for building the Cartesian coordinate representation of a molecule, as it can be converted into a Z-matrix[228] representation. Take ethane for example. Vertex 1 may be placed in space freely. The next vertex placed, vertex 4, has two degrees of freedom, since it must satisfy the bond length connecting atoms 1 and 4 (the H-C bond length). The third vertex placed, vertex 5, has one degree of freedom, since it must respect the 1-4-5, H-C-C bond angle. All vertices placed subsequently must respect bond lengths, bond angles, and bond dihedrals. For N atoms, there are $3N$ positional degrees of freedom, but $3N-6$ constraints, leaving 6 degrees of freedom corresponding to molecular translation and rotation. Program BUILDER applies this type of technique in building up molecules, with the additional caveat that all dihedrals are proper (improper dihedrals have three of the four atoms in the dihedral emanating from a common atom – used to keep rings planar in some force fields, for example[190]). For this reason, a placement order of vertex 1, 4, 2 is disallowed for ethane, since subsequent placement of vertex 3 or 5 would require use of an improper dihedral 1-4-2-3 or 1-4-2-5, where atom 4 is the common atom from which atoms 1, 2, and 3 or 5 emanate.

C.3 Velocity Initialization

Using the techniques of Section C.2, it is possible to build up an initial configuration for the atoms in a surfactant molecule through the use of a rooted tree, and specify the molecule topology through the analysis of the molecular graph. The sole remaining requirement for MD simulation is the specification of an initial set of velocities for each atom (when using the Velocity-Verlet algorithm described in Section C.5, or the Langevin Dynamics algorithm described in Section C.6 – other time integration schemes may require a different set of initial conditions).

Given knowledge of the system temperature, initial velocities can be generated randomly from the Maxwell-Boltzmann distribution, as follows:[202]

$$p\left(v_{x,i}\right) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left(-\frac{1}{2}\frac{m_i v_{x,i}^2}{k_B T}\right), \tag{C.3.1}$$

where $p\left(v_{x,i}\right)$ is a normal distribution for $v_{x,i}$ (the velocity of atom i in the x-direction), with zero mean and standard deviation $\sigma = \left(k_B T/m_i\right)^{1/2}$, m_i is the mass of atom i, k_B is the Boltzmann constant, and T is the specified system temperature. See the discussion of random number generation in Section C.6 below for a discussion on how to generate $v_{x,i}$ randomly from Eq. C.3.1. The velocities in the y and z directions follow a distribution identical to Eq. C.3.1, simply requiring substitution of $v_{y,i}$ or $v_{z,i}$ for $v_{x,i}$, respectively.[202]

C.4 Steepest-Descent Energy Minimization

To encourage a numerically-stable simulation, it is useful to precede all MD runs with an energy-minimization step. This typically involves application of a steepest-descent algorithm, which adjusts atom positions in a sequence, such that new positions are generated from current positions and forces, as follows:[190]

$$\begin{aligned} \underline{r}_i^{m+1} &= \underline{r}_i^m - \kappa^m \frac{\partial U\left(\{\underline{r}_i^m\}\right)}{\partial \underline{r}_i} \\ &= \underline{r}_i^m + \kappa^m \underline{F}_i\left(\{\underline{r}_i^m\}\right), \end{aligned} \tag{C.4.1}$$

where m indexes the steps in the sequence, starting with $m = 0$ (the initial state), $\underline{r}_i^m$ is the position of atom i at step m, $U(\{\underline{r}_i^m\})$ is the total configurational system potential (a function of the full set of positions $\{\underline{r}_i^m\}$ of all atoms in the system), $\underline{F}_i(\{\underline{r}_i^m\})$ is the net force vector acting on atom i (also a function of all atom positions), and κ^m is a positive, step-dependent parameter, equal to the ratio of the maximum change in position desired by the user per sequence step divided by the absolute value of the maximum magnitude force component encountered at the given step across all $\{\underline{F}_i\}$. Consequently, this ensures that all atoms in the system will have a well-defined, limited change in position, which ensures that atomic overlaps will be resolved in a stable fashion (compared to the alternative of unbounded displacement at each step).

C.5 Velocity-Verlet Dynamics

Given an expression for the system potential as a function of atomic positions, from which individual forces on atoms can be derived (see Section D.2 of Appendix D for the specific example of OPLS-AA[188]), an algorithm can be selected to integrate Newton's equations of motion from a supplied initial configuration. For this purpose, the Velocity-Verlet algorithm[282] is convenient, for two reasons. First, it involves evaluation of new positions and velocities at identical time points, as opposed to the Leap-Frog method, in which velocities are evaluated at time points offset from those for position by half a time step.[203] Second, it involves a convenient initial condition, requiring initial positions and velocities at a single point in time (in contrast to the Verlet algorithm,[308] which instead requires positions at two time points). The integrating equations for the Velocity-Verlet algorithm include (to be applied in this order):[282]

$$\underline{r}_i(t+\Delta t) = \underline{r}_i(t) + \underline{v}_i(t)\,\Delta t + \frac{1}{2}\underline{a}_i(t)\,\Delta t^2 \tag{C.5.1}$$

$$\underline{a}_i(t+\Delta t) = \frac{\underline{F}_i(t+\Delta t)}{m_i} = -\frac{1}{m_i}\frac{\partial U(\{\underline{r}_i(t+\Delta t)\})}{\partial \underline{r}_i} \tag{C.5.2}$$

$$\underline{v}_i(t+\Delta t) = \underline{v}_i(t) + \frac{1}{2}\left(\underline{a}_i(t) + \underline{a}_i(t+\Delta t)\right)\Delta t, \tag{C.5.3}$$

where Δt is the integrating time step; $\underline{r}_i$, $\underline{v}_i$, and $\underline{a}_i$, are the vectors for the position, velocity, and acceleration of atom i, $\underline{F}_i$ is the net force vector acting on atom i, m_i is the mass of atom i, and $U(\{\underline{r}_i\})$ is the total system configurational potential, a function of all the atomic positions in the system. Note that application of Eq. C.5.1 at the first time step requires an initial calculation of Eq. C.5.2 based on the initial condition (where $\Delta t = 0$). The above equations can also be recast in a slightly more convenient form:

$$\underline{v}_i\left(t+\frac{\Delta t}{2}\right) = \underline{v}_i(t)+\frac{1}{2}\underline{a}_i(t)\,\Delta t \tag{C.5.4}$$

$$\underline{r}_i(t+\Delta t) = \underline{r}_i(t)+\underline{v}_i\left(t+\frac{\Delta t}{2}\right)\Delta t \tag{C.5.5}$$

$$\underline{a}_i(t+\Delta t) = -\frac{1}{m_i}\frac{\partial U(\{\underline{r}_i(t+\Delta t)\})}{\partial \underline{r}_i} \tag{C.5.6}$$

$$\underline{v}_i(t+\Delta t) = \underline{v}_i\left(t+\frac{\Delta t}{2}\right)+\frac{1}{2}\underline{a}_i(t+\Delta t)\,\Delta t. \tag{C.5.7}$$

In addition to eliminating one redundant calculation, this form also allows in-place updates of acceleration within the program implementation, since Eq. C.5.7, the velocity update, no longer contains accelerations at two time points.

Errors due to Taylor series truncation in the derivation of the Velocity-Verlet algorithm yields errors which are $O(\Delta t^4)$ for positions and $O(\Delta t^2)$ for velocities.[203]

C.6 Langevin Dynamics

The development of libraries of surfactant configurations is efficiently accomplished using stochastic dynamics simulations of single monomers in vacuum, where random vectors are added to both positions and velocities during each time step of the integration. Note that these random vectors represent an approximation of the effect of solvent in jostling the monomer at a given temperature, without the computational cost associated with a full box of explicit solvent.[309] In Program BUILDER, I make use of the Langevin Dynamics algorithm, which has Velocity Verlet-like update

equations.[283] That is,

$$\underline{r}_i(t+\Delta t) = \underline{r}_i(t) + c_1\underline{v}_i(t)\,\Delta t + c_2\underline{a}_i(t)\,\Delta t^2 + \delta\underline{r}_{R,i} \quad \text{(C.6.1)}$$

$$\underline{v}_i(t+\Delta t) = c_0\underline{v}_i(t) + (c_1 - c_2)\,\underline{a}_i(t)\,\Delta t + c_2\underline{a}_i(t+\Delta t)\,\Delta t + \delta\underline{v}_{R,i}, \quad \text{(C.6.2)}$$

where $\delta\underline{r}_{R,i}$ and $\delta\underline{v}_{R,i}$ are the random vectors mentioned, the accelerations are evaluated as in Eq. C.5.2, and the coefficients c_0, c_1, and c_2 are computed as follows:[202]

$$c_0 = e^{-\gamma\Delta t} \quad \text{(C.6.3)}$$

$$c_1 = \frac{(1-c_0)}{\gamma\Delta t} \quad \text{(C.6.4)}$$

$$c_2 = \frac{(1-c_1)}{\gamma\Delta t}, \quad \text{(C.6.5)}$$

where γ is a friction coefficient (collision frequency), which, for the purposes of Program BUILDER and monomer configuration generation, is typically set to 5 ps^{-1}, following the example of Bogusz and coworkers in their study of octyl glucoside systems.[96]

The random displacements $\delta\underline{r}_{R,i}$ and $\delta\underline{v}_{R,i}$ in Eqs. C.6.1 and C.6.2, respectively, follow a bivariate normal distribution.[283,310] That is,

$$W\left(\delta\underline{r}_{R,i}, \delta\underline{v}_{R,i}\right) = \frac{1}{8\pi^3\left(FG-H^2\right)^{3/2}} \times \exp\left[-\frac{\left(G\left\|\delta\underline{r}_{R,i}\right\|^2 - 2H\left(\delta\underline{r}_{R,i}\cdot\delta\underline{v}_{R,i}\right) + F\left\|\delta\underline{v}_{R,i}\right\|^2\right)}{2\left(FG-H^2\right)}\right], \quad \text{(C.6.6)}$$

where

$$F = \sigma_r^2 = \Delta t\frac{k_BT}{m\gamma}\left[2 - \frac{\left(3-4e^{-\gamma\Delta t}+e^{-2\gamma\Delta t}\right)}{\gamma\Delta t}\right] \quad \text{(C.6.7)}$$

$$G = \sigma_v^2 = \frac{k_BT}{m}\left(1-e^{-2\gamma\Delta t}\right) \quad \text{(C.6.8)}$$

$$H = \sigma_r\sigma_v c_{rv} = \frac{k_BT}{m\gamma}\left(1-e^{-\gamma\Delta t}\right)^2, \quad \text{(C.6.9)}$$

and σ_r^2 is the variance in the position term, σ_v^2 is the variance in the velocity term, and c_{rv} is the correlation coefficient between the two randomized terms. Note that each component of the position and velocity random displacement vectors are correlated, but the individual components within each vector are independent.[283]

It then remains to generate $\delta\underline{r}_R$ and $\delta\underline{v}_R$ in accordance with Eq. C.6.6. Sections C.6.1-C.6.3 which follow briefly describe the procedure.

C.6.1 Generating Random Numbers from a Bivariate Gaussian Distribution

Given a pair of uncorrelated random numbers, (ζ_1, ζ_2), drawn from a univariate Gaussian distribution with zero mean and unit variance (see Section C.6.2 below), a correlated pair of random numbers from a bivariate Gaussian distribution, (ζ_1', ζ_2'), can be readily generated, via[202]

$$\zeta_1' = \sigma_1 \zeta_1 \tag{C.6.10}$$

$$\zeta_2' = \sigma_2 \left(c_{12}\zeta_1 + \left(1 - c_{12}^2\right)^{1/2} \zeta_2 \right), \tag{C.6.11}$$

where σ_1^2 and σ_2^2 are the variances of the bivariate distribution, and c_{12} is the correlation coefficient. Letting $1 = r$ and $2 = v$, and computing σ_1^2, σ_2^2, and c_{12} using Eqs. C.6.7-C.6.9 will allow us to populate the components of $\delta\underline{r}_R$ and $\delta\underline{v}_R$. Specifically, for each atom, we compute the correlated components of $\delta\underline{r}_R$ and $\delta\underline{v}_R$ as (ζ_1, ζ_2) individually (i.e., treating the x-axis, y-axis, and z-axis components independently).

C.6.2 Generating Random Numbers from a Normal (Gaussian) Distribution

The uncorrelated (ζ_1, ζ_2) pair needed above can be generated using an algorithm such as the polar method due to Marsaglia and Bray,[311] which generates two independent normal random numbers with each pass. Specifically, this involves:

(1) Generating two uniform random numbers, ξ_1 and ξ_2, from $(0, 1)$ using an

appropriate algorithm (see Section C.6.3 below).

(2) Creating the pair (v_1, v_2) from $v_1 = 2\xi_1 - 1$ and $v_2 = 2\xi_2 - 1$ (i.e., v_1 and v_2 are uniform random numbers drawn from $(-1, 1)$).

(3) If $v_1^2 + v_2^2 >= 1$, rejecting the pair, and returning to Step (1); otherwise, going to Step (4).

(4) Computing $\zeta_1 = v_1 S$ and $\zeta_2 = v_2 S$, where

$$S = \left(-\frac{2 \ln (v_1^2 + v_2^2)}{v_1^2 + v_2^2} \right)^{1/2} . \tag{C.6.12}$$

C.6.3 Generating Random Numbers from a Uniform Distribution

The generation of random numbers from a uniform distribution in $(0, 1)$ – required for Step (1) in Section C.6.2 – is highly non-trivial,[203] and it is important to avoid the bias found in many deficient algorithms.[312] For this purpose, Program BUILDER employs an open-source implementation of the Mersenne Twister algorithm,[313] which is widely recognized as a high-quality, pseudorandom number generator.[312] This generator has a variety of properties that are appealing, such as an exceptionally large period – the number of values in a generated random number sequence prior to its first repeat – of $2^{19937} - 1$, and is found in broad use, including, for example, in varied Monte Carlo applications,[314] and as the principal random number generator for the Python programming language.[315]

Appendix D

Force Expressions for Molecular Dynamics Simulations

D.1 Overview

As discussed in Appendix C, molecular dynamics (MD) simulations require the specification of a force field to model the behavior of molecules within a system as a function of time. The forces used in modeling are typically conservative, such that there exists a system potential function from which the forces on individual atoms can be calculated as

$$\underline{F}_i = \left(-\frac{\partial U}{\partial r_{x,i}}, -\frac{\partial U}{\partial r_{y,i}}, -\frac{\partial U}{\partial r_{z,i}} \right) = -\frac{\partial U}{\partial \underline{r}_i} \equiv -\underline{\nabla}_i U, \tag{D.1.1}$$

where U is the system potential function, $\underline{F}_i$ is the force vector operating on atom i, $\underline{r}_i = (r_{x,i}, r_{y,i}, r_{z,i})$ is the position vector of atom i, and use of the underbar will signify a vector throughout this Appendix. The system potential function, U, is configurational in nature: it is a function only of the $3N$ positional degrees of freedom for a system containing N atoms.[203] Note that position vectors such as $\underline{r}_i$ require specification of a frame-of-reference. In my work, this is the Cartesian coordinate system of the MD simulation box.

The total differential of the system potential function will be found to be quite

useful in deriving the force expressions in Section D.4 below. Since U is a function of the individual position vectors $\underline{r}_i$, dU is determined as follows:

$$dU = \sum_{i=1}^{N} d\underline{r}_i \cdot \left(\frac{\partial U}{\partial \underline{r}_i} \right) = \sum_{i=1}^{N} d\underline{r}_i \cdot (-\underline{F}_i) , \tag{D.1.2}$$

where Eq. D.1.1 has been used to arrive at the final expression on the right-hand side.

In this thesis, I make extensive use of the OPLS-AA force field of Jorgensen and coworkers[188] to carry out simulations of surfactant micelles in aqueous solution. This force field, as is common practice, includes bonded 1-2 (bond length), 1-3 (bond angle) and 1-4 (bond torsion) interactions and non-bonded Coulombic and van der Waals interactions. (Note that I use "1-M", with M = 2, 3, or 4, to indicate that two atoms are connected through $M - 1$ intervening bonds on a given molecule.) I use the OPLS-AA force field both in my own molecular dynamics code (i.e., Program BUILDER – see Appendix C) and in data production runs using the GROMACS software package.[189]

In this Appendix, I first present the individual contributions to the OPLS-AA system potential function, U, followed by a complete derivation of the associated forces through application of Eq. D.1.2 and direct calculation of dU. These forces are required to implement any MD time-integration algorithm, such as the Velocity-Verlet algorithm (see Appendix C), but they are not always fully documented – see, for example, the GROMACS manual, which is not comprehensive on this point. Therefore, the explicit use of these expressions in Program BUILDER warrants their presentation here.

D.2 OPLS-AA System Potential Function

The OPLS-AA force field consists of a set of parameters and functions which enable calculation of the system potential function, U, given a set of $3N$ atomic coordinates and molecular connectivity information (i.e., specification of how atoms are bonded).

The potential function takes the form[188]

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{Coulombic} + U_{vdW}, \tag{D.2.1}$$

comprising the following five potential-energy contributions (in Eqs. D.2.2-D.2.4, the first form represents the OPLS-AA potential-energy contribution, and the second form represents an equivalent expression used in GROMACS):

$$U_{bond} = \sum_{bonds} K_b \left(b - b_0\right)^2 = \sum_{bonds} \frac{k_b}{2} \left(b - b_0\right)^2 \tag{D.2.2}$$

the bond-stretching contribution, where K_b and k_b are force constants ($k_b = 2K_b$), b is the bond length, and b_0 is the equilibrium (often the *ab initio* geometry-optimized) bond length;

$$U_{angle} = \sum_{angles} K_\theta \left(\theta - \theta_0\right)^2 = \sum_{angles} \frac{k_\theta}{2} \left(\theta - \theta_0\right)^2 \tag{D.2.3}$$

the angle-stretching contribution, where K_θ and k_θ are force constants ($k_\theta = 2K_\theta$), θ is the bond angle, and θ_0 is the equilibrium (often the *ab initio* geometry-optimized) bond angle;

$$U_{dihedral} = \sum_{dihedrals} \sum_{n=1}^{4} \frac{V_n}{2} \left[1 + (-1)^{n+1} \cos\left(n\phi\right)\right] = \sum_{dihedrals} \sum_{n=0}^{5} C_n \cos^n\left(\phi - \pi\right) \tag{D.2.4}$$

the bond-torsion contribution, where V_n and C_n are series coefficients, and ϕ is the dihedral angle;

$$U_{Coulombic} = \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij} \left[\frac{q_i q_j e^2}{r_{ij}}\right] \tag{D.2.5}$$

the non-bonded, Coulombic electrostatics contribution (in cgs units, with the relative permittivity set to unity for explicit solvent simulations), where q is the partial charge of an atom, e is the elementary charge, r_{ij} is the distance between atoms i and j, and

f_{ij} is defined below; and:

$$U_{vdW} = \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij} \left[4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) \right] \tag{D.2.6}$$

the non-bonded, van der Waals contribution, in the form of a 12-6 Lennard-Jones potential,[199] where σ_{ij} and ϵ_{ij} are van der Waals parameters, and f_{ij} (also utilized in the Coulombic term) is a function designed to limit calculation of non-bonded interactions to atom pairs which are not directly bonded.[188] Specifically:

$$f_{ij} = \begin{cases} 0, \text{ if } i = j, \text{ or } i \text{ and } j \text{ are 1-2 or 1-3} \\ \frac{1}{2}, \text{ if } i \text{ and } j \text{ are 1-4} \\ 1, \text{ if } i \text{ and } j \text{ are 1-5 or greater, or on two different molecules} \end{cases} \tag{D.2.7}$$

where we can see that 1-4 atom pairs, which do have a bonded, torsional contribution to the system potential, are modeled as possessing a half-strength non-bonded contribution.

The van der Waals parameters σ_{ij} and ϵ_{ij} in Eq. D.2.6 are determined from single-atom properties using geometric mixing rules:[188]

$$\sigma_{ij} = (\sigma_i \sigma_j)^{1/2} \tag{D.2.8}$$

$$\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2} \tag{D.2.9}$$

In Eq. D.2.4 above, conversion between OPLS-AA series (specifically, Fourier series) and GROMACS series (specifically, Ryckaert-Bellemans series[316]) coefficients V_n and C_n, respectively, is readily accomplished using the formulas that I supply in Section D.5. Note that the units of various quantities in GROMACS differ from those often used in the literature.[190] Table D.1 summarizes these differences.

I now proceed with a review of some useful vector relationships, which I will make use of without further comment in deriving the force expressions.

Table D.1: Comparison of Units in OPLS-AA and GROMACS Potentials.

	OPLS-AA	GROMACS		OPLS-AA	GROMACS
K_b, k_b	(kcal/mol)/Å^2	(kJ/mol)/nm^2	V_i, C_n	kcal/mol	kJ/mol
b, b_0	Å	nm	ϕ	rad	rad
K_θ, k_θ	(kcal/mol)/rad^2	(kJ/mol)/rad^2	σ	Å	nm
θ, θ_0	degree (°)	degree (°)	ϵ	kcal/mol	kJ/mol

D.3 Review of Relevant Vector Relationships

In relating internal to external coordinates in the force-field potential expressions, various vector dot products, norms, and cross products are encountered. The force derivations will additionally require computation of derivatives of these products. Therefore, below, I summarize a few key results for reference, which can be derived or obtained readily from various engineering texts.[167,317]

For the purposes of molecular simulations, we will only require use of three-dimensional Cartesian vectors, which can be expressed as follows:

$$\underline{a} = a_x \underline{\hat{e}}_x + a_y \underline{\hat{e}}_y + a_z \underline{\hat{e}}_z, \tag{D.3.1}$$

where a_x, a_y, and a_z are the components of $\underline{a}$ in the x-axis, y-axis, and z-axis directions represented by the basis vectors $\underline{\hat{e}}_x$, $\underline{\hat{e}}_y$, and $\underline{\hat{e}}_z$.

D.3.1 Dot Product

For vectors $\underline{a}$ and $\underline{b}$, the scalar dot product is given by

$$\underline{a} \cdot \underline{b} \equiv a_x b_x + a_y b_y + a_z b_z. \tag{D.3.2}$$

The geometric interpretation of the dot product is

$$\underline{a} \cdot \underline{b} = |\underline{a}|\,|\underline{b}| \cos\theta, \tag{D.3.3}$$

or

$$\theta = \arccos \frac{\underline{a} \cdot \underline{b}}{|\underline{a}|\,|\underline{b}|}, \tag{D.3.4}$$

where θ is the unsigned and smallest angle between vectors $\underline{a}$ and $\underline{b}$ (note that $0 \leq \theta \leq \pi$, due to the range of arccos).

The differential of the dot product is

$$d\left(\underline{a} \cdot \underline{b}\right) = \underline{a} \cdot d\underline{b} + d\underline{a} \cdot \underline{b}. \tag{D.3.5}$$

Other important properties of the dot product include commutativity:

$$\underline{a} \cdot \underline{b} = \underline{b} \cdot \underline{a}, \tag{D.3.6}$$

compatibility with scalar multiplication:

$$s\left(\underline{a} \cdot \underline{b}\right) = \left(s\underline{a}\right) \cdot \underline{b} = \underline{a} \cdot \left(s\underline{b}\right), \tag{D.3.7}$$

and distributivity over vector addition:

$$\underline{a} \cdot \left(\underline{b} + \underline{c}\right) = \left(\underline{a} \cdot \underline{b}\right) + \left(\underline{a} \cdot \underline{c}\right). \tag{D.3.8}$$

D.3.2 Vector Norm

The magnitude, or norm, of a vector $\underline{a}$ can be expressed in terms of the dot product as follows:

$$|\underline{a}| = \sqrt{\underline{a} \cdot \underline{a}}. \tag{D.3.9}$$

The differential of the vector norm is given by

$$\begin{aligned} d\left(|\underline{r}|\right) &= d\left(\sqrt{\underline{r}\cdot\underline{r}}\right) \\ &= \frac{1}{2\,|\underline{r}|} d\left(\underline{r}\cdot\underline{r}\right) \\ &= \frac{1}{2\,|\underline{r}|}\left(d\underline{r}\cdot\underline{r}+\underline{r}\cdot d\underline{r}\right) \\ &= d\underline{r}\cdot\frac{\underline{r}}{|\underline{r}|}. \end{aligned} \tag{D.3.10}$$

Another important property is that norms are homogeneous of degree 1 (i.e., scalars can be "pulled out" of the norm). That is,

$$|s\underline{a}| = s\,|\underline{a}|\,. \tag{D.3.11}$$

D.3.3 Cross Product

For vectors $\underline{a}$ and $\underline{b}$, the vector cross product in three dimensions is given by

$$\underline{a}\times\underline{b} \equiv \left(a_y b_z - a_z b_y, a_z b_x - a_x b_z, a_x b_y - a_y b_x\right), \tag{D.3.12}$$

where the right-hand side is a vector in ordered-set notation, and use of the right-hand rule is implied.

The geometric interpretation of the cross product is

$$\underline{a}\times\underline{b} \equiv |\underline{a}|\,|\underline{b}|\sin\theta\,\widehat{n}, \tag{D.3.13}$$

or

$$\theta = \arcsin\frac{|\underline{a}\times\underline{b}|}{|\underline{a}|\,|\underline{b}|}, \tag{D.3.14}$$

where θ is the smallest angle between vectors $\underline{a}$ and $\underline{b}$ (note that $-\frac{\pi}{2}\le\theta\le\frac{\pi}{2}$, due to the range of arcsin). Choice of the normal unit vector, $\widehat{n}$, is by the right-hand rule, to agree with Eq. D.3.12.

The differential of the cross product is given by

$$d\left(\underline{a} \times \underline{b}\right) = d\underline{a} \times \underline{b} + \underline{a} \times d\underline{b}. \tag{D.3.15}$$

Other properties of the cross product include anticommutativity:

$$\underline{a} \times \underline{b} = -\left(\underline{b} \times \underline{a}\right), \tag{D.3.16}$$

compatibility with scalar multiplication:

$$s\left(\underline{a} \times \underline{b}\right) = \left(s\underline{a}\right) \times \underline{b} = \underline{a} \times \left(s\underline{b}\right), \tag{D.3.17}$$

and distributivity over vector addition:

$$\underline{a} \times \left(\underline{b} + \underline{c}\right) = \left(\underline{a} \times \underline{b}\right) + \left(\underline{a} \times \underline{c}\right). \tag{D.3.18}$$

D.3.4 Mixed Dot and Cross Product Expressions

Several mixed dot and cross product expressions will appear in the force derivations below. This section lists the relevant important expressions that will be used.

Scalar Triple Product

The scalar triple product is formulated as follows:

$$\underline{a} \cdot \left(\underline{b} \times \underline{c}\right) = \underline{c} \cdot \left(\underline{a} \times \underline{b}\right) = \underline{b} \cdot \left(\underline{c} \times \underline{a}\right), \tag{D.3.19}$$

with differential

$$\begin{aligned} d\left(\underline{a} \cdot \left(\underline{b} \times \underline{c}\right)\right) &= d\underline{a} \cdot \left(\underline{b} \times \underline{c}\right) + \underline{a} \cdot \left(\left(d\underline{b} \times \underline{c}\right) + \left(\underline{b} \times d\underline{c}\right)\right) \\ &= d\underline{a} \cdot \left(\underline{b} \times \underline{c}\right) + \underline{a} \cdot \left(d\underline{b} \times \underline{c}\right) + \underline{a} \cdot \left(\underline{b} \times d\underline{c}\right) \\ &= d\underline{a} \cdot \left(\underline{b} \times \underline{c}\right) + d\underline{b} \cdot \left(\underline{c} \times \underline{a}\right) + d\underline{c} \cdot \left(\underline{a} \times \underline{b}\right). \end{aligned} \tag{D.3.20}$$

Orthogonality of Cross Product and Original Vectors

The orthogonality of a cross product with either of the original two vectors involved in the cross product is clearly seen:

$$\underline{a} \cdot (\underline{a} \times \underline{b}) = \underline{b} \cdot (\underline{a} \times \underline{b}) = 0. \tag{D.3.21}$$

Vector Triple Product

The vector triple product is formulated as follows:

$$\underline{a} \times (\underline{b} \times \underline{c}) = \underline{b}\,(\underline{a} \cdot \underline{c}) - \underline{c}\,(\underline{a} \cdot \underline{b})\,, \tag{D.3.22}$$

from which it follows that

$$(\underline{a} \times \underline{b}) \times (\underline{c} \times \underline{d}) = \underline{c}\,((\underline{a} \times \underline{b}) \cdot \underline{d}) - \underline{d}\,((\underline{a} \times \underline{b}) \cdot \underline{c})\,, \tag{D.3.23}$$

with the special case

$$(\underline{a} \times \underline{b}) \times (\underline{b} \times \underline{c}) = \underline{b}\,((\underline{a} \times \underline{b}) \cdot \underline{c})\,.$$

Dot Product of Two Cross Products

The dot product of two cross products is formulated as follows:

$$\begin{aligned}
(\underline{a} \times \underline{b}) \cdot (\underline{c} \times \underline{d}) &= -\underline{d} \cdot (\underline{c} \times (\underline{a} \times \underline{b})) \\
&= \underline{d} \cdot (\underline{b}\,(\underline{c} \cdot \underline{a}) - \underline{a}\,(\underline{c} \cdot \underline{b})) \\
&= \underline{d} \cdot \underline{b}\,(\underline{c} \cdot \underline{a}) - \underline{d} \cdot \underline{a}\,(\underline{c} \cdot \underline{b}) \\
&= (\underline{a} \cdot \underline{c})\,(\underline{b} \cdot \underline{d}) - (\underline{a} \cdot \underline{d})\,(\underline{b} \cdot \underline{c})\,,
\end{aligned} \tag{D.3.24}$$

with differential

$$\begin{aligned} d\left((\underline{a}\times\underline{b})\cdot(\underline{c}\times\underline{d})\right) &= d\left(\underline{a}\times\underline{b}\right)\cdot(\underline{c}\times\underline{d}) + (\underline{a}\times\underline{b})\cdot d\left(\underline{c}\times\underline{d}\right) \\ &= (d\underline{a}\times\underline{b}+\underline{a}\times d\underline{b})\cdot(\underline{c}\times\underline{d}) + (\underline{a}\times\underline{b})\cdot(d\underline{c}\times\underline{d}+\underline{c}\times d\underline{d}) \\ &= (d\underline{a}\times\underline{b})\cdot(\underline{c}\times\underline{d}) + (\underline{a}\times d\underline{b})\cdot(\underline{c}\times\underline{d}) + \\ &\quad + (\underline{a}\times\underline{b})\cdot(d\underline{c}\times\underline{d}) + (\underline{a}\times\underline{b})\cdot(\underline{c}\times d\underline{d}) \\ &= -(\underline{c}\times\underline{d})\cdot(\underline{b}\times d\underline{a}) + (\underline{c}\times\underline{d})\cdot(\underline{a}\times d\underline{b}) + \\ &\quad -(\underline{a}\times\underline{b})\cdot(\underline{d}\times d\underline{c}) + (\underline{a}\times\underline{b})\cdot(\underline{c}\times d\underline{d}) \\ &= d\underline{a}\cdot(\underline{b}\times(\underline{c}\times\underline{d})) + d\underline{b}\cdot((\underline{c}\times\underline{d})\times\underline{a}) + \\ &\quad + d\underline{c}\cdot(\underline{d}\times(\underline{a}\times\underline{b})) + d\underline{d}\cdot((\underline{a}\times\underline{b})\times\underline{c}) \end{aligned} \quad , \qquad \text{(D.3.25)}$$

and the special case

$$\begin{aligned} d\left((\underline{a}\times\underline{b})\cdot(\underline{b}\times\underline{c})\right) = {} & d\underline{a}\cdot(\underline{b}\times(\underline{b}\times\underline{c})) + \\ & + d\underline{b}\cdot((\underline{b}\times\underline{c})\times\underline{a} + \underline{c}\times(\underline{a}\times\underline{b})) + \\ & + d\underline{c}\cdot((\underline{a}\times\underline{b})\times\underline{b}) \end{aligned} \quad . \qquad \text{(D.3.26)}$$

D.4 Force Derivations

D.4.1 Internal vs. External Coordinates

In the discussion of the OPLS-AA force field at the beginning of this Appendix, the potential function in Eq. D.2.1 was described as a sum of various interactions. A close examination of these interactions, given in Eqs. D.2.2–D.2.6, indicates that all the variables that affect the potential are "internal", implying that their values are independent of the simulation box frame-of-reference (i.e., the absolute coordinate system) selected. These variables include bond lengths (b_{ij}), bond angles (θ_{ijk}), bond dihedrals (ϕ_{ijkl}), and interatomic distances r_{ij}, for atoms in the system labeled i, j, k, and l. (I have added atom labels to b, θ, and ϕ at this stage for clarity.)

In contrast, the formulation of the force in Eq. D.1.1, as utilized during an actual MD simulation, is expressed in terms of the absolute coordinates of each atom (i.e.,

position vectors $\underline{r}_i$). For each of the OPLS-AA potential expressions, then, we require a conversion between internal and external coordinates.

D.4.2 Bond Vector Convention

In all the calculations that follow, I will continue to use the convention introduced at the beginning of this Appendix of representing atom positions by $\underline{r}$ subscripted with a single atom label (e.g., $\underline{r}_i$ for the position vector of atom i). We will find that there is a need for at most four atom labels, as indicated by the subscripting of the dihedral angle above, which we will consistently select to be i, j, k, and l. Vectors between atoms will be denoted using two labels, such as $\underline{r}_{ij}$, where a standard convention will be followed in which such a vector points *from* atom i *to* atom j. In other words,

$$\underline{r}_{ij} \equiv \underline{r}_j - \underline{r}_i, \tag{D.4.1}$$

and

$$\underline{r}_{ji} = \underline{r}_i - \underline{r}_j = -\underline{r}_{ij}. \tag{D.4.2}$$

The interatomic distances, r_{ij}, in Eqs. D.2.5 and D.2.6 are simply calculated as $|\underline{r}_{ij}|$, the norm of $\underline{r}_{ij}$.

Note that the GROMACS manual (as of Version 4.5.4) remains inconsistent on this point. Specifically, although their Equation 2.1 defines $\underline{r}_{ij}$ as we have, their Equation 4.13 has an incorrect sign for the Coulombic force.[190]

D.4.3 Bond Stretching Forces

The simplest bonded contribution to the system potential function is that due to bond stretching. In practice, bond lengths are often constrained in an MD data production run, using a constraint algorithm such as LINCS.[195] The purpose of such constraints is to remove fast oscillations from the system, thereby allowing an increase in the simulation time step without introducing numerical instabilities into the integration procedure.[190]

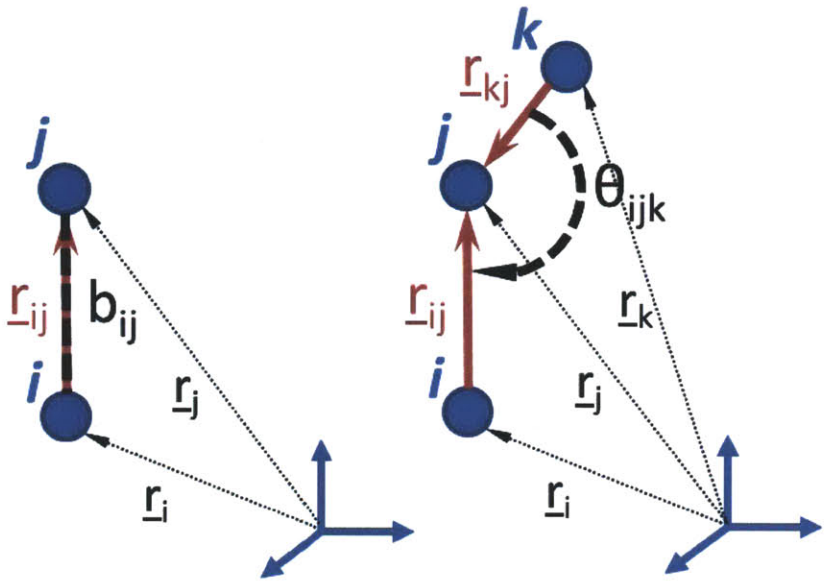


Figure D-1: Diagram illustrating the definition of the bond length, b_{ij}, and the bond angle, θ_{ijk} for consecutive atoms i, j, and (in the case of bond angle) k. Absolute position vectors (black dotted lines) relative to a fixed coordinate system (blue axes) and interatomic vectors (red solid lines) between atoms (blue filled circles) are illustrated for reference.

A typical bond length is pictured on the left in Fig. D-1. Denoted in Eq. D.2.2 as b, I will use b_{ij} here to indicate specifically that we are interested in the bond between atoms i and j. Likewise, the equilibrium bond length, b_0, will be labeled $b_{0,ij}$, and the force constant $k_{b,ij}$. In this notation, the potential contribution is given by

$$U_{bond(ij)} = \frac{1}{2} k_{b,ij} \left(b_{ij} - b_{0,ij}\right)^2 . \tag{D.4.3}$$

We now take the total derivative of Eq. D.4.3. This yields

$$dU_{bond(ij)} = k_{b,ij} \left(b_{ij} - b_{0,ij}\right) db_{ij} . \tag{D.4.4}$$

The bond length is calculated as

$$b_{ij} = \left|\underline{r}_{ij}\right|. \tag{D.4.5}$$

Therefore, its total derivative is given by

$$\begin{aligned} db_{ij} &= d\left|\underline{r}_{ij}\right| \\ &= d\underline{r}_{ij} \cdot \frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|}, \end{aligned} \tag{D.4.6}$$

where Eq. D.3.10 has been used. Substituting this result into Eq. D.4.4, we obtain

$$\begin{aligned} dU_{bond(ij)} &= k_{b,ij}\left(b_{ij} - b_{0,ij}\right) d\underline{r}_{ij} \cdot \frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|} \\ &= d\underline{r}_{i} \cdot \left(-k_{b,ij}\left(b_{ij} - b_{0,ij}\right)\frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|}\right) + \quad . \\ &\quad + d\underline{r}_{j} \cdot \left(+k_{b,ij}\left(b_{ij} - b_{0,ij}\right)\frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|}\right) \end{aligned} \tag{D.4.7}$$

Comparing Eq. D.4.7 to Eq. D.1.2, we extract the forces acting on each atom in bond ij. Specifically,

$$\underline{F}_{bond(ij),i} = +k_{b,ij}\left(b_{ij} - b_{0,ij}\right)\frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|}, \tag{D.4.8}$$

and

$$\underline{F}_{bond(ij),j} = -k_{b,ij}\left(b_{ij} - b_{0,ij}\right)\frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|}. \tag{D.4.9}$$

Clearly, $\underline{F}_{net} \equiv \underline{F}_{bond(ij),i} + \underline{F}_{bond(ij),j} = 0$, as desired, since these are internal forces.

D.4.4 Bond Angle Bending Forces

The next bonded contribution to the system potential function arises from angle bending. This is commonly modeled using a harmonic potential, although the actual potential could exhibit some asymmetry.

The bond angle is pictured on the right in Fig. D-1. Denoted in Eq. D.2.3 as θ, I will use θ_{ijk} here to indicate specifically that we are interested in atoms i, j, and k. Similarly, the equilibrium bond angle, θ_0, will be labeled $\theta_{0,ijk}$, and the force constant $k_{\theta,ijk}$. In this notation, the potential contribution is given by

$$U_{angle(ijk)} = \frac{1}{2} k_{\theta,ijk} \left(\theta_{ijk} - \theta_{0,ijk}\right)^2 . \tag{D.4.10}$$

Taking the total derivative, we obtain

$$dU_{angle(ijk)} = k_{\theta,ijk} \left(\theta_{ijk} - \theta_{0,ijk}\right) d\theta_{ijk}. \tag{D.4.11}$$

The bond angle is calculated by applying the dot product to the vectors $\underline{r}_{ij}$ and $\underline{r}_{kj}$, having a range of $0 \leq \theta_{ijk} \leq \pi$. Specifically,

$$\theta_{ijk} = \arccos\left(\frac{\underline{r}_{ij} \cdot \underline{r}_{kj}}{\left|\underline{r}_{ij}\right| \left|\underline{r}_{kj}\right|}\right). \tag{D.4.12}$$

Note the direction reversal of the second vector (i.e., $\underline{r}_{kj}$, rather than $\underline{r}_{jk}$): otherwise, we would be computing the bond angle supplement.

Letting

$$u \equiv \frac{\underline{r}_{ij} \cdot \underline{r}_{kj}}{\left|\underline{r}_{ij}\right| \left|\underline{r}_{kj}\right|} \tag{D.4.13}$$

in Eq. D.4.12, the total derivative can be found as follows:

$$\begin{aligned} d\theta_{ijk} &= d\arccos\left(u\right) \\ &= -\left(\frac{1}{\sqrt{1-u^2}}\right) du, \end{aligned} \tag{D.4.14}$$

where

$$\begin{aligned} du &= d\left(\frac{\underline{r}_{ij}\cdot\underline{r}_{kj}}{|\underline{r}_{ij}|\,|\underline{r}_{kj}|}\right) \\ &= \frac{1}{|\underline{r}_{ij}|\,|\underline{r}_{kj}|}\left[d\left(\underline{r}_{ij}\cdot\underline{r}_{kj}\right)-ud\left(|\underline{r}_{ij}|\,|\underline{r}_{kj}|\right)\right] \\ &= \frac{1}{|\underline{r}_{ij}|\,|\underline{r}_{kj}|}\left[\begin{array}{c}\left(d\underline{r}_{ij}\cdot\underline{r}_{kj}+d\underline{r}_{kj}\cdot\underline{r}_{ij}\right)+\\ -u\left(|\underline{r}_{kj}|\,d\,|\underline{r}_{ij}|+|\underline{r}_{ij}|\,d\,|\underline{r}_{kj}|\right)\end{array}\right]. \end{aligned} \tag{D.4.15}$$

Utilizing Eq. D.3.10, it follows that

$$\begin{aligned} |\underline{r}_{ij}|\,|\underline{r}_{kj}|\,du &= d\underline{r}_{ij}\cdot\underline{r}_{kj}+d\underline{r}_{kj}\cdot\underline{r}_{ij}+ \\ &\quad -u\left(d\underline{r}_{ij}\cdot|\underline{r}_{kj}|\,\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}+d\underline{r}_{kj}\cdot|\underline{r}_{ij}|\,\frac{\underline{r}_{kj}}{|\underline{r}_{kj}|}\right) \\ &= d\underline{r}_{ij}\cdot\left(\underline{r}_{kj}-u\,|\underline{r}_{kj}|\,\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}\right)+d\underline{r}_{kj}\cdot\left(\underline{r}_{ij}-u\,|\underline{r}_{ij}|\,\frac{\underline{r}_{kj}}{|\underline{r}_{kj}|}\right) \\ &= \left(d\underline{r}_{j}-d\underline{r}_{i}\right)\cdot\left(\underline{r}_{kj}-u\,|\underline{r}_{kj}|\,\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}\right)+ \\ &\quad +\left(d\underline{r}_{j}-d\underline{r}_{k}\right)\cdot\left(\underline{r}_{ij}-u\,|\underline{r}_{ij}|\,\frac{\underline{r}_{kj}}{|\underline{r}_{kj}|}\right) \\ &= d\underline{r}_{i}\cdot\left(u\,|\underline{r}_{kj}|\,\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}-\underline{r}_{kj}\right)+ \\ &\quad +d\underline{r}_{j}\cdot\left(\underline{r}_{kj}-u\,|\underline{r}_{kj}|\,\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}+\underline{r}_{ij}-u\,|\underline{r}_{ij}|\,\frac{\underline{r}_{kj}}{|\underline{r}_{kj}|}\right)+ \\ &\quad +d\underline{r}_{k}\cdot\left(u\,|\underline{r}_{ij}|\,\frac{\underline{r}_{kj}}{|\underline{r}_{kj}|}-\underline{r}_{ij}\right), \end{aligned} \tag{D.4.16}$$

and

$$dU_{angle(ijk)} = -\left(\frac{k_{\theta,ijk}\left(\theta_{ijk}-\theta_{0,ijk}\right)}{\sqrt{1-u^2}}\right)du. \tag{D.4.17}$$

Comparison of Eq. D.4.17 with Eq. D.1.2, including substitution of the various results above, yields the forces on each atom in bond angle ijk. Specifically,

$$\underline{F}_{angle(ijk),i} = +\left(\frac{k_{\theta,ijk}\left(\theta_{ijk}-\theta_{0,ijk}\right)}{\sqrt{1-u^2}}\right)\left(\frac{u\,|\underline{r}_{kj}|\,\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}-\underline{r}_{kj}}{|\underline{r}_{ij}|\,|\underline{r}_{kj}|}\right), \tag{D.4.18}$$

$$\underline{F}_{angle(ijk),j} = + \left(\frac{k_{\theta,ijk} \left(\theta_{ijk} - \theta_{0,ijk} \right)}{\sqrt{1-u^2}} \right) \left(\frac{\underline{r}_{kj} - u \left| \underline{r}_{kj} \right| \frac{\underline{r}_{ij}}{\left| \underline{r}_{ij} \right|} + \underline{r}_{ij} - u \left| \underline{r}_{ij} \right| \frac{\underline{r}_{kj}}{\left| \underline{r}_{kj} \right|}}{\left| \underline{r}_{ij} \right| \left| \underline{r}_{kj} \right|} \right), \tag{D.4.19}$$

and

$$\underline{F}_{angle(ijk),k} = + \left(\frac{k_{\theta,ijk} \left(\theta_{ijk} - \theta_{0,ijk} \right)}{\sqrt{1-u^2}} \right) \left(\frac{u \left| \underline{r}_{ij} \right| \frac{\underline{r}_{kj}}{\left| \underline{r}_{kj} \right|} - \underline{r}_{ij}}{\left| \underline{r}_{ij} \right| \left| \underline{r}_{kj} \right|} \right). \tag{D.4.20}$$

It can be readily seen that $\underline{F}_{net} \equiv \underline{F}_{angle(ijk),i} + \underline{F}_{angle(ijk),j} + \underline{F}_{angle(ijk),k} = 0$ holds, as required.

D.4.5 Bond Torsion Forces

The last bonded contribution that we are interested in is that due to bond torsion, represented by calculation of a dihedral angle. Specification of a dihedral angle requires four atoms, and therefore requires consideration of three interatomic vectors. Specifically,

$$\underline{r}_{ij} = d\underline{r}_j - d\underline{r}_i, \tag{D.4.21}$$

$$\underline{r}_{jk} = d\underline{r}_k - d\underline{r}_j, \tag{D.4.22}$$

and

$$\underline{r}_{kl} = d\underline{r}_l - d\underline{r}_k, \tag{D.4.23}$$

where the rotation is about the bond vector $\underline{r}_{jk}$. The planes containing $\left(\underline{r}_{ij}, \underline{r}_{jk}\right)$ and $\left(\underline{r}_{jk}, \underline{r}_{kl}\right)$ and sharing $\underline{r}_{jk}$ have normals

$$\underline{n}_1 = \underline{r}_{ij} \times \underline{r}_{jk}, \tag{D.4.24}$$

and

$$\underline{n}_2 = \underline{r}_{jk} \times \underline{r}_{kl}. \tag{D.4.25}$$

The dihedral angle is the signed angle between these two planes, ϕ_{ijkl}. The direction of positive dihedral angle is illustrated in Fig. D-2, which provides a few visual representations of the arrangements of atoms leading to three dihedral angles

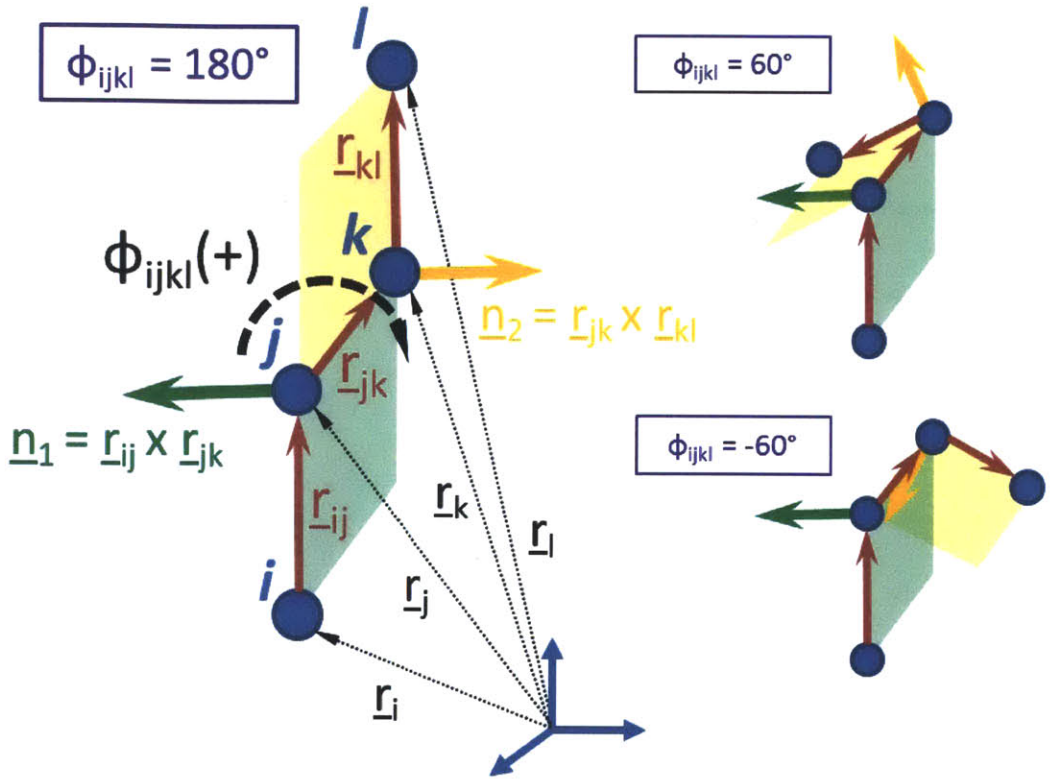


Figure D-2: Diagram illustrating the definition of the dihedral angle, ϕ_{ijkl}, for consecutive atoms i, j, k, and l, as the angle between the vectors normal to the planes comprising atoms i, j, and k (plane 1, with normal $\underline{n}_1$ – green) and atoms j, k, and l (plane 2, with normal $\underline{n}_2$ – orange). Absolute position vectors (black dotted lines) relative to a fixed coordinate system (blue axes) and interatomic vectors (red solid lines) between atoms (blue filled circles) are illustrated for reference.

commonly encountered in the torsional modeling of linear alkanes.[213] Briefly, if we construct a third normal, $\underline{n}_3 = \underline{r}_{jk} \times \underline{n}_1$ (not shown) to create a set of orthogonal axes $(\underline{r}_{jk}, \underline{n}_1, \underline{n}_3)$, rotation of vector $\underline{n}_2$ from the direction of $\underline{n}_1$, about $\underline{r}_{jk}$, and towards $\underline{n}_3$, yields a positive rotation.

Denoted in Eq. D.2.4 as ϕ, I will use ϕ_{ijkl} here to indicate specifically that we are interested in atoms i, j, k, and l. Similarly, the series coefficients C_n will be denoted $C_{n,ijkl}$. In this notation, the potential contribution (in Ryckaert-Bellemans form) is

given by

$$U_{dihedral(ijkl)} = \sum_{n=0}^{5} C_{n,ijkl} \cos^n \left(\phi_{ijkl} - \pi\right). \tag{D.4.26}$$

Taking the total derivative of Eq. D.4.26, we obtain

$$dU_{dihedral(ijkl)} = \left[\sum_{n=1}^{5} -C_{n,ijkl} n \cos^{n-1} \left(\phi_{ijkl} - \pi\right) \sin \left(\phi_{ijkl} - \pi\right)\right] d\phi_{ijkl}. \tag{D.4.27}$$

There are two geometric relationships that we can use to relate the dihedral angle to the normal vectors shown in Fig. D-2. Specifically,

$$\phi_{ijkl} = \arccos \frac{\underline{n}_1 \cdot \underline{n}_2}{|\underline{n}_1| \, |\underline{n}_2|}, \tag{D.4.28}$$

and

$$\phi_{ijkl} = \arcsin \frac{|\underline{n}_1 \times \underline{n}_2|}{|\underline{n}_1| \, |\underline{n}_2|}. \tag{D.4.29}$$

Recall, however, that the ranges of the inverse trigonometric functions limit the values that the calculated angle may obtain. Specifically, arccos allows an angle between 0 and π, while arcsin allows an angle between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. In order to obtain the signed dihedral that we are looking for, which may have a value from $-\pi$ to π, we instead make use of arctan as follows:

$$\begin{aligned} \phi_{ijkl} &= \arctan \left(\frac{\sin \phi_{ijkl}}{\cos \phi_{ijkl}}\right) \\ &= \arctan \left(\frac{|\underline{n}_1 \times \underline{n}_2|}{\underline{n}_1 \cdot \underline{n}_2}\right) \\ &= \arctan \left(\frac{y}{x}\right), \end{aligned} \tag{D.4.30}$$

where we have let $y \equiv |\underline{n}_1 \times \underline{n}_2|$ and $x \equiv \underline{n}_1 \cdot \underline{n}_2$. Although arctan as mathematically defined (with a single argument) has a range between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$, programmatic implementation as atan2(y, x) (with two arguments) yields the desired signed dihedral

ϕ_{ijkl}. The total differential is still that of arctan. That is,

$$\begin{aligned} d\phi_{ijkl} &= d\arctan\left(\frac{y}{x}\right) \\ &= \frac{1}{1+(y/x)^2} d\left(\frac{y}{x}\right) \\ &= \frac{x}{x^2+y^2}\left(dy - \frac{y}{x}dx\right). \end{aligned} \tag{D.4.31}$$

We must now compute dy and dx. Substituting the definitions of the normals into y and x, we obtain

$$\begin{aligned} y &= |\underline{n}_1 \times \underline{n}_2| \\ &= \left|\left(\underline{r}_{ij} \times \underline{r}_{jk}\right) \times \left(\underline{r}_{jk} \times \underline{r}_{kl}\right)\right| \\ &= \left|\underline{r}_{jk}\left(\left(\underline{r}_{ij} \times \underline{r}_{jk}\right) \cdot \underline{r}_{kl}\right)\right| \\ &= \left|\underline{r}_{jk}\right|\left(\left(\underline{r}_{ij} \times \underline{r}_{jk}\right) \cdot \underline{r}_{kl}\right) \\ &= \left|\underline{r}_{jk}\right| \underline{r}_{ij} \cdot \left(\underline{r}_{jk} \times \underline{r}_{kl}\right), \end{aligned} \tag{D.4.32}$$

and

$$\begin{aligned} x &= \underline{n}_1 \cdot \underline{n}_2 \\ &= \left(\underline{r}_{ij} \times \underline{r}_{jk}\right) \cdot \left(\underline{r}_{jk} \times \underline{r}_{kl}\right). \end{aligned} \tag{D.4.33}$$

We now take the total differentials, starting with y, which yields

$$dy = d\left(\left|\underline{r}_{jk}\right| \underline{r}_{ij}\right) \cdot \left(\underline{r}_{jk} \times \underline{r}_{kl}\right) + \left(\left|\underline{r}_{jk}\right| \underline{r}_{ij}\right) \cdot d\left(\underline{r}_{jk} \times \underline{r}_{kl}\right). \tag{D.4.34}$$

Proceeding term-by-term, the first term in Eq. D.4.34 yields

$$\begin{aligned}
d\left(\left|\underline{r}_{jk}\right|\underline{r}_{ij}\right)\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right) &= \left(\underline{r}_{ij}d\left|\underline{r}_{jk}\right|+\left|\underline{r}_{jk}\right|d\underline{r}_{ij}\right)\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\\
&= \underline{r}_{ij}d\left|\underline{r}_{jk}\right|\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)+d\underline{r}_{ij}\cdot\left|\underline{r}_{jk}\right|\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\\
&= \left(d\underline{r}_{jk}\cdot\frac{\underline{r}_{jk}}{\left|\underline{r}_{jk}\right|}\right)\left(\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)+d\underline{r}_{ij}\cdot\left|\underline{r}_{jk}\right|\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\\
&= d\underline{r}_{jk}\cdot\left(\frac{\underline{r}_{jk}}{\left|\underline{r}_{jk}\right|}\left(\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)\right)+d\underline{r}_{ij}\cdot\left|\underline{r}_{jk}\right|\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\\
&= d\underline{r}_{i}\cdot\left(-\left|\underline{r}_{jk}\right|\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)+ \qquad \text{(D.4.35)}\\
&\quad +d\underline{r}_{j}\cdot\left(\left|\underline{r}_{jk}\right|\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)-\tfrac{\underline{r}_{jk}}{\left|\underline{r}_{jk}\right|}\left(\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)\right)+\\
&\quad +d\underline{r}_{k}\cdot\left(\tfrac{\underline{r}_{jk}}{\left|\underline{r}_{jk}\right|}\left(\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)\right),
\end{aligned}$$

and the second term in Eq. D.4.34 yields

$$\begin{aligned}
\left(\left|\underline{r}_{jk}\right|\underline{r}_{ij}\right)\cdot d\left(\underline{r}_{jk}\times\underline{r}_{kl}\right) &= \left(\left|\underline{r}_{jk}\right|\underline{r}_{ij}\right)\cdot\left(\left(d\underline{r}_{jk}\times\underline{r}_{kl}\right)+\left(\underline{r}_{jk}\times d\underline{r}_{kl}\right)\right)\\
&= \left(-\left(\left|\underline{r}_{jk}\right|\underline{r}_{ij}\right)\cdot\left(\underline{r}_{kl}\times d\underline{r}_{jk}\right)+\left(\left|\underline{r}_{jk}\right|\underline{r}_{ij}\right)\cdot\left(\underline{r}_{jk}\times d\underline{r}_{kl}\right)\right)\\
&= d\underline{r}_{kl}\cdot\left(\left|\underline{r}_{jk}\right|\underline{r}_{ij}\times\underline{r}_{jk}\right)-d\underline{r}_{jk}\cdot\left(\left|\underline{r}_{jk}\right|\underline{r}_{ij}\times\underline{r}_{kl}\right)\\
&= d\underline{r}_{j}\cdot\left(\left|\underline{r}_{jk}\right|\underline{r}_{ij}\times\underline{r}_{kl}\right)+ \qquad \text{(D.4.36)}\\
&\quad +d\underline{r}_{k}\cdot\left(\underline{r}_{jk}\times\left|\underline{r}_{jk}\right|\underline{r}_{ij}+\underline{r}_{kl}\times\left|\underline{r}_{jk}\right|\underline{r}_{ij}\right)+\\
&\quad +d\underline{r}_{l}\cdot\left(\left|\underline{r}_{jk}\right|\underline{r}_{ij}\times\underline{r}_{jk}\right).
\end{aligned}$$

Combining these results, we obtain

$$\begin{aligned}
dy = \; & d\underline{r}_{i}\cdot\left(-\left|\underline{r}_{jk}\right|\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)+\\
& +d\underline{r}_{j}\cdot\left(\left|\underline{r}_{jk}\right|\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)-\tfrac{\underline{r}_{jk}}{\left|\underline{r}_{jk}\right|}\left(\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)+\left|\underline{r}_{jk}\right|\underline{r}_{ij}\times\underline{r}_{kl}\right)+\\
& +d\underline{r}_{k}\cdot\left(\tfrac{\underline{r}_{jk}}{\left|\underline{r}_{jk}\right|}\left(\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)+\underline{r}_{jk}\times\left|\underline{r}_{jk}\right|\underline{r}_{ij}+\underline{r}_{kl}\times\left|\underline{r}_{jk}\right|\underline{r}_{ij}\right)+\\
& +d\underline{r}_{l}\cdot\left(\left|\underline{r}_{jk}\right|\underline{r}_{ij}\times\underline{r}_{jk}\right).
\end{aligned} \qquad \text{(D.4.37)}$$

The total differential for x is given by

$$\begin{aligned} dx &= d\left(\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right) \\ &= d\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)+ \\ &\quad +d\underline{r}_{jk}\cdot\left(\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\times\underline{r}_{ij}+\underline{r}_{kl}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right)+ \\ &\quad +d\underline{r}_{kl}\cdot\left(\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\times\underline{r}_{jk}\right), \end{aligned} \tag{D.4.38}$$

which, after substituting the definitions of the interatomic vectors, yields

$$\begin{aligned} dx = \; & d\underline{r}_i\cdot\left(-\underline{r}_{jk}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)+ \\ & +d\underline{r}_j\cdot\left(\underline{r}_{jk}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)+\underline{r}_{ij}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)-\underline{r}_{kl}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right)+ \\ & +d\underline{r}_k\cdot\left(-\underline{r}_{ij}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)+\underline{r}_{kl}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)+\underline{r}_{jk}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right)+ \\ & +d\underline{r}_l\cdot\left(-\underline{r}_{jk}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right). \end{aligned} \tag{D.4.39}$$

Returning to Eq. D.4.31, and substituting in Eqs. D.4.37 and D.4.39, we find

$$\begin{aligned} & \frac{x^2+y^2}{x}d\phi_{ijkl} \\ = \; & d\underline{r}_i\cdot\left(-\left|\underline{r}_{jk}\right|\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)+\frac{y}{x}\left(\underline{r}_{jk}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)\right)+ \\ & +d\underline{r}_j\cdot\left(\begin{array}{c} -\frac{\underline{r}_{jk}}{\left|\underline{r}_{jk}\right|}\left(\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)+\left|\underline{r}_{jk}\right|\underline{r}_{jk}\times\underline{r}_{kl}+\left|\underline{r}_{jk}\right|\underline{r}_{ij}\times\underline{r}_{kl}+ \\ -\frac{y}{x}\left(\left(\underline{r}_{jk}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)+\left(\underline{r}_{ij}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)-\left(\underline{r}_{kl}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right)\right) \end{array}\right)+ \\ & +d\underline{r}_k\cdot\left(\begin{array}{c} \frac{\underline{r}_{jk}}{\left|\underline{r}_{jk}\right|}\left(\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)-\left|\underline{r}_{jk}\right|\underline{r}_{ij}\times\underline{r}_{jk}-\left|\underline{r}_{jk}\right|\underline{r}_{ij}\times\underline{r}_{kl}+ \\ -\frac{y}{x}\left(\left(\underline{r}_{jk}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right)+\left(\underline{r}_{kl}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right)-\left(\underline{r}_{ij}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)\right) \end{array}\right)+ \\ & +d\underline{r}_l\cdot\left(\left|\underline{r}_{jk}\right|\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)+\frac{y}{x}\left(\underline{r}_{jk}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right)\right). \end{aligned} \tag{D.4.40}$$

Finally, substituting this result into Eq. D.4.27, and comparing with Eq. D.1.2 yields

$$\underline{F}_{dihedral(ijkl),i}=-\frac{C_{ijkl}}{x^2+y^2}\left(x\left(-\left|\underline{r}_{jk}\right|\underline{r}_{jk}\times\underline{r}_{kl}\right)+y\left(\underline{r}_{jk}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)\right), \tag{D.4.41}$$

$$\underline{F}_{dihedral(ijkl),j} \tag{D.4.42}$$
$$= -\frac{C_{ijkl}}{x^2+y^2}\left(\begin{array}{c} x\left(-\frac{\underline{r}_{jk}}{|\underline{r}_{jk}|}\left(\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)+|\underline{r}_{jk}|\,\underline{r}_{jk}\times\underline{r}_{kl}+|\underline{r}_{jk}|\,\underline{r}_{ij}\times\underline{r}_{kl}\right)+ \\ +y\left(-\underline{r}_{jk}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)-\underline{r}_{ij}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)+\underline{r}_{kl}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right) \end{array}\right),$$

$$\underline{F}_{dihedral(ijkl),k} \tag{D.4.43}$$
$$= -\frac{C_{ijkl}}{x^2+y^2}\left(\begin{array}{c} x\left(+\frac{\underline{r}_{jk}}{|\underline{r}_{jk}|}\left(\underline{r}_{ij}\cdot\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)\right)-|\underline{r}_{jk}|\,\underline{r}_{ij}\times\underline{r}_{jk}-|\underline{r}_{jk}|\,\underline{r}_{ij}\times\underline{r}_{kl}\right)+ \\ +y\left(-\underline{r}_{jk}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)+\underline{r}_{ij}\times\left(\underline{r}_{jk}\times\underline{r}_{kl}\right)-\underline{r}_{kl}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right) \end{array}\right),$$

and

$$\underline{F}_{dihedral(ijkl),l} = -\frac{C_{ijkl}}{x^2+y^2}\left(x\left(|\underline{r}_{jk}|\,\underline{r}_{ij}\times\underline{r}_{jk}\right)+y\left(\underline{r}_{jk}\times\left(\underline{r}_{ij}\times\underline{r}_{jk}\right)\right)\right), \tag{D.4.44}$$

where C_{ijkl} is a function of ϕ_{ijkl} given by

$$C_{ijkl}\left(\phi_{ijkl}\right) \equiv \left[\sum_{n=1}^{5} -C_{n,ijkl} n \cos^{n-1}\left(\phi_{ijkl}-\pi\right)\sin\left(\phi_{ijkl}-\pi\right)\right]. \tag{D.4.45}$$

From inspection, $\underline{F}_{net} \equiv \underline{F}_{dihedral(ijkl),i}+\underline{F}_{dihedral(ijkl),j}+\underline{F}_{dihedral(ijkl),k}+\underline{F}_{dihedral(ijkl),l} = 0$, as required.

D.4.6 Improper Dihedral Forces

Improper dihedrals are a special class of dihedral functions used to maintain the planarity of atoms. Their contribution to the system potential function follows a harmonic potential form:

$$U_{dihedral(ijkl)} = \frac{1}{2}k_{\xi,ijkl}\left(\xi_{ijkl}-\xi_{0,ijkl}\right)^2, \tag{D.4.46}$$

where ξ_{ijkl} indicates the dihedral angle for atoms $ijkl$, computed precisely as in the previous section. In this case, the total derivative of Eq. D.4.46 is given by

$$dU_{dihedral(ijkl)} = k_{\xi,ijkl}\left(\xi_{ijkl}-\xi_{0,ijkl}\right)d\xi_{ijkl}, \tag{D.4.47}$$

and the forces follow Eqs. D.4.41–D.4.44, if C_{ijkl} is defined as follows:

$$C_{ijkl}\left(\xi_{ijkl}\right) \equiv \left[k_{\xi,ijkl}\left(\xi_{ijkl} - \xi_{0,ijkl}\right)\right]. \tag{D.4.48}$$

D.4.7 Coulombic Interaction Forces

In Program BUILDER, the electrostatic contribution to the system potential function is computed directly from pairwise computation of Coulombic interactions as follows:

$$U_{Coulombic(ij)}\left(\left|\underline{r}_{ij}\right|\right) = \frac{f_{ij}}{4\pi\varepsilon_0}\frac{q_i q_j}{\varepsilon_r}\frac{1}{\left|\underline{r}_{ij}\right|}, \tag{D.4.49}$$

where f_{ij} is defined as in Eq. D.2.7, $(4\pi\varepsilon_0)^{-1}$ has a value of 138.935485(9) $kJ\,mol^{-1}\,nm\,e^{-2}$ (where the unit e is the electronic charge), and the relative permittivity ε_r is set to 1 in explicit solvent simulations (i.e., the value for interactions in vacuum).[190] To simulate small molecules in vacuum using Langevin Dynamics to represent solvent collisions, Eq. D.4.49 can be readily evaluated for all pairs. However, for larger systems, such as micelles, the need to limit computational costs leads to the use of a simple cutoff and neighbor list to determine which pairs of atoms should be evaluated at a given time step during an MD run. Since the goal of Program BUILDER is to create initial configurations for production software such as GROMACS, no further corrections need to be applied. Note that GROMACS itself implements more advanced techniques for handling long-range electrostatic interactions, including particle-mesh Ewald (PME) summation.[196,197] The force expressions using this technique are beyond the scope of this Appendix, but the interested reader can consult the paper by Deserno and Holm[198] for more information.

Taking the total derivative of Eq. D.4.49, we obtain

$$\begin{aligned} dU_{Coulombic(ij)} &= -\frac{f_{ij}}{4\pi\varepsilon_0}\frac{q_i q_j}{\varepsilon_r}\frac{1}{\left|\underline{r}_{ij}\right|^2} d\left|\underline{r}_{ij}\right| \\ &= -\frac{f_{ij}}{4\pi\varepsilon_0}\frac{q_i q_j}{\varepsilon_r}\frac{1}{\left|\underline{r}_{ij}\right|^2}\left(d\underline{r}_{ij}\cdot\frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|}\right) \\ &= \begin{array}{l} d\underline{r}_i\cdot\left(\frac{f_{ij}}{4\pi\varepsilon_0}\frac{q_i q_j}{\varepsilon_r}\frac{1}{\left|\underline{r}_{ij}\right|^2}\frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|}\right)+ \\ +d\underline{r}_j\cdot\left(-\frac{f_{ij}}{4\pi\varepsilon_0}\frac{q_i q_j}{\varepsilon_r}\frac{1}{\left|\underline{r}_{ij}\right|^2}\frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|}\right). \end{array} \end{aligned} \tag{D.4.50}$$

Comparison with Eq. D.1.2 yields

$$\underline{F}_{Coulombic(ij),i} = -\frac{f_{ij}}{4\pi\varepsilon_0}\frac{q_i q_j}{\varepsilon_r}\frac{1}{\left|\underline{r}_{ij}\right|^2}\frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|}, \tag{D.4.51}$$

and

$$\underline{F}_{Coulombic(ij),j} = +\frac{f_{ij}}{4\pi\varepsilon_0}\frac{q_i q_j}{\varepsilon_r}\frac{1}{\left|\underline{r}_{ij}\right|^2}\frac{\underline{r}_{ij}}{\left|\underline{r}_{ij}\right|}, \tag{D.4.52}$$

where we can clearly see that $\underline{F}_{net} \equiv \underline{F}_{Coulombic(ij),i} + \underline{F}_{Coulombic(ij),j} = 0$, as required.

D.4.8 Lennard-Jones Interaction Forces

The van der Waals contribution to the system potential energy is computed in Program BUILDER, similarly to the Coulombic contribution. In other words, for small molecules, all pairwise contributions may be computed, while larger systems require use of a cutoff to reduce computational expense. Note that GROMACS employs a similar cutoff scheme, but provides the option to calculate long-range dispersion corrections to the energy and pressure. In this Thesis, such corrections are always used only during production runs.

The van der Waals contribution here takes a Lennard-Jones form. That is,

$$U_{LJ(ij)} = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{\left|\underline{r}_{ij}\right|}\right)^{12} - \left(\frac{\sigma_{ij}}{\left|\underline{r}_{ij}\right|}\right)^{6}\right). \tag{D.4.53}$$

Taking the total differential of Eq. D.4.53 yields

$$\begin{aligned}
dU_{LJ(ij)} &= 4\varepsilon_{ij}\left(12\left(\frac{\sigma_{ij}}{|\underline{r}_{ij}|}\right)^{11} d\left(\frac{\sigma_{ij}}{|\underline{r}_{ij}|}\right) - 6\left(\frac{\sigma_{ij}}{|\underline{r}_{ij}|}\right)^{5} d\left(\frac{\sigma_{ij}}{|\underline{r}_{ij}|}\right)\right) \\
&= 4\varepsilon_{ij}\left(-12\frac{\sigma_{ij}^{12}}{|\underline{r}_{ij}|^{13}} + 6\frac{\sigma_{ij}^{6}}{|\underline{r}_{ij}|^{7}}\right) d\left(|\underline{r}_{ij}|\right) \\
&= 4\varepsilon_{ij}\left(-12\frac{\sigma_{ij}^{12}}{|\underline{r}_{ij}|^{13}} + 6\frac{\sigma_{ij}^{6}}{|\underline{r}_{ij}|^{7}}\right)\left(d\underline{r}_{ij}\cdot\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}\right) \\
&= d\underline{r}_{ij}\cdot\left(4\varepsilon_{ij}\left(-12\frac{\sigma_{ij}^{12}}{|\underline{r}_{ij}|^{13}} + 6\frac{\sigma_{ij}^{6}}{|\underline{r}_{ij}|^{7}}\right)\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}\right) \\
&= d\underline{r}_{i}\cdot\left(4\varepsilon_{ij}\left(12\frac{\sigma_{ij}^{12}}{|\underline{r}_{ij}|^{13}} - 6\frac{\sigma_{ij}^{6}}{|\underline{r}_{ij}|^{7}}\right)\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}\right) + \\
&\quad + d\underline{r}_{j}\cdot\left(-4\varepsilon_{ij}\left(12\frac{\sigma_{ij}^{12}}{|\underline{r}_{ij}|^{13}} - 6\frac{\sigma_{ij}^{6}}{|\underline{r}_{ij}|^{7}}\right)\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}\right).
\end{aligned} \tag{D.4.54}$$

Comparison of Eq. D.4.54 with Eq. D.1.2 yields

$$\underline{F}_{LJ(ij),i} = -4\varepsilon_{ij}\left(12\frac{\sigma_{ij}^{12}}{|\underline{r}_{ij}|^{13}} - 6\frac{\sigma_{ij}^{6}}{|\underline{r}_{ij}|^{7}}\right)\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}, \tag{D.4.55}$$

and

$$\underline{F}_{LJ(ij),j} = +4\varepsilon_{ij}\left(12\frac{\sigma_{ij}^{12}}{|\underline{r}_{ij}|^{13}} - 6\frac{\sigma_{ij}^{6}}{|\underline{r}_{ij}|^{7}}\right)\frac{\underline{r}_{ij}}{|\underline{r}_{ij}|}, \tag{D.4.56}$$

where $\underline{F}_{net} \equiv \underline{F}_{LJ(ij),i} + \underline{F}_{LJ(ij),j} = 0$, as required.

D.4.9 Damped Mass-Spring Forces

A final force is implemented in Program BUILDER for use in generating micelle configurations. This force is non-conservative and is not used during production runs in GROMACS. The force in question corresponds to the damped mass-spring interaction, which consists of a conservative mass-spring force and a non-conservative damping force, which together provide some constraint on the motion of an atom to a region of space. The atom can be viewed as attached via a spring to any specified

point, p, such as a micelle center-of-mass. This damped force is expressed as follows:[167]

$$\underline{F}_{spring(ip)} = k_{s,ip}\left(\left|\underline{r}_{ip}\right| - l_{0,ip}\right)\frac{\underline{r}_{ip}}{\left|\underline{r}_{ip}\right|} + k_{d,ip}\underline{v}_{ip}, \tag{D.4.57}$$

where $\underline{r}_{ip} = \underline{r}_p - \underline{r}_i$ is the vector pointing from atom i to the point p, $\underline{v}_{ip} = \underline{v}_p - \underline{v}_i$ is a similarly defined relative velocity vector, $l_{0,ip}$ is the spring rest length, and $k_{s,ip}$ and $k_{d,ip}$ are the spring constant and damping coefficient, respectively.

The natural frequency of the (undamped) spring is given by[167]

$$\omega_{0,ip} = \sqrt{\frac{k_{s,ip}}{\widehat{m}_i}}, \tag{D.4.58}$$

where $\widehat{m}_i$ is the mass of atom i (see below for a discussion of units). The damping ratio is given by

$$\zeta_{ip} = \frac{k_{d,ip}}{2\widehat{m}_i\omega_{0,ip}}. \tag{D.4.59}$$

Critical damping occurs when $\zeta_{ip} = 1$, which provides the fastest return to equilibrium without oscillations. Recall that underdamping ($\zeta_{ip} < 1$) will still result in oscillations, while overdamping ($\zeta_{ip} > 1$) will more slowly return the system to equilibrium but will not yield oscillations. For critical damping,[167]

$$\zeta_{ip} = 1 = \frac{k_{d,ip}}{2\widehat{m}_i\omega_{0,ip}} = \frac{k_{d,ip}}{2\sqrt{\widehat{m}_i k_{s,ip}}}. \tag{D.4.60}$$

In this case,

$$k_{d,ip} = 2\sqrt{\widehat{m}_i k_{s,ip}}. \tag{D.4.61}$$

At this stage, it is useful to examine the units of the various terms in Eq D.4.57:

$$\left[\underline{F}_{spring(ip)} \text{ in } \frac{kJ}{mol\cdot nm},\ l_{0,ip} \text{ in } nm\right] \Rightarrow k_{s,ip} \text{ in } \frac{kJ}{mol\cdot nm^2}, \tag{D.4.62}$$

and

$$\left[\underline{F}_{spring(ip)} \text{ in } \frac{kJ}{mol\cdot nm},\ \underline{v}_{ip} \text{ in } \frac{nm}{ps}\right] \Rightarrow k_{d,ip} \text{ in } \frac{kJ\cdot ps}{mol\cdot nm^2}. \tag{D.4.63}$$

From the units of Eq. D.4.61, it follows that

$$\frac{kJ \cdot ps}{mol \cdot nm^2} [=] \sqrt{[\widehat{m}_i] \frac{kJ}{mol \cdot nm^2}}. \tag{D.4.64}$$

Rearranging Eq. D.4.64 yields

$$[\widehat{m}_i] [=] \frac{kJ \cdot ps^2}{mol \cdot nm^2}. \tag{D.4.65}$$

Equation D.4.65 provides us with the units of the mass term $\widehat{m}_i$:

$$\widehat{m}_i = \widehat{m}_i \left[\frac{kJ \cdot ps^2}{mol \cdot nm^2} \right]. \tag{D.4.66}$$

We now wish to relate $\widehat{m}_i$ to the atomic mass, m_i, which has units of g/mol. Specifically,

$$\begin{aligned} m_i \left[\frac{g}{mol} \right] &= \widehat{m}_i \left[\tfrac{kJ \cdot ps^2}{mol \cdot nm^2} \right] \times 10^3 \left[\tfrac{J}{kJ} \right] \times 1 \left[\tfrac{kg \cdot m^2/s^2}{J} \right] \times \\ &\quad \times 10^3 \left[\tfrac{g}{kg} \right] \times \left(10^{-12} \left[\tfrac{s}{ps} \right] \right)^2 \times \left(10^9 \left[\tfrac{nm}{m} \right] \right)^2 \\ &= \widehat{m}_i \left[\frac{kJ \cdot ps^2}{mol \cdot nm^2} \right] \times 1 \left[\left(\frac{g}{mol} \right) / \left(\frac{kJ \cdot ps^2}{mol \cdot nm^2} \right) \right]. \end{aligned} \tag{D.4.67}$$

Therefore, in terms of magnitudes, it follows that

$$\widehat{m}_i = m_i, \tag{D.4.68}$$

which implies that we may select our damping coefficient as

$$k_{i0}^d = 2\sqrt{m_i k_{i0}^s} \tag{D.4.69}$$

to achieve critical damping of our spring.

D.5 Converting Between Torsional Energy Series Types

The dihedral (bond torsion) contribution to the system potential energy in the OPLS-AA force field is cast as a Fourier series,[188] typically consisting of four coefficients:[318]

$$U(\phi) = \sum_{n=1}^{4} \frac{V_n}{2} \left[1 + (-1)^{n+1} \cos(n\phi)\right]. \tag{D.5.1}$$

Expanding Eq. D.5.1 yields

$$U(\phi) = \frac{V_1}{2}(1 + \cos(\phi)) + \frac{V_2}{2}(1 - \cos(2\phi)) + \frac{V_3}{2}(1 + \cos(3\phi)) + \frac{V_4}{2}(1 - \cos(4\phi)). \tag{D.5.2}$$

In contrast, in GROMACS, a Ryckaert-Bellemans expression[316] is used, as follows:

$$\begin{aligned} U(\phi) &= \sum_{n=0}^{5} C_n \cos^n(\phi - \pi) \\ &= \sum_{n=0}^{5} (-1)^n C_n \cos^n \phi \end{aligned} \tag{D.5.3}$$

Expanding Eq. D.5.3 yields

$$\begin{aligned} U(\phi) &= C_0 - C_1 \cos\phi + C_2 \cos^2\phi - C_3 \cos^3\phi + C_4 \cos^4\phi - C_5 \cos^5\phi \\ &= 1[C_0] + \cos\phi[-C_1] + \cos^2\phi[C_2] + \cos^3\phi[-C_3] + \\ &\quad + \cos^4\phi[C_4] + \cos^5\phi[-C_5]. \end{aligned} \tag{D.5.4}$$

In order to convert between these two expressions, we make use of the following useful trigonometric identities:

$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta, \tag{D.5.5}$$

$$\cos(2\phi) = 2\cos^2\phi - 1, \tag{D.5.6}$$

$$\sin(2\phi) = 2\sin\phi\cos\phi, \tag{D.5.7}$$

$$\cos^2\phi + \sin^2\phi = 1, \tag{D.5.8}$$

$$\begin{aligned}
\cos(3\phi) &= \cos(2\phi+\phi) \\
&= \cos(2\phi)\cos\phi - \sin(2\phi)\sin\phi \\
&= 2\cos^3\phi - \cos\phi - 2\sin^2\phi\cos\phi \\
&= 2\cos^3\phi - \cos\phi - 2\left(1-\cos^2\phi\right)\cos\phi \\
&= 4\cos^3\phi - 3\cos\phi,
\end{aligned} \tag{D.5.9}$$

and

$$\begin{aligned}
\cos(4\phi) &= \cos(2(2\phi)) \\
&= 2\cos^2(2\phi) - 1 \\
&= 2\left(2\cos^2\phi - 1\right)^2 - 1 \\
&= 8\cos^4\phi - 8\cos^2\phi + 1.
\end{aligned} \tag{D.5.10}$$

Returning to Eq. D.5.2, and utilizing Eqs. D.5.5-D.5.10 yields

$$\begin{aligned}
U(\phi) &= \tfrac{V_1}{2}(1+\cos(\phi)) + \tfrac{V_2}{2}(1-[2\cos^2\phi - 1]) + \\
&\quad + \tfrac{V_3}{2}(1+[4\cos^3\phi - 3\cos\phi]) + \tfrac{V_4}{2}(1-[8\cos^4\phi - 8\cos^2\phi + 1]) \\
&= 1\left[\tfrac{V_1}{2} + V_2 + \tfrac{V_3}{2}\right] + \cos(\phi)\left[\tfrac{V_1}{2} - \tfrac{3V_3}{2}\right] + \\
&\quad + \cos^2\phi\left[-V_2 + 4V_4\right] + \\
&\quad + \cos^3\phi\left[2V_3\right] + \cos^4\phi\left[-4V_4\right].
\end{aligned} \tag{D.5.11}$$

We can now compare Eq. D.5.11 to Eq. D.5.4 to obtain the coefficient relation-

ships. Specifically,

$$C_0 = V_2 + \frac{1}{2}(V_1 + V_3) \quad \text{(D.5.12a)}$$

$$C_1 = \frac{1}{2}(-V_1 + 3V_3) \quad \text{(D.5.12b)}$$

$$C_2 = -V_2 + 4V_4 \quad \text{(D.5.12c)}$$

$$C_3 = -2V_3 \quad \text{(D.5.12d)}$$

$$C_4 = -4V_4 \quad \text{(D.5.12e)}$$

$$C_5 = 0, \quad \text{(D.5.12f)}$$

which are also available in the GROMACS manual.[190]

The inversion of the above relations is also useful when comparing GROMACS values to literature values. Specifically,

$$V_1 = -\frac{3}{2}C_3 - 2C_1 \quad \text{(D.5.13a)}$$

$$V_2 = -C_2 - C_4 \quad \text{(D.5.13b)}$$

$$V_3 = -\frac{C_3}{2} \quad \text{(D.5.13c)}$$

$$V_4 = -\frac{C_4}{4}. \quad \text{(D.5.13d)}$$

Appendix E

Development of Force-Field Parameters

E.1 Overview

In Section D.2 of Appendix D, I introduced the OPLS-AA force field, developed by Jorgensen and coworkers[188] to model the behavior of organic molecules in liquid systems. Application of such a force field to a particular solution containing surfactant, solvent, and ions requires the specification of a set of parameters for: (i) all intramolecular, bonded interactions, including specifying equilibrium bond lengths and force constants (bond stretching), equilibrium bond angles and force constants (angle bending), and torsion potential series coefficients and phase angles (bond torsion), and (ii) all interatomic, non-bonded interactions, including specifying atomic partial charges (Coulombic interactions) and Lennard-Jones parameters (van der Waals interactions). (Note that, in the OPLS-AA model, atoms participating in the same bonds or bond angles are disqualified from non-bonded interactions with each other, and atoms participating in the same bond torsion are assigned a reduced – by one half – non-bonded interaction with each other.[188])

The values for all these parameters are determined from either ab initio calculations[98,191] or fitting to reproduce experimentally-measured properties, which include enthalpies of vaporization, bulk liquid densities, heat capacities, and isothermal com-

pressibilities.[188] In general, these parameters are found to be dependent not only on the elemental nature of the atoms participating in a given interaction, but also on their orbital hybridization and specific nearest neighbors (i.e., their chemical functionality). In fact, the parameters need not even be constants in principle – partial charges, which are often calculated using an electrostatic surface potential fit to ab initio data, can vary with molecule configuration[194]!

To reduce the number of parameters that need to be calculated or fit, a number of simplifying assumptions are made. For example, the partial charges mentioned above are usually set to average values based on the most likely configurations of a molecule (or even just of the geometry-minimized state).[194] Also, bonded interactions are typically considered on the basis of the atom types participating in the interaction, where a particular atom type represents an atom of a particular element and orbital hybridization connected to one or more of a subset of possible neighboring atoms. As a concrete example, OPLS-AA shares the atom types of the AMBER force field,[98] and the atom type CT represents any sp^3 (tetrahedral) carbon. Therefore, CT may represent the carbon in CH_4, RCH_3, R_2CH_2, R_3CH, or R_4C, regardless of R (a generic functional group). Of course, the particular nearest neighbors to CT in R (more precisely, their atom types) must still be considered when determining specific bonds, angles, and torsions involving these groups.[188]

Many various functional groups are represented through the existing atom types and specification of bonded and non-bonded parameters found in the original OPLS-AA force field.[188] In addition, other useful functional groups and related parameters have been added since (including carbohydrates,[238] heterocycles,[319] amines,[320] esters, nitriles, and nitro-containing compounds,[321] and even fluorinated alkanes[322]). Still, there were gaps in the force field for a number of years which corresponded to the chemical functionality encountered in certain common surfactant head groups, especially those involving non-zero net charges or dipoles. (This includes ionic and zwitterionic surfactants, containing sulfates, sulfonates, tetraalkyl ammoniums, and betaines.) Furthermore, due to the large number of combinations of atom types that must be considered when constructing a library of torsional parameters (recall

that these are four-body interactions), even common existing functional groups in the forcefield (e.g., ethers and alcohols) are not always fully specified when present in close proximity on a molecule. This is an issue, for example, with the terminus of ethoxylate head groups, where the ether oxygen (*OS* atom type) participates in a bond torsion with the terminal alcohol oxygen (*OH* atom type) with two *CT* carbon intermediaries. The resulting *OS-CT-CT-OH* dihedral is not present in the OPLS-AA force field (except as an entry for carbohydrates, in the context of a cyclic ring).

Fortunately, the development of new force-field parameters is a systematic process,[191] particularly for bonded interactions, where ab initio techniques are most often used.[188] There has also been recent interest in the literature in developing new force-field parameters for ionic liquids, which have much in common with charged surfactants, often possessing similar functionality and even alkyl attachments to the charged moieties. In this thesis, I make extensive use of the force-field parameters developed by Canongia Lopes and coworkers for tetraalkylammonium ions,[192] pyridinium ions,[193] and sulfate and sulfonate ions,[194] which were validated against experimental data for ionic liquids. These parameters have been most useful for the modeling of trimethylammonium, pyridinium, sulfate, sulfonate, and sulfobetaine surfactants. Furthermore, these parameters were developed specifically to be consistent with the OPLS-AA force field, by using similar procedures for deriving the parameters, and using existing OPLS-AA parameters to describe interactions involving alkyl groups which are distant from the functional group of interest (i.e., two or three carbons, or more, away).

Even with this recently expanded set of parameters, there are still cases that require new force-field parameter development, within the context of OPLS-AA. This includes the important linear alkylbenzene sulfonate class of surfactants. Here, the connectivity of the sulfonate directly to the benzene ring precludes use of the sulfonate parameters already available,[194] which were designed only to be attached to linear alkanes, not benzene rings. Note that linear alkylbenzene sulfonate surfactants with branching at the 4-position have indeed been previously parametrized, but in the CHARMM force field, rather than OPLS-AA.[269] Using these parameters directly,

however, could introduce inconsistencies which, when comparing results between different types of surfactants, with some modeled in OPLS-AA and some modeled in CHARMM, could complicate the analysis. (It would be difficult to quantify how much of an observed difference in, say, atomic group hydration was due to the use of the different force field, lacking explicit models for the molecule in both force fields.) Additionally, other classes of alkylbenzene sulfonate surfactants have *not* yet been studied, including those with alkyl chains at both the 2-position and 5-position. Hence, I identified a need for developing new force-field parameters for this system. Using the work of Canongia Lopes and coworkers[194] as a guide, I assembled several techniques from the literature and applied them to the linear alkylbenzene sulfonates. I will use this system as a case study in the following sections to illustrate the techniques.

E.2 Procedures for Force-Field Parametrization

E.2.1 Ab initio Calculations

All the techniques for force-field parametrization that I will present below depend upon ab initio geometry optimization, to determine either the global minimum potential energy configuration, or the energy-minimized configurations with certain internal coordinates (such as a dihedral angle) set to fixed values. Given an optimized geometry (i.e., a set of Cartesian coordinates for each atom), one also needs to be able to compute the configurational electronic potential energy (a single point energy calculation) and atomic partial charges (through electrostatic surface potential fitting). For all these tasks, I make use of the Gaussian 03 software package.[234]

Gaussian 03 implements a wide variety of model chemistries for the approximate solution of the Schrödinger equation

$$\mathbf{H}\psi = E\psi, \tag{E.2.1}$$

where $\mathbf{H}$ is a molecular Hamiltonian operator, ψ is the wavefunction (dependent on

electron and nuclei positions and time), and E is the energy of the molecule. Specifying a model chemistry involves: (i) choosing a level of theory, which refers to the degree of approximation made in solving a time-independent form of the Schrödinger equation under the Born-Oppenheimer approximation (e.g., Hartree-Fock theory, Møller-Plesset perturbation theory), and (ii) choosing a molecular orbital basis set. A molecular orbital (MO) is a function which represents the spatial distribution of an electron in a molecule. Using a Slater determinant, a closed-shell wavefunction ψ can be related to a set of $n/2$ MO's (for an even n electrons) and two spin functions (pairing two electrons per MO). Calculation of any particular MO is performed using a weighted sum of so-called one-electron basis functions centered on the positions of the atomic nuclei. These one-electron basis functions are themselves expressed as linear combinations of primitive gaussian functions.[323]

To summarize the above discussion, each MO, ϕ_i (for $i = 1$ to $n/2$), can be represented as

$$\phi_i = \sum_{\mu} c_{\mu,i}\chi_{\mu}, \tag{E.2.2}$$

where the $c_{\mu,i}$ are the weights (referred to as molecular orbital expansion coefficients) assigned to each one-electron basis function, χ_{μ}. The weights $c_{\mu,i}$ must be solved for using a variational principle during the course of calculation. The index μ spans all the basis functions included in the basis set for the atom center of interest. The χ_{μ}'s (also called contracted gaussians) are expressed in terms of primitive gaussians as follows:

$$\chi_{\mu} = \sum_{p} d_{\mu,p}g_p, \tag{E.2.3}$$

where the $d_{\mu,p}$ are contraction coefficients which weight the primitive gaussians, g_p. The weights $d_{\mu,p}$ are fixed by choice of basis set. The index p spans all the primitive gaussians included in the basis set.[323]

A complete basis set for a particular nuclear center then consists of: (i) specification of the number of g_p's to be used, (ii) provision of the precise functional form of each g_p (e.g., s, p, d, and f orbitals all exhibit differing radial-dependence), and (iii) declaration of the $d_{\mu,p}$ weights. The more primitive gaussians that are included, the

larger the computational cost. However, larger sets also lead to fewer restrictions on the position of electrons in space, more accurately representing the behavior of the real system.[234,323]

For the procedures that I will discuss shortly, I follow the strategy of Canongia Lopes and coworkers[191] in performing all geometry optimizations at the Hartree-Fock (HF) level of theory with a 6-31G(d) basis set. Single-point energy calculations and electrostatic surface potential fitting are performed using the Møller-Plesset (MP) perturbation theory,[324] which represents an improvement on HF theory by introducing the effects of electron correlations. Specifically, use of MP2 is recommended, which is the level of theory obtained by truncating the MP perturbation expansion after second order (providing the first non-zero correction to the electronic energy). It is also recommended that MP2 be used in combination with a relatively large basis set, the cc-pVTZ(-f) basis set, which is the triple-zeta, correlation-consistent basis set of Dunning,[325,326] modified to remove the f orbitals of heavy atoms and the d orbitals of hydrogen.[327] This basis set has been found to work very well with MP2 in computing conformational energies,[328] although it is not available by default with Gaussian 03. To create a custom basis set (.gbs) input file for cc-pVTZ(-f) that can be used with Gaussian, one can obtain the full cc-pVTZ basis set from the EMSL Basis Set Exchange[329,330] (online at https://bse.pnl.gov/bse/portal), and remove the specified orbitals by hand. (Gaussian 03 allows the use of such external, customized basis sets for any of its standard calculations.[234])

E.2.2 Bond-Stretching Parametrization

Using Gaussian 03,[234] directly following a geometry optimization of a molecule (keyword: OPT), a frequency calculation (keyword: FREQ) can be used to calculate force constants associated with the potential-energy minimized structure. These force constants are second derivatives of the electronic potential energy, E_{elec}, of a molecule with respect to the positional coordinates of its atoms. Note that these force constants have meaning only if the frequency calculation is performed using the same basis set and theoretical model as the geometry optimization.[323]

The force constants form a Cartesian Hessian, **H**, a $3N \times 3N$ matrix with elements $H_{ij} = \partial^2 E / \partial x_i \partial x_j$, where E is the potential energy of the molecule, and x_i and x_j are elements of $\underline{x}$, the vector containing the $3N$ total Cartesian coordinate variables for the N total atoms in the molecule (the elements of $\underline{x}$ are, in order, the x, y, and z-coordinates of atom 1, followed by those of atom 2, then atom 3, etc.). Since **H** is symmetric, it is stored as a lower-triangular matrix in practice, containing $3N\left(3N+1\right)/2$ total elements. Accessing **H** from the Gaussian 03 output requires converting the binary checkpoint file (.chk extension) generated by Gaussian 03 during the FREQ calculation into a formatted, plain-text file (.fchk extension) using the formchk.exe utility packaged with Gaussian 03 in its root installation directory. Within the .fchk file, the lower triangular form of **H** is reported as a vector under the heading "Cartesian Force Constants", where this vector, $\underline{h}$, has elements $(H_{11}, H_{21}, H_{22}, H_{31}, H_{32}, H_{33}, ...)$.

Seminario[1] has reported a useful method for extracting the force constants for bonded interactions from **H**, including those for bond stretching, angle bending, and even bond torsion. Note, however, that the bond-torsion force constants are limited in application, since they are derived using a harmonic potential, which is appropriate only for small perturbations from the geometry-optimized dihedral angles. That is, the full bond-torsion potential energy profile as a function of dihedral angle, which contains multiple energy minima, cannot be reproduced using this method, which admits only one minimum. A better treatment of bond torsion parametrization can be found in Section E.2.4 below.

In testing Seminario's approach, I have found the method to work quite well in terms of achieving close agreement with the bond-stretching force constants already found in the OPLS-AA force field. However, the angle bending force constants are not as accurately reproduced, and I prefer instead the use of an empirical correlation due to Halgren,[2,331] and discussed in Section E.2.3 below (also suggested by Canongia Lopes and coworkers[194]). Therefore, although Seminario's approach has been used by others to develop a full set of force-field parameters exclusively from the Hessian (see, for example, the program Hess2FF[332]), I use it only in the case of bond-stretching

parametrization, which I will now describe.

Starting with the Hessian

$$\mathbf{H} \equiv \begin{bmatrix} \frac{\partial^2 E}{\partial x_1 \partial x_1} & \frac{\partial^2 E}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 E}{\partial x_1 \partial x_{3N}} \\ \frac{\partial^2 E}{\partial x_2 \partial x_1} & \frac{\partial^2 E}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 E}{\partial x_2 \partial x_{3N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial x_{3N} \partial x_1} & \frac{\partial^2 E}{\partial x_{3N} \partial x_2} & \cdots & \frac{\partial^2 E}{\partial x_{3N} \partial x_{3N}} \end{bmatrix}, \tag{E.2.4}$$

the reaction force vector which arises due to displacements of the atoms of a molecule in space can be computed as follows:

$$\delta \underline{F} = [\delta F_1, \delta F_2 \ldots, \delta F_{3N}]^T = -\mathbf{H} \delta \underline{x}. \tag{E.2.5}$$

As an aside, the Hessian matrix $\mathbf{H}$ is clearly symmetric, since the order of partial derivatives may be exchanged, and $\mathbf{H}$ therefore admits an orthonormal basis of eigenvectors with corresponding eigenvalues, where each eigenvector represents the direction of displacement for a normal mode of vibration of the molecule and the eigenvalues can be construed as harmonic force constants. Seminario thus uses the symbol $[k]$ in place of $\mathbf{H}$, where k reminds the reader that $\mathbf{H}$ acts in this manner on $\delta \underline{x}$.

The reaction force, $\delta \underline{F}_p$, on a particular atom p $(1 \leq p \leq N)$ due to the displacement of all atoms in the system can be extracted from $\delta \underline{F}$ in Eq. E.2.5 as $\delta \underline{F}_p = \left[F_{3(p-1)+1}, F_{3(p-1)+2}, F_{3(p-1)+3}\right]^T$. Note that the vector $\delta \underline{F}_p$ can also be computed directly by taking a rectangular, $3 \times 3N$ submatrix of $[k]$, which I will symbolize as $[k_p]$, consisting of the three rows of $\mathbf{H}$ indexed by $3(p-1)+1$, $3(p-1)+2$, and $3(p-1)+3$. Furthermore, the reaction force on a particular atom p due to a displacement in atom q $(1 \leq q \leq N)$, $\delta \underline{F}_{pq}$, can be computed by taking a square, 3×3 submatrix of $[k_p]$, which I will call $[-k_{pq}]$, consisting of the values of the three columns of $[k_p]$ indexed by $3(q-1)+1$, $3(q-1)+2$, and $3(q-1)+3$. Then,

$$\delta \underline{F}_{pq} = [k_{pq}]\, \delta \underline{x}_q, \tag{E.2.6}$$

where $\delta\underline{x}_q$ is the displacement of atom q, corresponding to the $3(q-1)+1$, $3(q-1)+2$, and $3(q-1)+3$ entries of $\delta\underline{x}$. Note that I have absorbed the negative sign found, for example, in Eq. E.2.5 into the components of $[k_{pq}]$. This follows Seminario's notation.[1]

The 3×3 matrix $[k_{pq}]$ is not generally symmetric (or even normal), due to the presence of all the other atoms in the molecule. Therefore, its normalized eigenvectors, $\widehat{v}_m^{pq}$ (where $m = 1, 2, 3$) are not generally orthogonal.

A condition for atoms p and q to be considered bonded is that the three corresponding eigenvalues, λ_m^{pq}, are all positive. If this condition holds, and one assumes that the potential energy for a bond can be written as a harmonic potential. That is,

$$U_{b,pq} = \frac{1}{2} k_{b,pq} \left(b_{pq} - b_{0,pq}\right)^2, \tag{E.2.7}$$

where the subscript b indicates a bond, the subscript pq specifies the particular bond formed by atoms p and q, $k_{b,pq}$ is the bond force constant, b_{pq} is the bond length in a given configuration, and $b_{0,pq}$ is the equilibrium (geometry-optimized) bond length, then Seminario suggests the following form for the bond-stretching force constant:

$$k_{b,pq} = \sum_{m=1}^{3} \lambda_m^{pq} \left| \widehat{u}^{pq} \cdot \widehat{v}_m^{pq} \right|, \tag{E.2.8}$$

where $\widehat{u}^{pq}$ is the normalized bond vector pointing from atom p to atom q. That is, contributions to $k_{b,pq}$ arise from each eigenvector/eigenvalue pair.[1]

In Figure E-1, the molecule benzene sulfonate is presented in its geometry-optimized form, found using the model chemistry HF/6-31G(d) and the OPT keyword in Gaussian 03 (the snapshot shown is taken from the program GaussView, which accompanies Gaussian 03). These atomic coordinates and model chemistry were used to perform a FREQ calculation, as mentioned above, from which the Cartesian Hessian matrix was extracted from the checkpoint file. Application of Eq. E.2.8 with the eigenvalues and eigenvectors determined for submatrices of this Hessian yielded the results shown in Table E.1 (converting appropriately from the units used in Gaussian – Hartrees and

Table E.1: Results for Bond Force Constants in Benzene Sulfonate Using Seminario's Approach.[1]

Bond Atom Types	Labels	$k_b \left(\frac{kJ/mol}{nm^2}\right)$	$b_0\,(nm)$
SY2-CA	12,3	190450.2	0.1798
SY2-O2	12,13	558253.8	0.1452
SY2-O2	12,14	564415.1	0.1451
SY2-O2	12,15	558294.8	0.1452
CA-CA	3,4	353423.8	0.1387
CA-CA	3,2	353770.3	0.1387

Bohr – to the units used in GROMACS – kJ/mol and nm). The atom types in Table E.1 reflect the atom types used in OPLS-AA, while the labels follow the numbering shown in Figure E-1. The results displayed are the average values found by taking $(k_{b,pq} + k_{b,qp})/2$ – since $[k_{pq}]$ is only approximately symmetric, there is a minor deviation in the force constants. Furthermore, in applying these results to simulations in GROMACS, I average the results for all bonds with the same atom types, and use the resulting values as representive of all bonds of that type. Hence, for the $SY2$-$O2$ bond type, I use a value of $k_b = 560321.2\ \frac{kJ/mol}{nm^2}$ and $b_0 = 0.1451\ nm$. Finally, note that the CA-CA bond type is already specified in OPLS-AA. I have included the results here to demonstrate reasonable agreement with OPLS-AA. Here, the average value in the vicinity of the sulfur group is $k_b = 353597.1\ \frac{kJ/mol}{nm^2}$ and $b_0 = 0.1387\ nm$. In OPLS-AA, the values for a generic CA-CA bond are $k_b = 392459.2\ \frac{kJ/mol}{nm^2}$ and $b_0 = 0.14\ nm$ (obtained from the GROMACS OPLS-AA force-field files). To remain consistent with OPLS-AA, I make use of the OPLS-AA values whenever available.

It is interesting to compare the $SY2$-CA and $SY2$-$O2$ bond parametrizations in Table E.1 to analogous bonds in OPLS-AA. For the class of sulfonamides ($R_2NS(=O)_2R$), the sulfur has atom type SY and the oxygens OY. There is, in fact, an SY-CA bond specified: $k_b = 284512.0\ \frac{kJ/mol}{nm^2}$ and $b_0 = 0.177\ nm$ (obtained from GROMACS). Comparing to the $SY2$-CA bond parameters, where $k_b = 190450.2\ \frac{kJ/mol}{nm^2}$ and $b_0 = 0.1798\ nm$, we see that the $SY2$-CA bond is less stiff than the SY-CA bond, although the equilibrium bond lengths are approximately equal. There is also an SY-OY bond specified: $k_b = 585760.0\ \frac{kJ/mol}{nm^2}$ and $b_0 = 0.144\ nm$ (also obtained from GRO-

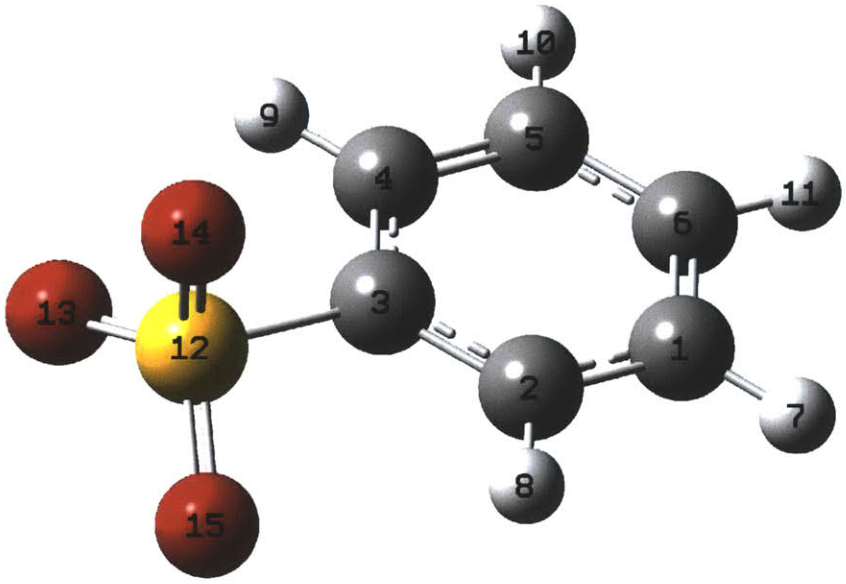


Figure E-1: Chemical structure and (arbitrary) atom numbering for benzene sulfonate using Gaussian 03. The atom numbers are referred to in the text when discussing force-field parameter development. (Color code: gray – carbon, white – hydrogen, yellow – sulfur, red – oxygen.)

MACS). Comparing to the $SY2$-$O2$ bond parameters, where $k_b = 560321.2\ \frac{kJ/mol}{nm^2}$ and $b_0 = 0.1451\ nm$, we see excellent agreement. Therefore, it appears that the sulfur-oxygen bond stiffness is relatively unaffected by the neighboring functional groups, while the sulfur-carbon bond is more sensitive.

E.2.3 Angle-Bending Parametrization

Halgren[2,331] proposes the following empirical correlation for the angle-bending force constant to be applied to an angle comprising atoms a, b, and c, which he has demonstrated fits experimental data for a test set of organic molecules to within 12% rms error:

$$k_\theta = 1.75 Z_a C_b Z_c \left(R_{ab} + R_{bc}\right)^{-1} \theta_{abc}^{-2} \exp\left(-2D\right), \tag{E.2.9}$$

where the Z and C constants depend on the atom element (see Table E.2 for a selection of relevant values from Halgren's work[2]), R_{ab} and R_{bc} are the equilibrium bond lengths between atoms a and b and atoms b and c, respectively (in Angstroms), θ_{abc} is the equilibrium bond angle (in radians), and D is calculated as follows:

$$D = \frac{\left(R_{ab} - R_{bc}\right)^2}{\left(R_{ab} + R_{bc}\right)^2}. \tag{E.2.10}$$

The units of k_θ in Eq. E.2.9 are $\left(mdyn\text{-}\mathring{A}\right)/rad^2$, which can be converted into GROMACS units of $\left(kJ/mol\right)/rad^2$ by multiplying k_θ by Avogadro's number and a unit conversion factor of $10^{-21}\ kJ/\left(mdyn\text{-}\mathring{A}\right)$. Note that special care must be taken when applying the constants in Table E.2 to angles containing carbon. Tetrahedral (sp^3 hybridized) carbon (e.g., in linear alkanes) has a C parameter of 0.97, while carbon in all other forms (e.g., in benzene) has a C value of 1.05. Certain constants in Table E.2 are not applicable for hydrogen, fluorine, and silicon, since these elements can only be found either in an angle endpoint (H and F) or in an angle center (Si).

Halgren points out that the expression in Eq. E.2.9 has four key dependencies: (1) a dependency on elemental composition of the angle, through the constants of Table

Table E.2: Atomic Constants To Be Used in Eq. E.2.9.

Atom	Z	C	Atom	Z	C	Atom	Z	C
H	1.44	-	N	2.67	1.06	Si	-	0.95
C (sp^3)	2.49	0.97	O	3.12	1.24	P	2.53	1.23
C	2.49	1.05	F	2.67	-	S	3.01	1.22

E.2, (2) a dependency on the size of the arms of the angle, through the $(R_{ab} + R_{bc})^{-1}$ term, such that k_θ increases as the sum of the bond lengths decreases, (3) a dependency on the equilibrium angle itself, through the θ_{abc}^{-2} term, such that k_θ increases as θ_{abc} decreases, and (4) a dependency on the asymmetry of the angle, through the $\exp(-2D)$ term, and, more specifically, the formulation of D. Regarding this last dependency, note that $R_{ab} = R_{bc}$ (bond length symmetry) yields $D = 0$, and $\exp(-2D) = 1$. For asymmetric cases, $0 < D < 1$ (since R_{ab}, $R_{bc} > 0$), such that $1 > \exp(-2D) > 1/e^2$. In other words, bond length asymmetries will decrease k_θ. Since usual differences in bond lengths are quite small, this effect is minor.

Using the appropriate values in Table E.2 and applying Eqs. E.2.9 and E.2.10, I have investigated several angles within benzene sulfonate (see Figure. E-1), which are summarized in Table E.3 below (equilibrium bond lengths and bond angles are computed from the geometry-optimized structure in Gaussian 03).

In Table E.3, note the force constant for the SY2-CA-CA angle, with $k_\theta = 572.2$ $(kJ/mol)/rad^2$. This force constant is reported by He and coworkers[269] in their CHARMM implementation as only 83.7 $(kJ/mol)/rad^2$, although they do not clearly describe their approach for this aspect of their force-field development. In order to confirm Halgren's approach, then, in view of this discrepancy, I selected a few existing k_θ OPLS-AA values to compare to, which are summarized in Table E.4. For these cases, equilibrium bond lengths and bond angles were drawn from the source given in brackets in the final, k_θ reference column. Note that, in all cases, there is excellent agreement between the Halgren values and the reported values. Since this includes an example with sulfur (alkyl sulfonate), I will utilize the Halgren results for benzene sulfonate.

Table E.3: Results for Angle Force Constants in Benzene Sulfonate Using Halgren's Correlation.[2]

Angle Atom Types	Z_a	C_b	Z_c	$R_{ab}(\AA)$	$R_{bc}(\AA)$	$\theta_{abc}\,(^\circ)$	$\theta_{abc}\,(rad)$	$k_\theta\left(\frac{kJ/mol}{rad^2}\right)$
O2-SY2-O2	3.12	1.22	3.12	1.451	1.451	114.1	1.991	1087.9
O2-SY2-CA	3.12	1.22	2.49	1.451	1.798	104.3	1.820	907.2
SY2-CA-CA	3.01	1.05	2.49	1.798	1.387	120.2	2.098	572.2

Table E.4: Additional Results for Angle Force Constants Using Halgren's Correlation.[2]

Angle Atom Types	Z_a	C_b	Z_c	$R_{ab}(\AA)$	$R_{bc}(\AA)$	$\theta_{abc}\,(^\circ)$	$\theta_{abc}\,(rad)$	$k_\theta\left(\frac{kJ/mol}{rad^2}\right)$	$k_\theta^{[\text{ref.}]}\left(\frac{kJ/mol}{rad^2}\right)$
CA-NA-CA	2.49	1.06	2.49	1.34	1.34	120.4	2.101	585.5	586[193]
HC-CT-HC	1.44	0.97	1.44	1.09	1.09	107.8	1.881	274.8	276.1[188]
CA-CA-CT	2.49	1.05	2.49	1.4	1.51	120.0	2.094	536.1	586.2[333]
CA-CA-CA	2.49	1.05	2.49	1.4	1.4	120.0	2.094	558.8	527.2[320]
O2-SY2-CT	3.12	1.22	2.49	1.455	1.792	104.5	1.824	904.9	870.0[194]

E.2.4 Bond-Torsion Parametrization

Parametrization of a missing bond torsional profile requires complete specification of all other parameters, including: (i) bond-stretching parameters, using either OPLS-AA values or values derived using Seminario's approach described in Section E.2.2, (ii) angle-bending parameters, using Halgren's empirical correlation when OPLS-AA values are unavailable, as described in Section E.2.3, (iii) all known dihedral angles, using OPLS-AA values, or dihedral angles obtained from other analogous organic molecules, and (iv) atom partial charges, which may be obtained from the CHELPG algorithm described in the next section. Note that van der Waals parameters (Lennard Jones σ and ϵ values) are assumed to be readily available – I have not yet found a need to develop these parameters myself for any surfactant molecule considered.

The procedure suggested by Canongia Lopes and coworkers,[194] and previously used by others in force-field development for OPLS-AA[188,322] involves conducting a relaxed potential energy surface scan using ab initio software such as Gaussian 03[234] about the dihedral of interest. This implies that the dihedral is advanced in regular degree intervals (for this work, I use 5°, although others have used broader intervals, such as 10°[322]), and a geometry optimization is conducted for the molecule as a whole at each step. (In contrast, a rigid potential energy surface scan starts with a geometry-optimized configuration, and rigidly rotates the dihedral through the same intervals.) A single-point energy calculation is then conducted at each step, yielding a curve for E_{elec}, the electronic potential energy as a function of dihedral angle. This is the quantum-mechanical (QM) torsional profile.

Then, using the molecular force-field parameters for the molecule (excluding the missing dihedral of interest) and the geometry-optimized coordinates for each dihedral step, a potential energy curve is generated (e.g., this can be accomplished in Program BUILDER – see Appendix C for a discussion of this Program). This is the molecular dynamics (MD) torsional profile.

Finally, the difference in the QM and MD torsional profiles yields the potential energy due to the missing torsional parameters. A least-squares fitting technique can

Table E.5: Ryckaert-Bellemans Potential Coefficients for the $O2$-$SY2$-CA-CA Dihedral.

Coefficient	Value (kJ/mol)	Coefficient	Value (kJ/mol)
C_0	3.29185	C_3	-0.35256
C_1	0.34603	C_4	12.99238
C_2	-16.27770	C_5	0.00000

be used to obtain the missing parameters. Recalculating the MD torsional profile then demonstrates close agreement with the QM values.

The curves for the missing dihedral $O2$-$SY2$-CA-CA in benzene sulfonate are shown in Figure E-2. Fitting the Ryckaert-Bellemans form for this bond-torsion potential energy,

$$U_{O2-SY2-CA-CA} = \sum_{n=0}^{5} C_n \cos^n(\phi - \pi), \tag{E.2.11}$$

to the data, recognizing that there are six total dihedrals of this type in benzene sulfonate and that their angles are offset as the $SY2$-CA bond is rotated, yields the coefficients listed in Table E.5.

As a final comment on dihedrals, note that the magnitude of the QC-MD curve in Figure E-2 is relatively small compared to typical curves encountered in the literature.[194] This is due to the fact that there are three oxygens present and offset from one another, such that, as one oxygen rotates into the plane of the aromatic ring, interacting sterically with the hydrogen attached to the aromatic carbon beta to the sulfonate functional group, the other two oxygens are in more favorable locations. As a result, the difference between the peaks and troughs of the QC-MD curve is minimal (less than 0.2 kJ/mol). When one examines the curve of the torsional potential of He and coworkers[269] in their CHARMM force-field representation for linear alkylbenzene sulfonate for a single $O2$-$SY2$-CA-CA dihedral, large values in the potential are encountered (order 30 kJ/mol). However, when all six dihedral angles are properly accounted for, and added together, the difference between the peaks and troughs is less than 0.25 kJ/mol, a very comparable result to the one obtained here.

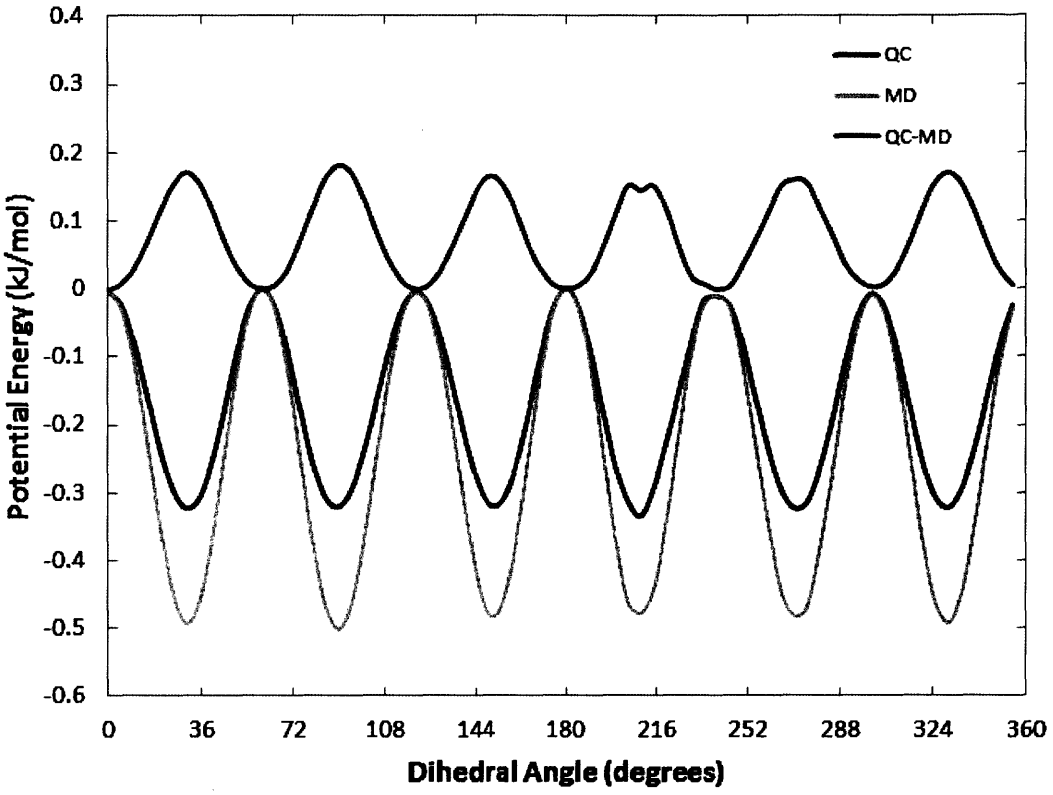


Figure E-2: O2-SY2-CA-CA torsional potential energy profile for benzene sulfonate. The QC curve is the quantum chemistry (ab initio) potential energy, the MD curve is the molecular dynamics potential energy (missing the dihedral of interest), and the QC-MD curve is the difference between the two (to be fit to develop the missing dihedral parameters). The QC curve has some small amount of error due to convergence to a tolerance – hence, small spurious features, such as the dip in the QC-MD curve will at times arise. The fitting function (a truncated cosine series) has a smoothing effect which eliminates these features.

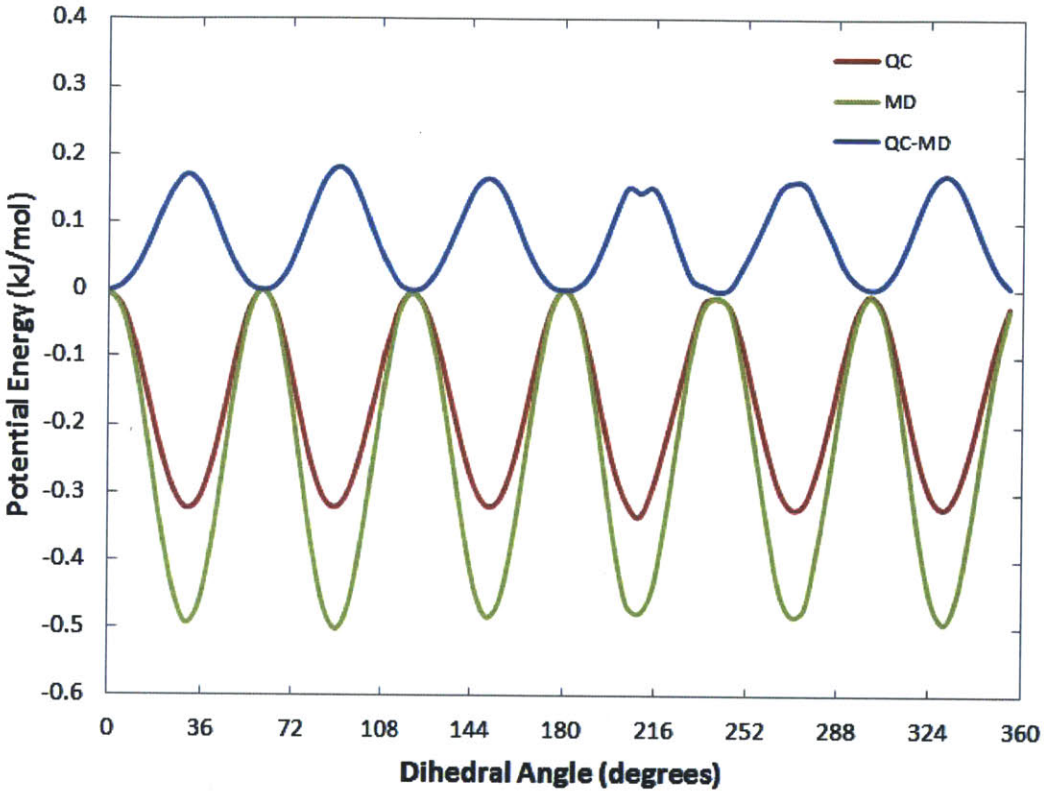


Figure E-2: O2-SY2-CA-CA torsional potential energy profile for benzene sulfonate. The QC curve is the quantum chemistry (ab initio) potential energy, the MD curve is the molecular dynamics potential energy (missing the dihedral of interest), and the QC-MD curve is the difference between the two (to be fit to develop the missing dihedral parameters). The QC curve has some small amount of error due to convergence to a tolerance – hence, small spurious features, such as the dip in the QC-MD curve will at times arise. The fitting function (a truncated cosine series) has a smoothing effect which eliminates these features.

Table E.6: Modified CHELPG Charges for Benzene Sulfonate (Net $q = -1$).

Label	Atom	q	Label	Atom	q	Label	Atom	q
1	C	-0.165	6	C	-0.087	11	H	+0.073
2	C	-0.026	7	H	+0.085	12	S	+1.363
3	C	-0.110	8	H	+0.095	13	O	-0.739
4	C	-0.026	9	H	+0.095	14	O	-0.739
5	C	-0.165	10	H	+0.085	15	O	-0.739

For the benzene sulfonate molecule depicted in Figure E-1, the CHELPG charges computed using Gaussian 03 are those listed in Table E.6 (in units of elementary charge). These atomic charges have been slightly modified by averaging together the charges for atoms which should have equal values by symmetry arguments (e.g., groups like carbons 1 and 5, hydrogens 7 and 10, or oxygens 13, 14, and 15 in Figure E-1). The CHELPG charges in these case differed by less than 0.01 in the case of the oxygens, less than 0.001 in the case of the carbons, and less than 0.0001 in the case of the hydrogens. Therefore, the effect of using average atomic charge values is negligible.

Note that the charges in Table E.6 are used only for developing the dihedral parameters for the missing *O2-SY2-CA-CA* dihedral described in Section E.2.4. For production runs with linear alkylbenzene sulfonates, the bond-stretching, angle-bending, and bond-torsion parameters developed using benzene sulfonate are combined with CHELPG charges developed specifically for the alkylbenzene sulfonate of interest. This follows the principle of using more basic molecules to parametrize the bonded interaction parameters needed for the analysis of larger molecules.[98,335] I have developed three distinct sets of charges for alkylbenzene sulfonate molecules: (i) one for a linear alkane attached at the 4-position of the ring, (ii) one for a branched alkane attached at the 4-position of the ring (branching at the alpha carbon), and (iii) one for a linear alkane attached at the 2-position and 5-position of the ring. These systems are described in detail in Chapter 6.

Appendix F

Molecular Property Determination

F.1 Overview

For a predictive model of surfactant micellization behavior to be useful, it must be able to distinguish between similar surfactant chemical structures, since small changes in structure and atomic composition can demonstrably impact the aggregation process, as evidenced by a shifting in the critical micelle concentration (cmc) and changes to the shape and size distribution of micelles formed in solution at various total surfactant concentrations.[12,13] In the molecular-thermodynamic (MT) theory presented in Chapter 2, which I have helped develop and refine in order to accommodate the modeling of branched surfactants, various molecular parameters, or descriptors, are used to uniquely describe each surfactant considered. Molecular descriptors have been used in a variety of fields, such as pharmaceutical science and chemistry, to compare and contrast molecules, draw correlations between molecular structure and activity, and design novel molecules possessing some pre-determined desired property. These descriptors can be experimental or theoretical in origin, but always have a well-defined procedure for their measurement or calculation.[336]

In many applications, molecular descriptors are used to develop quantitative structure-activity relationships (QSAR's), where a statistical technique, such as Principle Component Analysis (PCA), is used to create a model based on an existing experimental dataset which gives preferential weighting to those descriptors which

best explain the data.[337] In contrast, the use of molecular descriptors within an MT theory arises from a phenomenological thought process utilized to compute the free energy of micellization, g_{mic}, for any possible micelle present in solution. The value of g_{mic} represents the difference between reference-state chemical potentials of the micelle as a whole and of its individual constituents (i.e., surfactant molecules and bound counterions) in their dispersed states. Once g_{mic} is known, the aggregation behavior of the solution can be fully determined.[81]

The specifics of this thought process are thoroughly discussed in Section 2.3.1 of Chapter 2, but a key point of relevance to this Appendix is that the thought process involves a conceptual division of a surfactant molecule into so-called head (hydrophilic) and tail (hydrophobic) regions. The specific molecular descriptors required for MT modeling, then, are related to these two regions. They include: (i) the cross-sectional area of the surfactant head, a_h, (ii) the valence, z, and the location (relative to the micelle core-water interface), d_z, of each charge in the surfactant head (one charge for anionic and cationic surfactants, and two charges for zwitterionic surfactants), (iii) the length of the surfactant head, l_h, (iv) the solubility of the surfactant tail in water, S_{tail}, (v) the planar interfacial tension of the surfactant tail or its analogous oil against water, σ_0, (vi) a Tolman distance, δ, used to provide curvature corrections to σ_0, (vii) the area at the micelle core-water interface shielded by the cross-sectional area of the bond connecting head and tail, a_0, and (viii) the molecular volume of the tail, v_{tail}.

This Appendix includes a summary of methods used in determining values for several of the above descriptors. In Section F.2, I tabulate a few useful properties related to the elements most commonly found in surfactant molecules. In Section F.3, I tabulate water densities at various temperatures of interest, and provide corresponding molarities. In Section F.4, I describe my development of a simple group-contribution model for molecular volume based on fitting to a sizable set of molecular volumes derived from D_4^{20} specific gravity data (i.e., liquid densities at $20°C$ divided by the reference density of water at $4°C$). In Section F.5, I describe several possible group-contribution models for tail solubility. In Section F.6, I discuss the procedure that I

Table F.1: Atomic Weights and Van der Waals Radii for the Elements Most Commonly Found in Surfactant Molecules

Element	Symbol	A_r*	R_{VdW}†	Element	Symbol	A_r*	R_{VdW}†
Hydrogen	H	1.008	0.117	Fluorine	F	18.998	0.130
Carbon	C	12.011	0.175	Silicon	Si	28.085	0.210
Nitrogen	N	14.007	0.155	Phosphorus	P	30.974	0.180
Oxygen	O	15.999	0.140	Sulfur	S	32.06	0.180

* Relative atomic weight (amu).[3]
† Van der Waals radius (nm). Values from Gavezzotti[338] (H, C, N, O, and F) and Bondi[339] (Si, P, and S).

use for estimating the tail interfacial tension against water for linear and branched alkyl tails. Finally, in Section F.7, I provide a listing of packing polynomials appropriate for the linear and branched surfactant tails considered in this Thesis.

F.2 Atomic Weights and Radii

In Table F.1, I list the elements most commonly found in surfactant molecules (excluding their counterions), together with their associated relative atomic weights and van der Waals radii. In addition to the elements commonly found in most organic materials (hydrogen, carbon, nitrogen, and oxygen), the element fluorine is found in fluorinated surfactants, the element silicon is found in siloxane surfactants, the element phosphorus is found in classes of surfactants such as the phosphates, and the element sulfur is found in classes of surfactants including the sulfates, sulfonates, and sulfoxides.[12] The relative atomic weights are used in Section F.4 to convert from density to molecular volume, and were used in the preparation of Section F.5 to convert experimental solubility data from a weight basis such as ppm (g solute/$10^6 g$ solution) to molar concentration, where necessary. The van der Waals radii are used in the calculation of solvent accessible surface areas based on computer simulation results, as described in Chapter 3.

Table F.2: Density and Molarity of Water at Atmospheric Pressure[3]

T (°C)	ρ_{H_2O} (g/cm^3)	$\mathcal{M}_{H_2O}$ (mol/L)	T (°C)	ρ_{H_2O} (g/cm^3)	$\mathcal{M}_{H_2O}$ (mol/L)
4	0.9999749	55.508	35	0.9940326	55.178
20	0.9982067	55.410	40	0.9922152	55.077
25	0.9970470	55.345	45	0.99021	54.966
30	0.9956488	55.268	60	0.98320	54.577

F.3 Liquid Water Density and Molarity

The densities and molarities of liquid water at various temperatures are summarized in Table F.2 for standard atmospheric pressure ($1\ atm = 101325\ Pa$) and temperatures following the ITS-90 temperature scale.[3,340] The density of water at $4°C$ was used to prepare the list of densities in Table F.5 from D_4^{20} data. The molarity of water at $25°C$ was used to compute the group contributions to the transfer free energy in Table F.9.

The molarity of water in Table F.2 is computed from the density as follows:

$$\mathcal{M}_{H_2O}\left[\frac{mol}{L}\right] = \frac{\rho_{H_2O}\left[\frac{g}{cm^3}\right]}{MW_{H_2O}\left[\frac{g}{mol}\right]} \times 1000\left[\frac{cm^3}{L}\right], \tag{F.3.1}$$

where ρ_{H_2O} is the liquid density of water, and $MW_{H_2O} = 18.015\frac{g}{mol}$ is its molecular weight (determined using the values in Table F.1).

F.4 Surfactant Tail Liquid Density

The packing free-energy contribution to g_{mic}, described in Section 2.3.7 of Chapter 2 and throughout Chapter 4, requires knowledge of the distribution of volume of a chain (the surfactant tail) in space as a function of the conformation of that chain. This is most readily accomplished using an additive group-contribution method, where each atomic group in the chain has a specified volume, located at a particular point in space corresponding to some convenient center for the group (typically the center of

mass of the heavy atom). This is represented symbolically as

$$v(r) = \sum_a \delta(r - r_a)\, v_a, \tag{F.4.1}$$

where $v(r)$ is the volume of the chain at radial position r, a indexes the atomic groups, v_a is the total volume assigned to atomic group a, r_a is the radial position of atomic group a, and δ is the Dirac delta function (r_a is restricted to lie between $r = 0$, the radial position of the center of mass of the micelle, and $r = R$, the radius of the micelle). Note that the presence of the delta function means that $v(r)$ has units of volume per length, such that integration across the radial dimension of the micelle yields the total tail volume. Specifically,

$$\int_0^R v(r)\, dr = \sum_a v_a = v_{tail}. \tag{F.4.2}$$

Note also that, in this discussion, "radius" refers to the radial distance from the center of a spherical micelle, the perpendicular distance from the central axis of an infinite cylindrical micelle along a cross-section, or the perpendicular distance from the midplane of an infinite bilayer micelle.

Strictly speaking, in Eq. F.4.1 the values of v_a for individual atomic groups can be functions of the set of groups present in the molecule and their specific connectivity. As a result, some approaches to predict liquid density make use of topological indices, which can be calculated from the molecular graph of a compound.[341] Other approaches correlate densities to other experimentally-determined properties, including critical points.[342] Here, I will take a simpler approach, where I will assume that the v_a's are only functions of the group identity. That is, the volume for an atomic group such as CH_2 will be assumed to be a constant across all types of molecules.

The CRC Handbook of Tables for Organic Compound Identification[265] provides a wealth of D_4^{20} specific gravity data for linear and branched hydrocarbons, alkylbenzenes, alcohols, ethers, and many other liquid compounds. The molecular volume of

each of these compounds at $T = 20°C$, $v_m^{20°C}$, can be related to their D_4^{20} values by

$$v_m^{20°C}\left[\mathring{A}^3\right] = \frac{MW_m\left[\frac{g}{mol}\right]}{D_4^{20}\rho_{H_2O}^{4°C}\left[\frac{g}{cm^3}\right]}\frac{10^{24}\left[\frac{\mathring{A}^3}{cm^3}\right]}{N_{AV}\left[\frac{1}{mol}\right]}, \tag{F.4.3}$$

where MW_m is the molecular weight of molecule m (determined using Table F.1), $\rho_{H_2O}^{4°C}$ is the density of water at $4°C$ (see Table F.2), and N_{AV} is Avogadro's number ($6.022 \times 10^{23}\ mol^{-1}$). Note that $\rho_m^{20°C} = D_4^{20}\rho_{H_2O}^{4°C}$ is the density of molecule m at $T = 20°C$.

Assuming that the molecular volume, v_m, is related to the v_a values as

$$v_m = \sum_{a(m)} v_a, \tag{F.4.4}$$

where $a(m)$ indicates the set of atomic groups in molecule m (similar to Eq. F.4.2), one can fit values of v_a to reproduce either $v_m^{20°C}$ or $\rho_m^{20°C}$ (the results were found to be essentially the same using a least-squares approach). The results of fitting to reproduce $\rho_m^{20°C}$ are summarized in Table F.3 and Figure F-1. The full data set used for fitting can be found in Table F.5. Note that previous modeling efforts focusing on solely linear surfactants, and thus fitting to linear hydrocarbons, arrived at best-fit values for v_a of $54\mathring{A}^3$ for CH_3 and $27\mathring{A}^3$ for CH_2. If one fixes these values, then fits the remaining groups represented in Table F.3, the result is Figure F-2. While the fit for linear hydrocarbons is found to be excellent, the fitting for branched hydrocarbons is especially poor, due to the assignation of a large volume to each terminal methyl group. It is advisable, then, to fit the full set of molecules (as in Table F.3), rather than proceed subset-by-subset.

As a final comment, since specific gravity is a function of temperature, v_m (and thus the set of v_a) cannot be expected to hold at all temperatures. The change in volume with temperature at constant pressure is normally expressed as a volumetric coefficient of thermal expansion:

$$\alpha_{v,m} = \frac{1}{v_m}\left(\frac{\partial v_m}{\partial T}\right)\bigg|_P. \tag{F.4.5}$$

Table F.3: Group Contributions to Molecular Volume Obtained for Common Tail Groups at $20°C$

Class	Moiety	$v_a\ \left(\mathring{A}^3\right)$	Class	Moiety	$v_a\ \left(\mathring{A}^3\right)$
Aliphatic	CH3	44	Aromatic	CH	24
Aliphatic	CH2	29	Aromatic	C	6
Aliphatic	CH	14	Ether	O	10
Aliphatic	C	0	Alcohol	OH	17

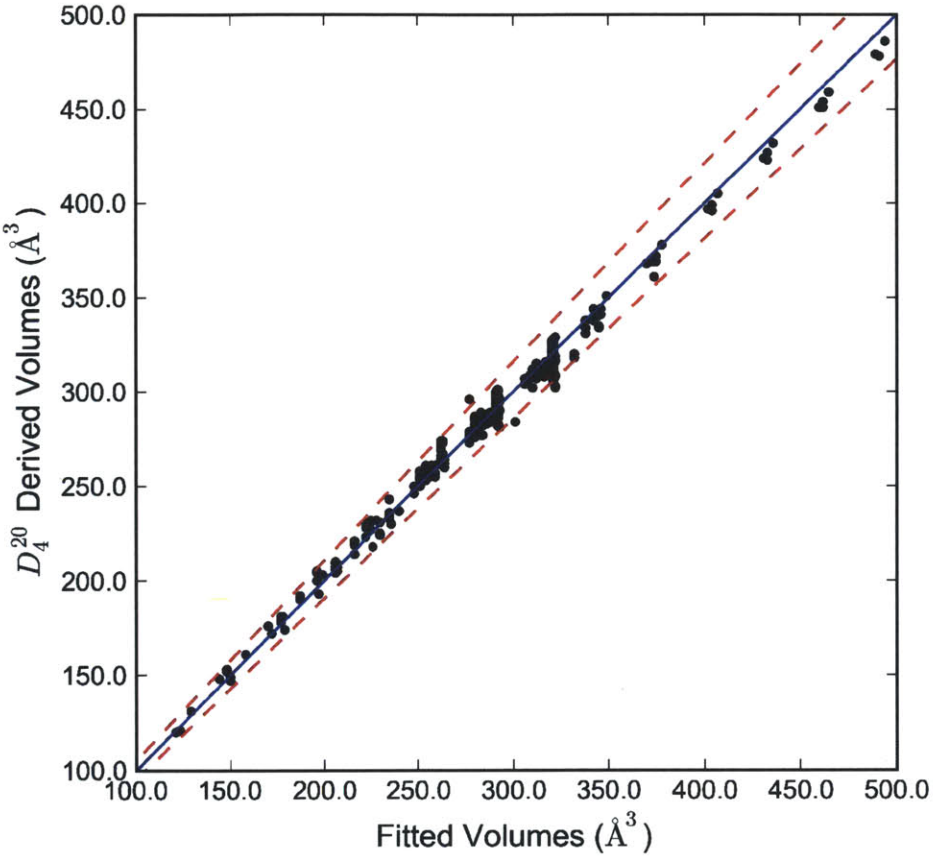


Figure F-1: Comparison of molecular volumes derived from experimental D_4^{20} data with molecular volumes obtained from an additive group-contribution method proposed in the text, where the full set of molecules in Table F.5 are fit simultaneously to obtain the group contributions. Each black dot represents one molecule, a point on the blue line represents equivalence between experimentally-derived and fitted volumes, and a point lying between the red dashed lines and the blue line has less than 5% error relative to the experimentally-derived volume. Note the good agreement obtained without recourse to more advanced group-contribution methods (e.g., those involving topological indices).

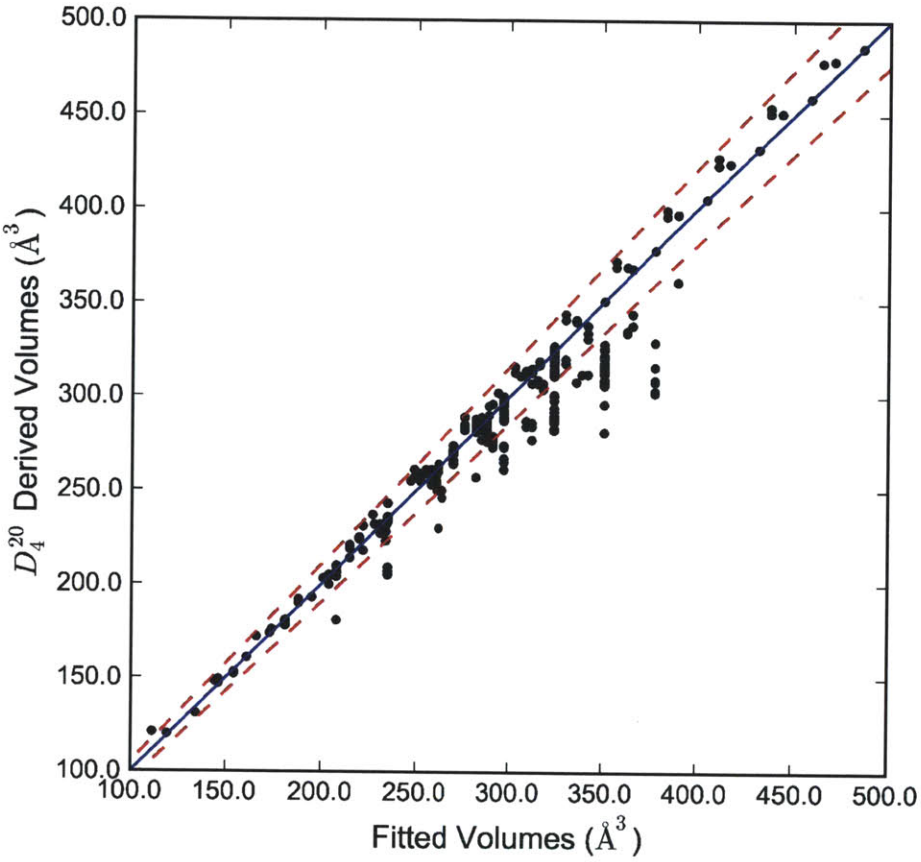


Figure F-2: Comparison of molecular volumes derived from experimental D_4^{20} data with molecular volumes obtained from an additive group-contribution method proposed in the text, where linear hydrocarbons are fit prior to the remaining entries in Table F.5 to obtain the group contributions. Each black dot represents one molecule, a point on the blue line represents equivalence between experimentally-derived and fitted volumes, and a point lying between the red dashed lines and the blue line has less than 5% error relative to the experimentally-derived volume. Note the overestimation for a large set of points: these correspond to branched hydrocarbons possessing multiple terminal methyl groups, indicating too large of a group contribution for CH_3.

Table F.4: Values of $10^3\alpha_{v,m}$ for Various Temperatures[4]

Molecule	$20°C$	$50°C$	$100°C$	Molecule	$20°C$	$50°C$	$100°C$
Hexane	1.36	1.51	-	Dodecane	0.98	1.02	1.1
Octane	1.14	1.225	1.42	Tetradecane	0.89	0.96	1.035
Decane	1.02	1.095	1.21	Hexadecane	0.855	0.92	0.985

The approximate volume change, Δv_m, for a change in temperature ΔT is

$$\Delta v_m \approx \alpha_{v,m} v_m \Delta T, \tag{F.4.6}$$

Table F.4 contains representative values for $\alpha_{v,m}$ for linear alkanes. It can be seen that $\alpha_{v,m}$ for liquid hydrocarbons (and other liquids) is typically around $1 \times 10^{-3}/°C$. (Although $\alpha_{v,m}$ does change with temperature, the change is fairly minor.) Consider that the maximum value for v_a in Table F.3 is $44\mathring{A}^3$. For $\Delta T = 5°C$, Δv_m due to this group is $0.2\mathring{A}^3$, a minor error. For $\Delta T = 40°C$, the error is $\Delta v_m = 1.8\mathring{A}^3$, about 4% error. Consider a spherical micelle of aggregation number 45 possessing a C_{11} tail. The error in aggregation number is approximately 4% at this higher temperature.

Table F.5: Density Fitting Results

#*	Molecule	$\rho_{liq,model}$†	$\rho_{liq,CRC}$†	$\frac{\Delta\rho}{\rho_{liq,CRC}}$	$\underline{V}_{model}$‡	$\underline{V}_{CRC}$‡	$\Delta\underline{V}$‡
Alkane Gases/Liquids Table (Table I(a), pg. 3 in Ref [265])							
42	2,2,4-trimethylpentane	0.721	0.692	0.042	263	274	-11
47	2,2-dimethylhexane	0.721	0.695	0.037	263	273	-10
48	2,5-dimethylhexane	0.724	0.694	0.044	262	274	-12
50	2,4-dimethylhexane	0.724	0.700	0.034	262	271	-9
51	2,2,3-trimethylpentane	0.721	0.716	0.007	263	265	-2
53	3,3-dimethylhexane	0.721	0.710	0.016	263	267	-4
54	2,3,4-trimethylpentane	0.724	0.719	0.007	262	264	-2
56	2,3,3-trimethylpentane	0.721	0.726	-0.007	263	261	2
57	2,3-dimethylhexane	0.724	0.712	0.017	262	266	-4
58	3-ethyl-2-methylpentane	0.724	0.719	0.007	262	264	-2
61	2-methylheptane	0.724	0.698	0.037	262	272	-10
62	4-methylheptane	0.724	0.705	0.028	262	269	-7
63	3,4-dimethylhexane	0.724	0.719	0.007	262	264	-2
65	3-ethyl-3-methylpentane	0.721	0.727	-0.008	263	261	2
66	3-ethylhexane	0.724	0.714	0.015	262	266	-4
67	3-methylheptane	0.724	0.706	0.026	262	269	-7
76	2,2,4,4-tetramethylpentane	0.727	0.719	0.010	293	296	-3
79	2,2,5-trimethylhexane	0.729	0.707	0.031	292	301	-9
82	*n*-octane	0.724	0.703	0.031	262	270	-8
84	2,2,4-trimethylhexane	0.729	0.716	0.019	292	298	-6
87	2,4,4-trimethylhexane	0.729	0.724	0.008	292	294	-2
89	2,3,5-trimethylhexane	0.732	0.722	0.014	291	295	-4
91	2,2-dimethylheptane	0.729	0.710	0.027	292	300	-8
92	2,2,3,4-tetramethylpentane	0.729	0.739	-0.013	292	288	4
93	2,4-dimethylheptane	0.732	0.716	0.022	291	297	-6
95	2,2,3-trimethylhexane	0.729	0.729	0.000	292	292	0
96	4-ethyl-2-methylhexane	0.732	0.723	0.012	291	295	-4
97	3-ethyl-2,2-dimethylpentane	0.729	0.735	-0.007	292	290	2
98	4,4-dimethylheptane	0.729	0.725	0.006	292	294	-2
99	2,6-dimethylheptane	0.732	0.709	0.032	291	300	-9
100	2,5-dimethylheptane	0.732	0.715	0.024	291	298	-7
101	3,5-dimethylheptane	0.732	0.723	0.012	291	295	-4
104	2,4-dimethyl-3-ethylpentane	0.732	0.738	-0.008	291	289	2
106	3,3-dimethylheptane	0.729	0.725	0.006	292	294	-2
107	2,2,5,5-tetramethylhexane	0.734	0.719	0.021	322	329	-7
108	2,3,3-trimethylhexane	0.729	0.738	-0.012	292	289	3
109	3-ethyl-2-methylhexane	0.732	0.731	0.001	291	291	0
111	2,3,4-trimethylhexane	0.732	0.739	-0.010	291	288	3
114	2,2,3,3-tetramethylpentane	0.727	0.757	-0.039	293	281	12
115	4-ethyl-3-methylhexane	0.732	0.742	-0.014	291	287	4

Continued on Next Page

Table F.5 – Continued from Previous Page

#*	Molecule	$\rho_{liq,model}$ †	$\rho_{liq,CRC}$ †	$\frac{\Delta\rho}{\rho_{liq,CRC}}$	$\underline{V}_{model}$ ‡	$\underline{V}_{CRC}$ ‡	$\Delta\underline{V}$ ‡
116	3,3,4-trimethylhexane	0.729	0.745	-0.021	292	286	6
117	2,3-dimethylheptane	0.732	0.726	0.008	291	293	-2
118	3,4-dimethylheptane	0.732	0.731	0.001	291	291	0
119	3-ethyl-3-methylhexane	0.729	0.741	-0.016	292	287	5
120	4-ethylheptane	0.732	0.730	0.003	291	292	-1
121	2,3,3,4-tetramethylpentane	0.729	0.755	-0.034	292	282	10
122	2,3-dimethyl-3-ethylpentane	0.729	0.754	-0.033	292	282	10
125	4-methyloctane	0.732	0.720	0.017	291	296	-5
127	3-ethylheptane	0.732	0.727	0.007	291	293	-2
128	2-methyloctane	0.732	0.713	0.026	291	299	-8
130	3-methyloctane	0.732	0.721	0.016	291	296	-5
131	2,4,6-trimethylheptane	0.738	0.722	0.022	320	327	-7
133	3,3-diethylpentane	0.729	0.754	-0.032	292	283	9
134	2,2-dimethyl-4-ethylhexane	0.736	0.733	0.004	321	322	-1
135	2,2,4-trimethylheptane	0.736	0.727	0.012	321	325	-4
136	2,2,4,5-tetramethylhexane	0.736	0.735	0.001	321	321	0
137	2,2,5-trimethylheptane	0.736	0.726	0.014	321	325	-4
138	2,2,6-trimethylheptane	0.736	0.719	0.023	321	328	-7
139	2,2,3,5-tetramethylhexane	0.736	0.738	-0.002	321	320	1
144	*n*-nonane	0.732	0.718	0.020	291	297	-6
149	2,5,5-trimethylheptane	0.736	0.737	-0.001	321	321	0
150	2,4,4-trimethylheptane	0.736	0.733	0.004	321	322	-1
151	2,3,3,5-tetramethylhexane	0.736	0.746	-0.013	321	317	4
152	2,2,4,4-tetramethylhexane	0.734	0.747	-0.018	322	316	6
155	2,2,3,4-tetramethylhexane	0.736	0.755	-0.025	321	313	8
156	2,2-dimethyloctane	0.736	0.724	0.016	321	326	-5
157	3-ethyl-2,2,4-trimethylpentane	0.736	0.757	-0.028	321	312	9
158	3,3,5-trimethylheptane	0.736	0.743	-0.009	321	318	3
159	2,3,6-trimethylheptane	0.738	0.734	0.005	320	322	-2
160	2,4-dimethyloctane	0.738	0.725	0.019	320	326	-6
162	*d,l*-2,5-dimethyloctane	0.738	0.737	0.002	320	321	-1
166	2,3,5-trimethylheptane	0.738	0.741	-0.004	320	319	1
167	2,5-dimethyl-3-ethylhexane	0.738	0.741	-0.004	320	319	1
168	2,4,5-trimethylheptane	0.738	0.741	-0.004	320	319	1
169	2,4-dimethyl-3-isopropylpentane	0.738	0.758	-0.026	320	312	8
170	2,2,3-trimethylheptane	0.736	0.742	-0.008	321	318	3
171	2,4-dimethyl-4-ethylhexane	0.736	0.747	-0.015	321	316	5
172	2,2-dimethyl-3-ethylhexane	0.736	0.749	-0.017	321	315	6
173	2,2,3,4,4-pentamethylpentane	0.734	0.767	-0.043	322	308	14
175	5-ethyl-2-methylheptane	0.738	0.736	0.003	320	321	-1
176	2,7-dimethyloctane	0.738	0.724	0.020	320	326	-6
177	3,6-dimethyloctane	0.738	0.736	0.003	320	321	-1

Continued on Next Page

Table F.5 – Continued from Previous Page

#*	Molecule	$\rho_{liq,model}$ †	$\rho_{liq,CRC}$ †	$\frac{\Delta\rho}{\rho_{liq,CRC}}$	$\underline{V}_{model}$ ‡	$\underline{V}_{CRC}$ ‡	$\Delta\underline{V}$ ‡
178	3,5-dimethyloctane	0.738	0.736	0.003	320	321	-1
179	4-isopropylheptane	0.738	0.741	-0.004	320	319	1
180	2,3,3-trimethylheptane	0.736	0.749	-0.017	321	316	5
181	4-ethyl-2-methylheptane	0.738	0.736	0.003	320	321	-1
182	2,6-dimethyloctane	0.738	0.728	0.014	320	324	-4
183	2,2,3,3-tetramethylhexane	0.734	0.764	-0.040	322	309	13
185	4,4-dimethyloctane	0.736	0.735	0.002	321	322	-1
186	2,3,4,5-tetramethylhexane	0.738	0.757	-0.025	320	312	8
187	5-ethyl-3-methylheptane	0.738	0.737	0.002	320	321	-1
188	3,3-dimethyloctane	0.736	0.739	-0.004	321	320	1
189	4,5-dimethyloctane	0.738	0.746	-0.010	320	317	3
190	3,4-diethylhexane	0.738	0.754	-0.021	320	313	7
191	4-propylheptane	0.738	0.736	0.003	320	321	-1
194	2,3,4,4-tetramethylhexane	0.736	0.764	-0.036	321	309	12
195	2,3,4-trimethylheptane	0.738	0.751	-0.017	320	315	5
196	3-isopropyl-2-methylhexane	0.738	0.751	-0.017	320	315	5
198	3-ethyl-3-methylheptane	0.736	0.750	-0.019	321	315	6
199	2,4-dimethyl-3-ethylhexane	0.738	0.759	-0.027	320	311	9
200	3,4,4-trimethylheptane	0.736	0.757	-0.028	321	312	9
201	3,3,4-trimethylheptane	0.736	0.757	-0.028	321	312	9
202	3,4,5-trimethylheptane	0.738	0.759	-0.027	320	311	9
203	2,3-dimethyl-4-ethylhexane	0.738	0.759	-0.027	320	311	9
205	2,3-dimethyloctane	0.738	0.738	0.001	320	320	0
207	2,3,3,4-tetramethylhexane	0.736	0.769	-0.043	321	307	14
208	3,3-dimethyl-4-ethylhexane	0.736	0.764	-0.037	321	309	12
209	5-methylnonane	0.738	0.733	0.008	320	323	-3
210	4-methylnonane	0.738	0.732	0.008	320	323	-3
211	3-ethyl-2-methylheptane	0.738	0.746	-0.010	320	317	3
212	3,4-dimethyloctane	0.738	0.746	-0.010	320	317	3
215	2,2,3,3,4-pentamethylpentane	0.734	0.780	-0.059	322	303	19
217	3,3-diethylhexane	0.736	0.767	-0.040	321	308	13
218	2-methylnonane	0.738	0.728	0.014	320	325	-5
220	3-ethyl-4-methylheptane	0.738	0.753	-0.019	320	314	6
221	4-ethyl-3-methylheptane	0.738	0.753	-0.019	320	314	6
222	4-ethyl-4-methylheptane	0.736	0.752	-0.021	321	314	7
224	3-methylnonane	0.738	0.733	0.007	320	322	-2
225	3-ethyloctane	0.738	0.740	-0.002	320	319	1
226	4-ethyloctane	0.738	0.740	-0.002	320	319	1
227	3-ethyl-2,2,3-trimethylpentane	0.734	0.781	-0.060	322	303	19
231	2,3-dimethyl-3-ethylhexane	0.736	0.765	-0.038	321	309	12
233	3,4-dimethyl-3-ethylhexane	0.736	0.772	-0.047	321	306	15
234	3,3,4,4-tetramethylhexane	0.734	0.782	-0.062	322	302	20

Continued on Next Page

Table F.5 – Continued from Previous Page

#*	Molecule	$\rho_{liq,model}$ †	$\rho_{liq,CRC}$ †	$\frac{\Delta\rho}{\rho_{liq,CRC}}$	$\underline{V}_{model}$ ‡	$\underline{V}_{CRC}$ ‡	$\Delta\underline{V}$ ‡
239	*n*-decane	0.738	0.730	0.011	320	324	-4
247	*n*-undecane	0.744	0.740	0.005	349	351	-2
254	*n*-dodecane	0.748	0.749	0.000	378	378	0
261	*n*-tridecane	0.752	0.756	-0.006	407	405	2
265	*n*-tetradecane	0.756	0.763	-0.009	436	432	4
268	*n*-pentadecane	0.759	0.768	-0.013	465	459	6
Alkane Solids Table (Table I(b), pg. 8 in Ref [265])							
2	*n*-hexadecane	0.761	0.773	-0.016	494	486	8
Aromatic Liquids Table (Table IV(a), pg. 35 in Ref [265])							
1	benzene	0.901	0.879	0.025	144	148	-4
2	toluene	0.900	0.867	0.038	170	176	-6
3	ethylbenzene	0.886	0.867	0.022	199	203	-4
4	1,4-xylene	0.899	0.861	0.045	196	205	-9
5	1,3-xylene	0.899	0.864	0.041	196	204	-8
6	1,2-xylene	0.899	0.880	0.022	196	200	-4
7	isopropylbenzene (cumene)	0.875	0.862	0.016	228	232	-4
8	*n*-propylbenzene	0.875	0.862	0.016	228	232	-4
9	1-ethyl-3-methylbenzene	0.887	0.865	0.026	225	231	-6
10	1-ethyl-4-methylbenzene	0.887	0.861	0.030	225	232	-7
11	1,3,5-trimethylbenzene (mesitylene)	0.899	0.865	0.039	222	231	-9
12	1-ethyl-2-methylbenzene	0.887	0.881	0.007	225	227	-2
13	*tert*-butylbenzene	0.864	0.866	-0.003	258	257	1
14	1,2,4-trimethylbenzene	0.899	0.876	0.027	222	228	-6
15	isobutylbenzene	0.867	0.853	0.016	257	261	-4
16	*sec*-butylbenzene	0.867	0.862	0.006	257	259	-2
17	3-isopropyl-1-methylbenzene	0.877	0.861	0.019	254	259	-5
18	1,2,3-trimethylbenzene	0.899	0.894	0.005	222	223	-1
21	4-isopropyl-1-methylbenzene	0.877	0.854	0.028	254	261	-7
22	2-isopropyl-1-methylbenzene	0.877	0.877	0.001	254	254	0
23	1,3-diethylbenzene	0.877	0.864	0.016	254	258	-4
24	1-methyl-3-propylbenzene	0.877	0.861	0.019	254	259	-5
26	*n*-butylbenzene	0.867	0.860	0.008	257	259	-2
27	1-methyl-4-propylbenzene	0.877	0.858	0.022	254	260	-6
28	1,2-diethylbenzene	0.877	0.880	-0.003	254	253	1
29	1,4-diethylbenzene	0.877	0.862	0.018	254	259	-5
30	1,3-dimethyl-5-ethylbenzene	0.888	0.865	0.027	251	258	-7
31	1-methyl-2-propylbenzene	0.877	0.874	0.004	254	255	-1
32	2,2-dimethyl-1-phenylpropane	0.858	0.858	0.000	287	287	0
33	1,4-dimethyl-2-ethylbenzene	0.888	0.877	0.012	251	254	-3
35	3-methyl-2-phenylbutane	0.861	0.870	-0.011	286	283	3
37	1,3-dimethyl-4-ethylbenzene	0.888	0.876	0.013	251	254	-3
38	3-tert-butyl-1-methylbenzene	0.867	0.866	0.001	284	284	0

Continued on Next Page

Table F.5 – Continued from Previous Page

#*	Molecule	$\rho_{liq,model}$ †	$\rho_{liq,CRC}$ †	$\frac{\Delta\rho}{\rho_{liq,CRC}}$	$\underline{V}_{model}$ ‡	$\underline{V}_{CRC}$ ‡	$\Delta\underline{V}$ ‡
39	1,2-dimethyl-4-ethylbenzene	0.888	0.874	0.015	251	255	-4
40	1,3-dimethyl-2-ethylbenzene	0.888	0.890	-0.003	251	250	1
41	3-phenylpentane	0.861	0.865	-0.005	286	285	1
42	1-ethyl-3-isopropylbenzene	0.870	0.859	0.013	283	287	-4
43	2-methyl-2-phenylbutane	0.823	0.874	-0.058	301	284	17
44	4-tert-butyl-1-methylbenzene	0.867	0.861	0.007	284	286	-2
45	1-ethyl-2-isopropylbenzene	0.870	0.888	-0.020	283	277	6
46	2-phenylpentane	0.861	0.858	0.004	286	287	-1
47	1,2-dimethyl-3-ethylbenzene	0.888	0.892	-0.005	251	250	1
48	3-sec-butyl-1-methylbenzene	0.870	0.858	0.014	283	287	-4
49	3-isobutyl-1-methylbenzene	0.870	0.854	0.019	283	288	-5
50	*d*-2-methyl-1-phenylbutane	0.861	0.862	-0.001	286	286	0
51	1,3-dimethyl-5-isopropylbenzene	0.879	0.859	0.023	280	287	-7
53	4-isobutyl-1-methylbenzene	0.870	0.852	0.021	283	289	-6
54	2-sec-butyl-1-methylbenzene	0.870	0.873	-0.004	283	282	1
55	2-isobutyl-1-methylbenzene	0.870	0.865	0.006	283	285	-2
56	1,4-dimethyl-2-isopropylbenzene	0.879	0.874	0.006	280	282	-2
57	1-ethyl-4-isopropylbenzene	0.870	0.858	0.013	283	287	-4
58	*d,l*-2-methyl-1-phenylbutane	0.861	0.859	0.002	286	287	-1
59	1,2,3,5-tetramethylbenzene	0.899	0.890	0.010	248	250	-2
60	3-methyl-1-phenylbutane	0.861	0.856	0.006	286	288	-2
61	1,3-dimethyl-2-isopropylbenzene	0.879	0.890	-0.012	280	277	3
62	1,3-dimethyl-4-isopropylbenzene	0.879	0.869	0.012	280	283	-3
64	4-*sec*-butyl-1-methylbenzene	0.870	0.865	0.006	283	285	-2
65	2-*tert*-butyl-1-methylbenzene	0.867	0.890	-0.026	284	277	7
66	3,5-diethyl-1-methylbenzene	0.879	0.863	0.019	280	285	-5
67	2-butyl-1-methylbenzene	0.870	0.872	-0.003	283	282	1
68	1-ethyl-3-propylbenzene	0.870	0.861	0.011	283	286	-3
69	1,2-dimethyl-4-isopropylbenzene	0.879	0.870	0.011	280	283	-3
71	1,2-dimethyl-3-isopropylbenzene	0.879	0.888	-0.010	280	277	3
72	1-ethyl-2-propylbenzene	0.870	0.874	-0.005	283	282	1
73	1,3-di-isopropylbenzene	0.864	0.856	0.009	312	315	-3
74	1,2-diethyl-4-methylbenzene	0.879	0.876	0.003	280	281	-1
75	1,2-di-isopropylbenzene	0.864	0.877	-0.015	312	307	5
76	1,4-dimethyl-2-propylbenzene	0.879	0.872	0.009	280	282	-2
77	1,2,3,4-tetramethylbenzene	0.899	0.905	-0.007	248	246	2
78	1-ethyl-4-propylbenzene	0.870	0.859	0.012	283	286	-3
79	3-butyl-1-methylbenzene	0.870	0.859	0.013	283	287	-4
80	2,4-diethyl-1-methylbenzene	0.879	0.875	0.005	280	281	-1
81	*n*-pentylbenzene	0.861	0.858	0.003	286	287	-1
82	3-methyl-3-phenylpentane	0.853	0.875	-0.026	316	308	8
83	1,3-dimethyl-5-tert-butyl-benzene	0.869	0.864	0.006	310	312	-2

Continued on Next Page

Table F.5 – Continued from Previous Page

#*	Molecule	$\rho_{liq,model}$†	$\rho_{liq,CRC}$†	$\frac{\Delta\rho}{\rho_{liq,CRC}}$	$\underline{V}_{model}$‡	$\underline{V}_{CRC}$‡	$\Delta\underline{V}$‡
84	1,3-dimethyl-4-propylbenzene	0.879	0.872	0.008	280	282	-2
85	1,2-diethyl-3-methylbenzene	0.879	0.891	-0.013	280	276	4
86	4-butyl-1-methylbenzene	0.870	0.857	0.015	283	287	-4
87	2,5-diethyl-1-methylbenzene	0.879	0.876	0.004	280	281	-1
89	1,3-diethyl-2-propylbenzene	0.866	0.886	-0.022	338	331	7
90	2,6-diethyl-1-methylbenzene	0.879	0.891	-0.013	280	276	4
91	1,2-dimethyl-4-propylbenzene	0.879	0.871	0.009	280	282	-2
92	1,3-dimethyl-5-propylbenzene	0.879	0.861	0.021	280	286	-6
93	2-methyl-3-phenylpentane	0.855	0.868	-0.014	315	311	4
95	1,4-di-isopropylbenzene	0.864	0.857	0.008	312	315	-3
96	1,2-dimethyl-3-propylbenzene	0.879	0.886	-0.008	280	278	2
97	1-*tert*-butyl-4-ethylbenzene	0.861	0.863	-0.003	313	312	1
98	*d,l*-3-phenylhexane	0.855	0.860	-0.005	315	313	2
99	2-ethyl-1,3,5-trimethylbenzene	0.889	0.883	0.007	277	279	-2
100	3-ethyl-4-isopropyl-1-methylbenzene	0.872	0.872	0.000	309	309	0
101	5-ethyl-1,2,4-trimethylbenzene	0.889	0.833	0.067	277	296	-19
102	6-ethyl-1,2,4-trimethylbenzene	0.889	0.890	-0.001	277	277	0
103	2-phenylhexane	0.855	0.860	-0.005	315	313	2
104	2-methyl-1-phenylpentane	0.855	0.862	-0.008	315	312	3
105	4-isopropyl-1-propylbenzene	0.864	0.861	0.003	312	313	-1
107	5-ethyl-1,2,3-trimethylbenzene	0.889	0.886	0.003	277	278	-1
108	3-ethyl-1,2,4-trimethylbenzene	0.889	0.895	-0.007	277	275	2
109	1,2,4-triethylbenzene	0.872	0.879	-0.008	309	307	2
113	4-ethyl-1,2,3-trimethylbenzene	0.889	0.902	-0.015	277	273	4
114	1,4-dipropylbenzene	0.864	0.856	0.009	312	315	-3
115	3-methyl-1-phenylpentane	0.855	0.860	-0.006	315	313	2
116	2-propyl-1,3,5-trimethylbenzene	0.881	0.878	0.003	306	307	-1
118	3-*tert*-butyl-1-isopropylbenzene	0.856	0.851	0.006	342	344	-2
119	1-methyl-3-pentylbenzene	0.864	0.859	0.005	312	314	-2
120	4-*tert*-butyl-1-isopropylbenzene	0.856	0.866	-0.012	342	338	4
121	2-methyl-2-phenylhexane	0.849	0.874	-0.029	345	335	10
122	2,4-di-isopropyl-1-methylbenzene	0.866	0.866	0.000	338	338	0
123	3-methyl-3-phenylhexane	0.849	0.878	-0.033	345	334	11
124	*n*-hexylbenzene	0.855	0.857	-0.002	315	314	1
125	3-phenylheptane	0.851	0.861	-0.011	344	340	4
126	2,6-di-isopropyl-1-methylbenzene	0.866	0.877	-0.012	338	334	4
127	5-propyl-1,2,4-trimethylbenzene	0.881	0.887	-0.007	306	304	2
130	2-phenylheptane	0.851	0.861	-0.012	344	340	4
139	1,4-di-*sec*-butylbenzene	0.854	0.859	-0.006	370	368	2
141	3-ethyl-3-phenylhexane	0.845	0.875	-0.034	374	361	13
145	*n*-heptylbenzene	0.851	0.859	-0.010	344	341	3
146	1-methylnapthalene	1.027	1.020	0.006	230	231	-1

Continued on Next Page

Table F.5 – Continued from Previous Page

#*	Molecule	$\rho_{liq,model}$ †	$\rho_{liq,CRC}$ †	$\frac{\Delta\rho}{\rho_{liq,CRC}}$	$\underline{V}_{model}$ ‡	$\underline{V}_{CRC}$ ‡	$\Delta\underline{V}$ ‡
151	2-ethylnaphthalene	1.002	0.992	0.010	259	261	-2
152	1-ethylnaphthalene	1.002	1.008	-0.006	259	257	2
153	1,7-dimethylnaphthalene	1.013	1.011	0.002	256	256	0
154	1,6-dimethylnaphthalene	1.013	1.003	0.010	256	259	-3
155	1,3-dimethylnaphthalene	1.013	1.006	0.007	256	258	-2
156	*n*-octylbenzene	0.847	0.856	-0.010	373	369	4
158	1-isopropylnaphthalene	0.982	0.996	-0.014	288	284	4
159	1,4-dimethylnaphthalene	1.013	1.017	-0.003	256	255	1
160	1,1-diphenylethane	0.976	1.003	-0.027	310	302	8
161	2-isopropylnaphthalene	0.982	0.979	0.002	288	289	-1
162	2-propylnaphthalene	0.982	0.977	0.005	288	289	-1
163	1-propylnaphthalene	0.982	0.992	-0.010	288	285	3
164	1,3,7-trimethylnaphthalene	1.003	1.007	-0.004	282	281	1
165	1-isopropyl-7-methylnaphthalene	0.975	0.983	-0.009	314	311	3
166	*n*-nonylbenzene	0.844	0.856	-0.014	402	397	5
167	2-butylnaphthalene	0.965	0.967	-0.002	317	316	1
168	2-*tert*-butylnaphthalene	0.927	0.969	-0.043	332	318	14
169	1-*tert*-butylnaphthalene	0.927	0.963	-0.037	332	320	12
170	1-butylnaphthalene	0.965	0.977	-0.012	317	313	4
172	*n*-decylbenzene	0.841	0.856	-0.016	431	424	7
173	1-pentylnaphthalene	0.952	0.966	-0.014	346	341	5
174	2-pentylnaphthalene	0.952	0.956	-0.005	346	344	2
175	*n*-undecylbenzene	0.839	0.855	-0.019	460	451	9
176	1-hexylnaphthalene	0.940	0.957	-0.017	375	369	6
177	2-hexylnaphthalene	0.940	0.948	-0.008	375	372	3
178	*n*-dodecylbenzene	0.837	0.855	-0.021	489	479	10
179	1-heptylnaphthalene	0.930	0.949	-0.020	404	396	8
180	2-heptylnaphthalene	0.930	0.941	-0.011	404	399	5
181	*n*-tridecylbenzene	0.835	0.855	-0.023	518	506	12
182	1-octylnaphthalene	0.922	0.943	-0.022	433	423	10
183	2-octylnaphthalene	0.922	0.936	-0.015	433	427	6
184	1-nonylnaphthalene	0.914	0.937	-0.024	462	451	11
185	2-nonylnaphthalene	0.914	0.930	-0.016	462	454	8
186	1-decylnaphthalene	0.908	0.932	-0.026	491	478	13
Alcohol Liquids Table (Table VI(a), pg. 80 in Ref [265])							
7	2-butanol	0.832	0.807	0.031	148	153	-5
8	2-methyl-2-butanol	0.822	0.809	0.017	178	181	-3
10	2-methyl-1-propanol	0.832	0.802	0.037	148	153	-5
12	*d,l*-3-methyl-2-butanol	0.827	0.818	0.011	177	179	-2
13	3-pentanol	0.827	0.820	0.008	177	178	-1
14	1-butanol	0.832	0.810	0.027	148	152	-4
15	*d,l*-2-pentanol	0.827	0.809	0.022	177	181	-4

Continued on Next Page

Table F.5 – Continued from Previous Page

#*	Molecule	$\rho_{liq,model}$ †	$\rho_{liq,CRC}$ †	$\frac{\Delta\rho}{\rho_{liq,CRC}}$	$\underline{V}_{model}$ ‡	$\underline{V}_{CRC}$ ‡	$\Delta\underline{V}$ ‡
16	3,3-dimethyl-2-butanol	0.820	0.818	0.001	207	207	0
17	2,3-dimethyl-2-butanol	0.820	0.821	-0.001	207	207	0
19	2-methyl-2-pentanol	0.820	0.813	0.008	207	209	-2
20	2-methyoxyethanol	0.980	0.965	0.015	129	131	-2
21	2-methyl-3-pentanol	0.824	0.825	-0.001	206	206	0
23	2-methyl-1-butanol	0.827	0.819	0.009	177	179	-2
25	*d,l*-4-methyl-2-pentanol	0.824	0.807	0.020	206	210	-4
26	3-methyl-1-butanol	0.827	0.809	0.022	177	181	-4
29	2-ethoxyethanol	0.947	0.930	0.019	158	161	-3
30	3-hexanol	0.824	0.818	0.006	206	207	-1
31	2,2-dimethyl-1-butanol	0.820	0.828	-0.010	207	205	2
32	1-pentanol	0.827	0.815	0.015	177	180	-3
34	2,4-dimethyl-3-pentanol	0.821	0.829	-0.009	235	233	2
36	2-isopropoxyethanol	0.925	0.903	0.024	187	192	-5
37	3-ethyl-3-pentanol	0.818	0.839	-0.025	236	230	6
41	2-methyl-1-pentanol	0.824	0.821	0.004	206	207	-1
42	2-ethylbutanol	0.824	0.833	-0.012	206	204	2
44	2-*n*-propoxyethanol	0.925	0.911	0.015	187	190	-3
46	3-methyl-1-pentanol	0.824	0.824	-0.001	206	206	0
47	4-methyl-1-pentanol	0.824	0.813	0.013	206	209	-3
48	*d,l*-4-heptanol	0.821	0.818	0.003	235	236	-1
49	1-hexanol	0.824	0.819	0.006	206	207	-1
50	*d,l*-2-heptanol	0.821	0.817	0.005	235	236	-1
51	2-isobutoxyethanol	0.909	0.890	0.021	216	221	-5
52	2-*sec*-butoxyethanol	0.909	0.897	0.013	216	219	-3
53	2,4-dimethyl-1-pentanol	0.821	0.793	0.035	235	243	-8
55	2-methyl-1-hexanol	0.821	0.827	-0.007	235	233	2
57	*d,l*-4-methyl-1-hexanol	0.821	0.824	-0.003	235	234	1
61	2-*n*-butoxyethanol	0.909	0.919	-0.011	216	214	2
63	2,6-dimethyl-4-heptanol	0.818	0.811	0.008	293	295	-2
75	1-heptanol	0.821	0.822	-0.002	235	235	0
80	*d,l*-2-octanol	0.819	0.820	-0.002	264	264	0
87	2-ethyl-1-hexanol	0.819	0.833	-0.016	264	260	4
96	2-(2-methoxyethoxy)-ethanol	1.013	1.033	-0.020	197	193	4
98	1-octanol	0.819	0.825	-0.007	264	262	2
100	2-(2-ethoxyethoxy)-ethanol	0.986	1.021	-0.035	226	218	8
101	glycol (1,2-ethanediol)	1.120	1.114	0.006	92	93	-1
104	benzyl alcohol	1.044	1.045	-0.001	172	172	0
105	*d,l*-1,3-butanediol	0.998	1.005	-0.008	150	149	1
106	*d,l*-2-decanol	0.816	0.825	-0.011	322	319	3
107	1-nonanol	0.818	0.827	-0.011	293	290	3
108	1,3-propanediol	1.044	1.054	-0.009	121	120	1

Continued on Next Page

Table F.5 – Continued from Previous Page

#*	Molecule	$\rho_{liq,model}$†	$\rho_{liq,CRC}$†	$\frac{\Delta\rho}{\rho_{liq,CRC}}$	$\underline{V}_{model}$‡	$\underline{V}_{CRC}$‡	$\Delta\underline{V}$‡
111	1-phenyl-*n*-propyl alcohol	0.983	1.004	-0.020	230	225	5
116	α-isopropylbenzyl alcohol	0.963	0.977	-0.014	259	255	4
118	2-(2-*n*-butoxyethoxy)-ethanol	0.949	0.955	-0.007	284	282	2
119	1,4-butanediol	0.998	1.017	-0.019	150	147	3
121	1-decanol	0.816	0.829	-0.016	322	317	5
123	3-phenylpropanol	0.983	1.008	-0.024	230	224	6
124	1,5-pentanediol	0.966	0.992	-0.026	179	174	5
128	2-benzyloxyethanol	1.053	1.068	-0.014	240	237	3
131	glycerol	1.243	1.261	-0.014	123	121	2

*Index of molecule in Ref. [265].
†Units of g/cm^3.
‡Units of $Å^3$.

F.5 Surfactant Tail Aqueous Solubility

F.5.1 Introduction

In Section 2.3.3 of Chapter 2, I demonstrated that the transfer free-energy contribution to the free energy of micellization for a particular surfactant is related to the hypothetical solubility limit of the tail of that surfactant in aqueous solution. This relationship is described by Eq. 2.3.15, which I reproduce here for convenience:

$$g_{tr} = k_B T \ln X^{sat}_{tail,aq}, \tag{F.5.1}$$

where k_B is the Boltzmann constant, T is the absolute system temperature, and $X^{sat}_{tail,aq}$ is the mole fraction of the surfactant tail at saturation. Assuming that the surfactant solution is dilute, Eq. F.5.1 can be written as

$$g_{tr} = k_B T \ln \left(\frac{S^{sat}_{tail,aq}}{\mathcal{M}_{H_2O}} \right), \tag{F.5.2}$$

where $S^{sat}_{tail,aq}$ is the solubility of the tail in molar units (henceforth notated simply as S), and $\mathcal{M}_{H_2O}$ is the molarity of water at the temperature of interest (see Table F.2 for values for $\mathcal{M}_{H_2O}$).

I describe the solubility limit of the surfactant tail here as "hypothetical" since a surfactant tail is not a complete or whole molecule. For example, a linear alkyl tail possesses only one true terminus (i.e., one methyl group), unlike an oil of the same length (which possesses two methyl groups), since one end of the tail is attached to the surfactant head. As a result, direct experimental values for S for surfactant tails do not exist. Instead, S must be calculated based on a group-contribution method for solubility, where the group contributions have been determined by fitting to solubility data for a large set of organic molecules. In the discussion below, I focus on four promising models: (i) a model for linear alkyl tails derived from the regression equations of Abraham for the Gibbs free energy of solution (MHA),[239] (ii) a variation of the MHA model with different parameter values, which I label MHA2, (iii) a group-contribution method due to Klopman, Wang, and Balthasar (KWB),[5] and (iv) a group-contribution method due to Kühne, Ebert, Kleint, Schmidt, and Schüürmann (KEKSS).[6] Our group has previously used the MHA model for linear alkanes. However, extension to branched surfactants requires further parameter fitting to accommodate nodal carbon groups (i.e., CH and C). In contrast, the KWB and KEKSS models are already equipped with the group contributions for these groups, in addition to providing parameters for a variety of other interesting functional groups (including ethers and alcohols).

Although I focus on four specific models below, there are many others in the literature that have merit. However, these typically incorporate additional molecular descriptors, such as octanol-water partition coefficients and melting points,[343-345] hydrogen-bond acidity and basicity descriptors,[346] or topological indices,[347,348] which complicate the modeling. The use of a model which depends on physical properties is also not recommended for the reason that the surfactant tail is not a whole molecule, and such properties would, as a consequence, also have to be estimated.

F.5.2 Abraham (MHA/MHA2) Models for Aqueous Solubility

From experimental data for linear alkanes up to octane, Abraham developed the following correlation for the Gibbs free energy of solution of the n-alkanes at $25°C$:[239]

$$\Delta G_s^0 \left(liq \rightarrow aq\right) = 2.399 + N\left(0.887 \pm 0.010\right) \left[\frac{kcal}{mol}\right], \tag{F.5.3}$$

where N is the number of carbons in the alkane. Equation F.5.3 represents the difference in reference-state chemical potentials between an oil at infinite dilution in aqueous solution and the same oil in a pure bulk phase. This is the opposite process to that described by g_{tr}, and the two expressions are related via

$$\frac{g_{tr}^{MHA}}{k_B T} = -\frac{\Delta G_s^0 \left(liq \rightarrow aq\right) \left[\frac{kcal}{mol}\right] \cdot 4184 \left[\frac{J}{kcal}\right]}{RT}, \tag{F.5.4}$$

where R is the gas constant ($8.314 \frac{J}{mol-K}$), and $T = 298.15K$. Inserting Eq. F.5.3 into Eq. F.5.4 and evaluating yields

$$\frac{g_{tr}^{MHA}}{k_B T} = -4.049 + N\left(-1.497\right). \tag{F.5.5}$$

Assuming that g_{tr} can be represented by a fully-additive group contribution method, the MHA estimate for g_{tr} is obtained as

$$\frac{g_{tr}^{MHA}}{k_B T} = \sum_a \frac{g_{tr,a}^{MHA}}{k_B T}. \tag{F.5.6}$$

One can then determine the group contributions for CH_3 and CH_2 from Eq. F.5.5, by first rearranging and factoring terms to obtain:

$$\frac{g_{tr}^{MHA}}{k_B T} = 2\left(-3.522\right) + \left(N - 2\right)\left(-1.497\right). \tag{F.5.7}$$

Since each n-alkane has two methyl groups and $N - 2$ methylene groups, comparing Eqs. F.5.6 and F.5.7 yields $g_{tr,CH_3}^{MHA} = -3.522 k_B T$ and $g_{tr,CH_2}^{MHA} = -1.497 k_B T$. These

are the values traditionally used by our group to determine the solubility of linear alkyl tails,[81] and I will refer to the use of these values and Eq. F.5.6 as the MHA model.

In order to apply this model to branched surfactants and alkyl benzene surfactants, additional experimental data must be introduced in order to fit the additional group parameters required. In Section F.5.5 below and, in particular, Table F.9, I present a dataset used for evaluating the various models for solubility in this Appendix. To expand the parameter set of the MHA model, I used a least-squares fitting approach to attempt to reproduce these data, holding the values for CH_3 and CH_2 fixed to the results obtained from the Abraham correlation in Eq. F.5.3. The results of this fitting are presented in Table F.6, together with an alternative model, which I label MHA2, which I developed using the same set of experimental data, but allowing for the fresh fitting of all parameters, including CH_3 and CH_2. This has the benefit of including some longer-chain alkanes in the fitting procedure.

Note that the result of parameter fitting for both the MHA and MHA2 models indicates positive group contributions for CH and C alkane groups. Since each branching of an alkane chain introduces additional methyl groups, and methyl groups have the most negative $g_{tr,a}$ values in the set, this seems to indicate that this value represents too large a contribution in general to the solubility of a molecule. In the KWB and KEKSS models below, the presence of a constant term in addition to the group contributions results in a much smaller contribution from the methyl groups, and negative group contributions from all carbon-based atomic groups. For the reproduction of whole-molecule solubility data, no one model can be considered right or wrong: positive contributions simply indicate a different partitioning of the solubility amongst the groups present. However, since the surfactant tail is not a whole molecule, the replacement of a terminal methyl with a methylene, for example, can have a very strong, model-dependent effect on g_{tr}, thereby affecting micellization predictions. I will revisit this discussion in Section F.5.5.

Table F.6: Parameters for the Abraham Models (MHA and MHA2) at 25°C

Description	Group	$g_{tr,a}^{MHA}/k_BT^*$	$g_{tr,a}^{MHA2}/k_BT^\dagger$
alkane	CH_3	-3.522‡	-4.250
alkane	CH_2	-1.497‡	-1.195
alkane	CH	0.816	2.078
alkane	C	3.395	5.544
aromatic	CH	-1.268	-1.359
aromatic	C	0.914	1.733

* Parameters not marked with ‡are fit to the data in Table F.9.
† All parameters in this column are fit to the data in Table F.9.
‡ Values derived from Ref. [239] (see text).

F.5.3 Klopman, Wang, and Balthasar (KWB) Model for Aqueous Solubility

Klopman, Wang, and Balthasar (KWB) developed a model for aqueous solubility which includes group-contribution parameters for a variety of useful functional groups.[5] Their model consists of an additive group-contribution piece, a constant term applied to all molecules, and an adjustment term which modifies the constant based on the overall class of molecule (specifically, whether it is an alkane or other type of hydrocarbon). Their parameters were fit using linear regression to a database of 469 compounds, from which approximately 25 compounds (about 5% of the set) were randomly withheld during multiple fitting trials to cross-validate the fit.[5]

The KWB model can be expressed as

$$\log S^{KWB} = C_0^{KWB} + A^{KWB} + \sum_a logS_a^{KWB}, \tag{F.5.8}$$

where C_0^{KWB} is the additive constant applied to all molecules, A^{KWB} represents the species-dependent adjustment term, and $logS_a^{KWB}$ is the group contribution of group a to the base-10 logarithm of the solubility S^{KWB} (where S^{KWB} has units of mol/m^3).

Working with Eq. F.5.2, one can derive an expression which relates g_{tr} and

$\log S^{KWB}$:

$$\frac{g_{tr}^{KWB}}{k_B T} = -\ln\left(1000\mathcal{M}_{H_2O}\right) + \left(\ln 10\right)\log S^{KWB}, \tag{F.5.9}$$

where $\mathcal{M}_{H_2O}$ is the molarity of water at the temperature of interest ($25°C$) and the factor of 1000 is needed to convert to mol/m^3.

Substituting Eq. F.5.8 into Eq. F.5.9 and grouping constant and group-additive terms yields:

$$\frac{g_{tr}^{KWB}}{k_B T} = \left[-\ln\left(1000\mathcal{M}_{H_2O}\right) + \left(\ln 10\right)\left(C_0^{KWB} + A^{KWB}\right)\right] + \sum_a \left[\left(\ln 10\right) logS_a^{KWB}\right]. \tag{F.5.10}$$

Drawing an analogy to Eq. F.5.6, one can construct a group contribution model for g_{tr}^{KWB} as follows

$$\frac{g_{tr}^{KWB}}{k_B T} = \frac{g_{tr,0}^{KWB}}{k_B T} + \sum_a \frac{g_{tr,a}^{KWB}}{k_B T}, \tag{F.5.11}$$

where the constant $g_{tr,0}$ term was not present in the MHA model.

Comparing Eqs. F.5.10 and F.5.11, one finds that

$$\frac{g_{tr,0}^{KWB}}{k_B T} = -\ln\left(1000\mathcal{M}_{H_2O}\right) + \left(\ln 10\right)\left(C_0^{KWB} + A^{KWB}\right), \tag{F.5.12}$$

and

$$\frac{g_{tr,a}^{KWB}}{k_B T} = \left(\ln 10\right) logS_a^{KWB}. \tag{F.5.13}$$

In Table F.7, I have collected values for $\log S_a^{KWB}$ of most relevance to this work (including a few extra useful groups, such as the alkenes, for future reference). The constant and adjustment terms can be read from the last three rows of the Table, where the "$\log S_a^{KWB}$" value is to be interpreted as $\left(\ln 10\right) C_0^{KWB}$ and $\left(\ln 10\right) A^{KWB}$, respectively. For $25°C$, from Table F.2, one has that $\mathcal{M}_{H_2O} = 55.345\ mol/L$. Using the appropriate C_0^{KWB} and A^{KWB} values in Table F.7, and applying Eq. F.5.12, one obtains $g_{tr,0}^{KWB}/k_B T = -6.256$ for alkanes (linear or branched) and $g_{tr,0}^{KWB}/k_B T = -3.311$ for all other hydrocarbons (e.g, hydrocarbons including aromatic rings or other functional groups).

Table F.7: Parameters Derived from the Solubility Model of Klopman, Wang, and Balthasar (KWB)[5] at $25°C$

#*	Description	Group	$logS_a^{KWB\dagger}$	$g_{tr,a}^{KWB}/k_BT^{\ddagger}$
1	alkane	CH_3	-0.336	-0.774
2	alkane	CH_2	-0.573	-1.319
3	alkane	CH	-0.606	-1.395
4	alkane	C	-0.785	-1.808
5	alkene	$=CH_2$	-0.687	-1.582
6	alkene	=CH-	-0.323	-0.744
7	alkene	=C-(-)	-0.335	-0.770
14	aromatic	CH	-0.369	-0.850
15	aromatic	C	-0.494	-1.138
26	alcohol (primary)	OH	1.464	3.371
27	alcohol (secondary)	OH	1.563	3.599
28	alcohol (tertiary)	OH	1.089	2.506
35	ether	O	0.852	1.961
41	amine (secondary)	CONH	0.193	0.445
66	A (alkanes)[§]		-1.539	-3.543
67	A (all other hydrocarbons)[§]		-0.260	-0.598
	C_0 constant[§]		3.565	8.209

* Entry number in Table I of Ref. [5].

† Group contribution to the base-10 logarithm of the solubility (units of mol/m^3).

‡ Group contribution to g_{tr}, calculated according to Eq. F.5.13.

§ Constant terms: C_0 is applied to all molecules, the "alkane adjustment" is applied to molecules possessing only alkane groups, and the "other adjustment" is applied to all other hydrocarbons.

F.5.4 Kühne, Ebert, Kleint, Schmidt, and Schüürmann (KEKSS) Model for Aqueous Solubility

The group contribution model for the prediction of the aqueous solubility of organic molecules developed by Kühne, Ebert, Kleint, Schmidt, and Schüürmann (KEKSS) was designed to accommodate solutes that have either a liquid or solid pure state, and is based on a database of 351 liquids and 343 solids.[6] In its most complete form, the KEKSS model is formulated as

$$\log S^{KEKSS} = C_0^{KEKSS} + \sum_a logS_a^{KEKSS} + b^{KEKSS}\left(T_M, \Delta S_M\right) \cdot \left(T_M - T\right), \quad \text{(F.5.14)}$$

where C_0^{KEKSS} is a constant applied to all molecules, $logS_a^{KEKSS}$ is the group contribution of group a to the base-10 logarithm of the solubility S^{KEKSS} (where S^{KEKSS} has units of mol/L), $b^{KEKSS}\left(T_M, \Delta S_M\right)$ is a parameter which depends on the aromaticity of the solute, T_M is the melting temperature of the solute, ΔS_M is the entropy of fusion, and T is the absolute system temperature. The term containing b^{KEKSS} in Eq. F.5.14 is applicable only if $T_M > T$, and taken to be zero otherwise. Since the oils analogous to the tails of the surfactants studied in this Thesis are liquids at $25°C$ (oils of chain length C_{16} and below are liquids at this temperature[265]), a simplified version of Eq. F.5.14 holds:

$$\log S^{KEKSS} = C_0^{KEKSS} + \sum_a logS_a^{KEKSS}. \quad \text{(F.5.15)}$$

Similar to the case with the KWB model, once can relate g_{tr}^{KEKSS} to $\log S^{KEKSS}$. Since the units of S_a^{KEKSS} are mol/L, the following expression for g_{tr}^{KEKSS} is applicable (derived from Eq. F.5.2):

$$\frac{g_{tr}^{KEKSS}}{k_B T} = -\ln \mathcal{M}_{H_2O} + (\ln 10) \log S^{KEKSS}. \quad \text{(F.5.16)}$$

As with the KWB model, one can cast this expression in terms of constant and

group-additive contributions as

$$\frac{g_{tr}^{KEKSS}}{k_B T} = \frac{g_{tr,0}^{KEKSS}}{k_B T} + \sum_a \frac{g_{tr,a}^{KEKSS}}{k_B T}. \tag{F.5.17}$$

Inserting Eq. F.5.15 into Eq. F.5.16, and grouping terms, one finds, in comparing the grouped terms to the terms on the right-hand side of Eq. F.5.17, that

$$\frac{g_{tr,0}^{KEKSS}}{k_B T} = -\ln \mathcal{M}_{H_2O} + (\ln 10)\, C_0, \tag{F.5.18}$$

and

$$\frac{g_{tr,a}^{KEKSS}}{k_B T} = (\ln 10)\, log S_a^{KEKSS} \tag{F.5.19}$$

In Table F.8, I provide a summary of values for $\log S_a^{KEKSS}$ and $g_{tr,a}^{KEKSS}/k_B T$ (obtained via Eq. F.5.19) for atomic groups of particular interest in the study of surfactant systems. Note that, in the KEKSS model, group contributions are defined for various fragments, such as an "H attached to any C" and a "C aliphatic single bonded".[6] A given atomic group may contain multiple such fragments. Additionally, there are various correction terms to be applied based on either the type of fragment present in a group or the nature of the group as a whole. In preparing Table F.8, I have added these various terms together in advance, to enable presentation of a single value for each atomic group. The formulas for these additions are also provided in Table F.8, and the specific arguments refer to the numbered entries in Table 4 of the original paper[6] (e.g., #4 refers to Entry 4, "C aliphatic single bonded"). Note that, in contrast to the KWB model presented in Section F.5.3, alkyl groups attached to an aromatic ring are considered to have different properties than alkyl groups in a non-aromatic hydrocarbon (at least for CH_3 and CH_2 groups). This will have ramifications in the study of alkyl benzenes, which I discuss in Section F.5.5 below.

Finally, Eq. F.5.18 is applicable to all molecules; there is no adjustment term as with the KWB model, since adjustments are made within the group contributions themselves. At $25°C$, $\mathcal{M}_{H_2O} = 55.345\ mol/L$ (from Table F.2), and $(\ln 10)\, C_0 = 0.984$ (from the last entry in Table F.8), such that $g_{tr,0}^{KEKSS}/k_B T = -3.030$.

Table F.8: Parameters Derived from the Solubility Model of Kühne, Ebert, Kleint, Schmidt, and Schüürmann (KEKSS)[6] for $T = 25°C$

Formula*	Description	Group	$logS_a^{KEKSS}$†	$g_{tr,a}^{KEKSS}/k_BT$‡
#4+3(#1)+#53	alkane	CH_3	-0.823	-1.895
#4+2(#1)+#52	alkane	CH_2	-0.652	-1.501
#4+#1+#50	alkane	CH	-0.310	-0.713
#4+2(#50)	alkane	C	-0.154	-0.354
#4+3(#1)	alkane (arom.)§	CH_3	-0.393	-0.905
#4+2(#1)	alkane (arom.)§	CH_2	-0.466	-1.073
#4+#1+#50	alkane (arom.)§	CH	-0.310	-0.713
#4+2(#50)	alkane (arom.)§	C	-0.154	-0.354
#3+2(#1)	alkene	$=CH_2$	-0.416	-0.957
#3+#1	alkene	=CH-	-0.488	-1.124
#3+#50	alkene	=C-(-)	-0.332	-0.765
#5+#1	aromatic	CH	-0.353	-0.813
#5	aromatic	C	-0.426	-0.980
#15	alcohol (primary)	OH	1.092	2.514
#16	alcohol (secondary)	OH	1.212	2.791
#17	alcohol (tertiary)	OH	1.074	2.472
#20	ether	O	0.821	1.891
#58	C_0 constant¶		0.427	0.984

* Formula for $logS_a^{KEKSS}$ based on entry numbers in Table 4 of Ref. [6].
† Group contribution to the base-10 logarithm of the solubility (units of mol/L).
‡ Group contribution to g_{tr}, calculated from $logS_a^{KEKSS}$ according to Eq. F.5.19.
§ Values for alkane groups in alkyl aromatic compounds.
¶ Constant term is applied once per molecule.

F.5.5 Comparison of the Solubility Models

The individual $g_{tr,a}$ values presented in Tables F.6, F.7, and F.8 cannot be directly compared between models, since the MHA model lacks a constant term, and the KWB and KEKSS models have differing units for S (i.e., mol/m^3 and mol/L, respectively) and differing approaches to adjusting for aromaticity. As a result, only a comparison of whole-molecule g_{tr} values is informative. In Table F.9, I provide a collection of experimental values for the aqueous solubility of various organic compounds at $25°C$ (in mass-based units of mg/L), together with the corresponding g_{tr}/k_BT values obtained from

$$\frac{g_{tr}}{k_BT} = \ln\left(\frac{S\left[\frac{mg}{L}\right]}{1000\left[\frac{mg}{g}\right] MW\left[\frac{g}{mol}\right] \mathcal{M}_{H_2O}\left[\frac{mol}{L}\right]}\right), \tag{F.5.20}$$

where $\mathcal{M}_{H_2O}$ is the molarity of water (obtained from Table F.2 for $T = 298.15K$), and MW is the molecular weight of the solute, obtained from the atomic weights in Table F.1. The remaining columns of Table F.9 provide the model values for the MHA, MHA2, KWB, and KEKSS models, along with the errors $\Delta^{\text{model}} = g_{tr}^{\text{model}} - g_{tr}^{\text{expt}}$. Note that, as discussed in Section F.5.2, the MHA and MHA2 models contain parameters obtained by fitting to this dataset.

The collection of results in Table F.9 indicate that the MHA, KWB, and KEKSS models fit the n-alkanes reasonably well (with less than 1 k_BT unit of error) up to and including n-dodecane; in fact, it is worthwhile to note that the variation in the experimental data itself is at times as large as 0.5 k_BT. The MHA2 model presents more error in the linear alkanes in exchange for reducing the overall sum of squares error by about a third relative to the MHA model. This represents a limitation in a model lacking a constant term, since the KWB and KEKSS models have better fits to the whole set while maintaining a good agreement with the n-alkanes up to and including n-dodecane. It is also interesting to note that the MHA model fits the experimental data well for these chain lengths, despite the fact that the Abraham correlation from which the CH_2 and CH_3 contributions were derived was based on chain lengths only up to n-octane.[239] This indicates that each additional CH_2 group

added to a chain does, in fact, contribute approximately additively to g_{tr}.

The breakdown in all models for C_{14} and C_{16} chains is perhaps not surprising, given the break in the experimental solubility trend as a function of chain length observed between C_{12} and C_{14}, and the fact that all the solubility models discussed here are strictly linear. Experimentally, it is quite difficult to measure the solubilities of long chain alkanes with accuracy, since small hydrocarbon colloids can form in these systems which can skew the results towards an increased observed solubility.[349] If the experimental values are indeed accurate, then the MHA2 model demonstrates the best agreement. The reason for this is that the MHA2 model has CH_3 and CH_2 contributions fit with a dataset containing these long chains (i.e., the dataset of Table F.9).

It is clear that caution should be exercised in modeling surfactants with very long chain lengths. The transfer contribution could introduce a substantial error into micellization predictions if the selected model is poor. Fortunately, most of the linear surfactants studied in Chapter 5 have chain lengths between 8 and 12. As already mentioned, it appears the models perform well in this regime.

Examining the branched alkane entries in Table F.9, one observes that all four models in general fit quite well. The one exception is the fit of 4-methyloctane, which is consistently underpredicted across all four models, although the MHA and KEKSS models are in better agreement than the MHA2 and KWB models. In general, the deviation observed for 4-methyloctane may simply be a result of each of the four models being developed from datasets containing fairly low-molecular-weight branched alkanes. Unfortunately, there is a dearth of experimental data for higher-molecular-weight alkane compounds in the literature, which could have otherwise been used to improve the fitting or further validate the models.

The MHA, MHA2, and KWB models all demonstrate good agreement for all the alkyl benzenes considered. The KEKSS model appears to deviate considerably for the butyl benzenes, which is concerning, since these are still relatively low-molecular-weight compounds. It could be that the separate handling of alkyl groups in alkanes and in alkyl benzenes within the KEKSS model is problematic (in contrast to the

KWB model, where a constant adjustment term is used instead of modifications at the group-level).

Since many of the compounds in Table F.9 were used in the fitting of the four models (I fit the MHA and MHA2 models myself to this Table, while many of the compounds were also used by the authors of the KWB[5] and KEKSS[6] models), I selected a test set of compounds particularly relevant to the modeling of alkyl benzene sulfonate surfactant tails. These linear and branched alkyl benzenes are presented in Table F.10 with experimental solubilities measured by Sherblom and coworkers.[350] The observation that the KEKSS model does not handle alkyl benzenes well is reinforced with this dataset, where g_{tr} values are greatly underpredicted for decane, undecane, and dodecane-based alkyl benzenes (errors around 3 k_BT). The MHA model is acceptable for decane-based alkyl benzenes, but increases in error as the chain length increases, while the MHA2 model is better for dodecane, but worse for decane. In fact, the only model which appears to consistently provide reliable estimates for g_{tr} (with errors less than 1 k_BT up to and including dodecane-based alkyl benzenes) is the KWB model.

Similar to the case with long alkanes, these models tend to break down for tridecane and tetradecane-based alkyl benzenes: there appears to be a non-linear break in the curve representing solubility as a function of chain length, which cannot be well-represented by the four linear models considered here.

In conclusion, the KWB model appears to provide the most accurate predictions of whole-molecule solubility. However, it is important to remember that surfactant tails are not whole molecules, due to their connection with the surfactant head. As a result, I have chosen to implement all four solubility models within the micellization framework implemented in Program CS-PREDICT (see Section 2.4.1 of Chapter 2 for a description of this Program). Further comparison of these models in terms of their effect on micellization predictions is provided in Chapter 5 for linear surfactants and Chapter 6 for branched surfactants.

Table F.9: Comparison of Solubility Models I

Ref.*	Molecule	S^{expt}†	g_{tr}^{expt}‡	g_{tr}^{MHA}§	Δ^{MHA}¶	g_{tr}^{MHA2}§	Δ^{MHA2}¶	g_{tr}^{KWB}§	Δ^{KWB}¶	g_{tr}^{KEKSS}§	Δ^{KEKSS}¶
Linear Alkanes (Compiled in Table 2 of Ref. [351])											
352	*n*-pentane	3.84E+01	-11.552	-11.535	0.017	-12.084	-0.532	-11.761	-0.208	-11.322	0.230
353		3.94E+01	-11.527		-0.008		-0.558		-0.234		0.205
278		4.75E+01	-11.340		-0.195		-0.744		-0.421		0.018
352	*n*-hexane	9.47E+00	-13.129	-13.032	0.097	-13.279	-0.150	-13.080	0.049	-12.823	0.307
353		9.44E+00	-13.133		0.101		-0.147		0.053		0.310
278		1.24E+01	-12.863		-0.169		-0.416		-0.217		0.040
352	*n*-heptane	2.92E+00	-14.457	-14.529	-0.072	-14.474	-0.017	-14.399	0.057	-14.323	0.133
353		2.23E+00	-14.725		0.196		0.251		0.326		0.402
278		3.36E+00	-14.317		-0.212		-0.157		-0.083		-0.007
352	*n*-octane	6.58E-01	-16.078	-16.026	0.052	-15.669	0.409	-15.718	0.360	-15.824	0.254
353		4.30E-01	-16.504		0.478		0.836		0.786		0.680
278		8.47E-01	-15.825		-0.201		0.156		0.107		0.001
353	*n*-nonane	1.22E-01	-17.882	-17.523	0.359	-16.863	1.019	-17.037	0.845	-17.324	0.558
354		2.19E-01	-17.292		-0.231		0.429		0.255		-0.032
354	*n*-decane	5.18E-02	-18.839	-19.020	-0.181	-18.058	0.781	-18.357	0.482	-18.825	0.014
349	*n*-dodecane	3.69E-03	-21.662	-22.014	-0.352	-20.447	1.214	-20.995	0.667	-21.826	-0.165
349	*n*-tetradecane	2.19E-03	-22.334	-25.008	-2.674	-22.837	-0.503	-23.633	-1.299	-24.827	-2.493
349	*n*-hexadecane	8.97E-04	-23.360	-28.002	-4.642	-25.226	-1.866	-26.272	-2.912	-27.829	-4.469
Branched Alkanes (Compiled in Table 2 of Ref. [351])											
352	2-methylbutane (isopentane)	4.77E+01	-11.336	-11.247	0.089	-11.867	-0.531	-11.291	0.045	-10.929	0.407
353		4.79E+01	-11.332		0.085		-0.535		0.041		0.402
278		4.95E+01	-11.299		0.052		-0.568		0.008		0.370
352	2-methylpentane	1.38E+01	-12.756	-12.744	0.012	-13.062	-0.306	-12.610	0.146	-12.430	0.326
353		1.30E+01	-12.816		0.072		-0.246		0.205		0.386
278		1.57E+01	-12.627		-0.117		-0.435		0.017		0.197

Continued on Next Page

Table F.9 – Continued from Previous Page

Ref.*	Molecule	S^{expt}†	g_{tr}^{expt}‡	g_{tr}^{MHA}§	Δ^{MHA}¶	g_{tr}^{MHA2}§	Δ^{MHA2}¶	g_{tr}^{KWB}§	Δ^{KWB}¶	g_{tr}^{KEKSS}§	Δ^{KEKSS}¶
353	3-methylpentane	1.28E+01	-12.831	-12.744	0.087	-13.062	-0.230	-12.610	0.221	-12.430	0.401
278		1.78E+01	-12.496		-0.248		-0.566		-0.114		0.066
353	3-methylhexane	2.63E+00	-14.561	-14.241	0.320	-14.256	0.304	-13.929	0.631	-13.931	0.630
278		4.94E+00	-13.932		-0.309		-0.324		0.003		0.002
353	2-methylhexane	2.53E+00	-14.599	-14.241	0.358	-14.256	0.343	-13.929	0.670	-13.931	0.669
352	2,2-dimethylbutane	1.83E+01	-12.468	-12.190	0.278	-12.651	-0.183	-12.479	-0.010	-12.465	0.003
353		2.11E+01	-12.327		0.137		-0.324		-0.152		-0.139
278		2.37E+01	-12.211		0.021		-0.440		-0.268		-0.254
353	2,3-dimethylbutane	1.90E+01	-12.431	-12.456	-0.025	-12.844	-0.413	-12.141	0.290	-12.037	0.394
278		2.24E+01	-12.267		-0.189		-0.577		0.127		0.230
353	2,2-dimethylpentane	4.39E+00	-14.050	-13.687	0.363	-13.846	0.204	-13.798	0.252	-13.966	0.084
353	2,3-dimethylpentane	5.23E+00	-13.873	-13.953	-0.080	-14.039	-0.166	-13.460	0.414	-13.538	0.335
352	2,4-dimethylpentane	4.05E+00	-14.130	-13.953	0.177	-14.039	0.091	-13.460	0.671	-13.538	0.592
353		4.40E+00	-14.048	-13.953	0.095		0.009		0.588		0.510
278		5.48E+00	-13.827	-13.953	-0.126		-0.212		0.367		0.289
353	3,3-dimethylpentane	5.92E+00	-13.750	-13.687	0.063	-13.846	-0.096	-13.798	-0.048	-13.966	-0.216
353	3-methylheptane	7.90E-01	-15.896	-15.738	0.158	-15.451	0.445	-15.249	0.647	-15.431	0.465
352	2,2,4-trimethylpentane	2.43E+00	-14.771	-14.896	-0.126	-14.823	-0.052	-14.647	0.123	-15.074	-0.303
278		2.04E+00	-14.945	-14.896	0.049		0.122		0.298		-0.129
353	2,3,4-trimethylpentane	1.36E+00	-15.355	-15.162	0.193	-15.016	0.339	-14.309	1.046	-14.646	0.709
278		2.29E+00	-14.830	-15.162	-0.332		-0.186		0.520		0.184
352	2,2,5-trimethylhexane	1.15E+00	-15.639	-16.393	-0.754	-16.018	-0.379	-15.966	-0.328	-16.574	-0.936
278		5.38E-01	-16.395	-16.393	0.001		0.377		0.428		-0.180
353	4-methyloctane	1.15E-01	-17.941	-17.235	0.706	-16.646	1.295	-16.568	1.373	-16.932	1.009
Aromatics											
278	benzene	1.75E+03	-7.812	-7.606	0.207	-8.151	-0.339	-8.409	-0.597	-7.907	-0.094
278	toluene	5.71E+02	-9.097	-8.946	0.150	-9.310	-0.213	-9.471	-0.375	-8.979	0.117

Continued on Next Page

Table F.9 – Continued from Previous Page

Ref.*	Molecule	$S^{expt\dagger}$	$g_{tr}^{expt\ddagger}$	$g_{tr}^{MHA\S}$	$\Delta^{MHA\P}$	$g_{tr}^{MHA2\S}$	$\Delta^{MHA2\P}$	$g_{tr}^{KWB\S}$	$\Delta^{KWB\P}$	$g_{tr}^{KEKSS\S}$	$\Delta^{KEKSS\P}$
278	ethylbenzene	1.76E+02	-10.413	-10.443	-0.030	-10.505	-0.092	-10.791	-0.377	-10.052	0.361
278	*o*-xylene	2.12E+02	-10.228	-10.287	-0.059	-10.469	-0.241	-10.534	-0.306	-10.052	0.176
278	*m*-xylene	1.62E+02	-10.502	-10.287	0.215	-10.469	0.033	-10.534	-0.032	-10.052	0.450
278	*p*-xylene	1.84E+02	-10.369	-10.287	0.082	-10.469	-0.100	-10.534	-0.165	-10.052	0.317
279	isopropylbenzene	6.51E+01	-11.534	-11.652	-0.118	-11.482	0.053	-11.640	-0.106	-10.598	0.936
279	1,2,4-trimethylbenzene	5.88E+01	-11.636	-11.628	0.008	-11.628	0.008	-11.597	0.039	-11.125	0.511
279	1,2,3-trimethylbenzene	7.50E+01	-11.393	-11.628	-0.234	-11.628	-0.234	-11.597	-0.203	-11.125	0.268
279	1,3,5-trimethylbenzene	4.81E+01	-11.838	-11.628	0.210	-11.628	0.210	-11.597	0.241	-11.125	0.713
279	*n*-butylbenzene	1.18E+01	-13.356	-13.437	-0.082	-12.894	0.462	-13.429	-0.073	-12.198	1.158
279	*sec*-butylbenzene	1.75E+01	-12.956	-13.149	-0.193	-12.677	0.279	-12.959	-0.003	-11.671	1.285
279	*tert*-butylbenzene	2.94E+01	-12.439	-12.595	-0.156	-12.266	0.174	-12.827	-0.388	-11.144	1.295

*Reference number for the literature source of the experimental data.

†Experimental value for the aqueous solubility. Units of mg/L.

‡Derived value for the experimental transfer free energy. Units of k_BT.

§Model value for the transfer free energy. Units of k_BT. See text for a description of the models.

¶Difference between model and experimental transfer free energy values. Units of k_BT.

Table F.10: Comparison of Solubility Models II

Molecule*	$S^{expt\dagger}$	$g_{tr}^{expt\ddagger}$	$g_{tr}^{MHA\S}$	$\Delta^{MHA\P}$	$g_{tr}^{MHA2\S}$	$\Delta^{MHA2\P}$	$g_{tr}^{KWB\S}$	$\Delta^{KWB\P}$	$g_{tr}^{KEKSS\S}$	$\Delta^{KEKSS\P}$
n-decylbenzene	11.4	-22.303	-22.419	-0.116	-20.062	2.241	-21.344	0.959	-18.634	3.669
2-phenyldecane	26	-21.479	-22.131	-0.653	-19.845	1.634	-20.874	0.605	-18.107	3.371
3-phenyldecane	38	-21.099	-22.131	-1.032	-19.845	1.254	-20.874	0.225	-18.107	2.992
4-phenyldecane	36	-21.153	-22.131	-0.978	-19.845	1.309	-20.874	0.279	-18.107	3.046
5-phenyldecane	35	-21.182	-22.131	-0.950	-19.845	1.337	-20.874	0.307	-18.107	3.074
2-phenylundecane	8	-22.657	-23.628	-0.971	-21.039	1.618	-22.193	0.464	-19.180	3.477
3-phenylundecane	12	-22.252	-23.628	-1.376	-21.039	1.212	-22.193	0.059	-19.180	3.072
4-phenylundecane	9	-22.540	-23.628	-1.089	-21.039	1.500	-22.193	0.346	-19.180	3.359
5-phenylundecane	10	-22.434	-23.628	-1.194	-21.039	1.395	-22.193	0.241	-19.180	3.254
6-phenylundecane	11	-22.339	-23.628	-1.289	-21.039	1.299	-22.193	0.146	-19.180	3.159
2-phenyldodecane	4	-23.351	-25.125	-1.775	-22.234	1.116	-23.512	-0.162	-20.253	3.098
3-phenyldodecane	7	-22.791	-25.125	-2.334	-22.234	0.557	-23.512	-0.721	-20.253	2.538
4-phenyldodecane	5	-23.127	-25.125	-1.998	-22.234	0.893	-23.512	-0.385	-20.253	2.874
5-phenyldodecane	5	-23.127	-25.125	-1.998	-22.234	0.893	-23.512	-0.385	-20.253	2.874
6-phenyldodecane	4	-23.351	-25.125	-1.775	-22.234	1.116	-23.512	-0.162	-20.253	3.098
2-phenyltridecane	4	-23.351	-26.622	-3.272	-23.429	-0.078	-24.831	-1.481	-21.326	2.025
3-phenyltridecane	4	-23.351	-26.622	-3.272	-23.429	-0.078	-24.831	-1.481	-21.326	2.025
4-phenyltridecane	4	-23.351	-26.622	-3.272	-23.429	-0.078	-24.831	-1.481	-21.326	2.025
5-phenyltridecane	4	-23.351	-26.622	-3.272	-23.429	-0.078	-24.831	-1.481	-21.326	2.025
6-phenyltridecane	4	-23.351	-26.622	-3.272	-23.429	-0.078	-24.831	-1.481	-21.326	2.025

Continued on Next Page

Table F.10 – Continued from Previous Page

Molecule*	S^{expt}†	g_{tr}^{expt}‡	g_{tr}^{MHA}§	Δ^{MHA}¶	g_{tr}^{MHA2}§	Δ^{MHA2}¶	g_{tr}^{KWB}§	Δ^{KWB}¶	g_{tr}^{KEKSS}§	Δ^{KEKSS}¶
2-phenyltetradecane	4	-23.351	-28.119	-4.769	-24.624	-1.273	-26.151	-2.800	-22.399	0.952
3-phenyltetradecane	5	-23.127	-28.119	-4.992	-24.624	-1.496	-26.151	-3.023	-22.399	0.729
4-phenyltetradecane	4	-23.351	-28.119	-4.769	-24.624	-1.273	-26.151	-2.800	-22.399	0.952
5-phenyltetradecane	5	-23.127	-28.119	-4.992	-24.624	-1.496	-26.151	-3.023	-22.399	0.729
6-phenyltetradecane	4	-23.351	-28.119	-4.769	-24.624	-1.273	-26.151	-2.800	-22.399	0.952

*Systems from Tables IV and V of Ref. [350].

†Experimental value for the aqueous solubility. Units of $nmol/L$.

‡Derived value for the experimental transfer free energy. Units of k_BT.

§Model value for the transfer free energy. Units of k_BT. See text for a description of the models.

¶Difference between model and experimental transfer free energy values. Units of k_BT.

F.6 Surfactant Tail-Water Interfacial Tension

The interfacial tension of a surfactant tail against water is needed in order to evaluate the interfacial free-energy contribution to the free energy of micellization, a penalty arising from the area of the hydrophobic core remaining exposed to water after transfer of the surfactant tails into the core droplet. From Section 2.3.6 of Chapter 2, this term has the following form (shown for a single-component system for simplicity):

$$f_{int} = \sigma\left(a - a_0\right), \tag{F.6.1}$$

where σ is the curvature-corrected interfacial tension, a is the total interfacial area per surfactant, and a_0 is the interfacial area per surfactant shielded from contact with water by the cross-sectional area of the bond connecting surfactant head and tail. The curvature-corrected interfacial tension is related to the planar interfacial tension by[81,144–146]

$$\sigma = \frac{\sigma_0}{1 + (S-1)\,\delta/l_c}, \tag{F.6.2}$$

where σ_0 is the planar interfacial tension, S is a shape factor, δ is the Tolman length, and l_c is the core-minor radius of the micelle (see Section 2.3.6 of Chapter 2 for more information). Of these four terms, σ_0 and δ are dependent on the specific surfactant tail under consideration, with δ postulated to be related to molecular size (e.g., Puvvada and Blankschtein relate δ to the fully-extended chain length[81] – for which one could use the value of r_{rel} reported in the packing polynomials in Table F.17 of Section F.7). Assuming that δ can be calculated from geometric arguments, it remains to calculate σ_0.

In previous modeling efforts for linear surfactants, a correlation has been developed for the temperature and chain-length dependence of σ_0 for the n-alkanes from experimental data.[10] It is then assumed that the value of σ_0 for an n-alkane of equivalent number of carbons to the linear alkyl tail holds for the tail in question.[81] The

correlation is

$$\sigma_0 = \frac{1.381\,(117.99 + 57.868N - T\,(0.059N + 0.1768))}{N + 2.4}, \tag{F.6.3}$$

where N is the total number of carbons in the tail/oil and T is the absolute system temperature. Application of σ_0 to dodecane at $T = 293.15K$ yields $\sigma_0 = 53.04$ dyn/cm and at $T = 298.15K$ yields $\sigma_0 = 52.61$ dyn/cm, as compared to the experimental values of 52.78 dyn/cm and 52.46 dyn/cm, respectively.[10]

While Eq. F.6.3 is especially useful for linear alkyl tails, it is not clear what sort of relationship holds for branched alkanes, especially given the general lack of interfacial tension data in the literature.

When interfacial tensions are unknown, one common equation for their estimation is the Girifalco-Good-Fowkes equation,[355,356] which is most generally written as

$$\sigma_{0,AB} = \gamma_{0,A} + \gamma_{0,B} - 2\sqrt{\gamma^d_{0,A}\gamma^d_{0,B}}, \tag{F.6.4}$$

where $\sigma_{0,AB}$ is the planar interfacial tension between species A and B, $\gamma_{0,A}$ and $\gamma_{0,B}$ are the planar surface tensions of pure species A and B in equilibrium with their vapors, respectively, and $\gamma^d_{0,A}$ and $\gamma^d_{0,B}$ are the London dispersion contributions to $\gamma_{0,A}$ and $\gamma_{0,B}$, respectively. Substituting $A = water$ and $B = oil$, and making the common assumption that $\gamma^d_{0,oil} \approx \gamma_{0,oil}$ (i.e., that the surface tensions of oils are purely dispersive in nature),[356] one obtains

$$\sigma_0 = \gamma_{0,water} + \gamma_{0,oil} - 2\sqrt{\gamma^d_{0,water}} \cdot \sqrt{\gamma_{0,oil}}. \tag{F.6.5}$$

For water, Fowkes determined an average value of $\gamma^d_{0,water} \approx 21.8$ dyn/cm at $20°C$ based on fitting to experimental data for a small set of known interfacial tensions.[356]

Equation F.6.5 is presumed to be an improvement on the original Girifalco-Good equation, which involves only the various γ but also a fitting parameter, Φ:[357]

$$\sigma_{0,AB} = \gamma_{0,A} + \gamma_{0,B} - 2\Phi\sqrt{\gamma_{0,A}\gamma_{0,B}}, \tag{F.6.6}$$

Table F.11: Surface Tension of Water at Various Temperatures[7]

T (°C)	$\gamma_{0,water}$ (dyn/cm)	T (°C)	$\gamma_{0,water}$ (dyn/cm)
20	72.75	40	69.60
25	71.99	50	67.94
30	71.20	60	66.24

where Φ is dependent on the functionality of the molecule of interest, with recommended values provided by Demond and Lindner of $\Phi = 0.5595$ for aliphatic hydrocarbons and $\Phi = 0.7006$ for aromatic hydrocarbons.[358] For the oil/water system, Eq. F.6.6 is simply

$$\sigma_0 = \gamma_{0,water} + \gamma_{0,oil} - 2\left(\Phi\sqrt{\gamma_{0,water}}\right)\sqrt{\gamma_{0,oil}}, \tag{F.6.7}$$

and it is clear through comparison of Eqs. F.6.5 and F.6.7 that

$$\gamma^d_{0,water} = \Phi^2\gamma_{0,water}. \tag{F.6.8}$$

Using the Demond and Lindner value of $\Phi = 0.5595$[358] and $\gamma_{0,water} = 72.75$ dyn/cm (see Table F.11), Eq. F.6.8 yields $\gamma^d_{0,water} = 22.8$ dyn/cm, a slightly larger number than that proposed by Fowkes.[356] This is likely a better value to use, if using an averaged value, since the Demond and Lindner fitting was based on a larger set of molecules (including alkenes and branched alkanes).

In order to check that $\gamma^d_{0,water}$ can be viewed as a constant, and determine appropriate values for this term at temperatures other than $20°C$, I prepared Table F.12 from surface tension data prepared by the American Petroleum Institute (Project 44)[8] and Table F.13 from interfacial tension data collected by Zeppieri and coworkers[9] and Aveyard and Haydon for C_{14} and C_{16} oils.[10] Values for $\gamma^d_{0,water}$ were then calculated using a rearrangement of Eq. F.6.5:

$$\gamma^d_{0,water} = \left(\frac{\gamma_{0,water} + \gamma_{0,oil} - \sigma_0}{2\sqrt{\gamma_{0,oil}}}\right)^2, \tag{F.6.9}$$

as summarized in Table F.14.

From Table F.14, it is clear that there is a monotonic decrease in $\gamma^d_{0,water}$ not only

Table F.12: Surface Tensions of Oils at Various Temperatures (dyn/cm)[8]

Molecule	$20°C$	$25°C$	$30°C$	$40°C$	$50°C$	$60°C$
n-hexane	18.42	17.90	17.38	16.36	15.36	14.33
n-heptane	20.30	19.80	19.30	18.32	17.37	16.43
n-octane	21.76	21.26	20.77	19.80	18.86	17.94
n-nonane	22.92	22.44	21.95	21.00	20.08	19.16
n-decane	23.92	23.44	22.96	22.02	21.09	20.18
n-undecane	24.74	24.27	23.80	22.88	21.96	21.07
n-dodecane	25.44	24.98	24.51	23.60	22.70	21.83
n-tridecane	26.10	25.60	25.20	24.30	23.40	22.50
n-tetradecane	26.60	26.20	25.70	24.80	24.00	23.10
n-pentadecane	27.10	26.70	26.20	25.30	24.50	23.70
n-hexadecane	27.60	27.10	26.70	25.80	25.00	24.20

Table F.13: Interfacial Tensions of Oils Against Water at Various Temperatures (dyn/cm)[9,10]

Molecule	$20°C$	$25°C$	$30°C$	$40°C$	$50°C$	$60°C$
n-hexane	50.80	50.38	49.96	48.92	48.13	
n-heptane	51.24	50.71	50.30	49.38	48.55	
n-octane	51.64	51.16	50.74	49.84	48.95	48.32
n-nonane	52.06	51.63	51.21	50.27	49.36	48.82
n-decane	52.33	51.98	51.51	50.53	49.78	49.21
n-undecane	52.56	52.25	51.82	50.95	50.11	
n-dodecane	52.87	52.55	52.14	51.24	50.43	50.00
n-tridecane						
n-tetradecane	53.32	52.92	52.46			
n-pentadecane						
n-hexadecane	53.77	53.30	52.90			

Table F.14: Calculated Values for $\gamma^d_{0,water}$ at Various Temperatures (dyn/cm)

Molecule	$20°C$	$25°C$	$30°C$	$40°C$	$50°C$	$60°C$
n-hexane	22.12	21.80	21.45	20.97	20.13	
n-heptane	21.53	21.31	20.93	20.27	19.45	
n-octane	21.11	20.83	20.46	19.76	18.99	17.92
n-nonane	20.74	20.41	20.03	19.36	18.61	17.46
n-decane	20.55	20.14	19.81	19.17	18.26	17.15
n-undecane	20.40	19.95	19.59	18.85	18.02	
n-dodecane	20.18	19.75	19.36	18.65	17.81	16.60
n-tridecane						
n-tetradecane	19.91	19.56	19.21			
n-pentadecane						
n-hexadecane	19.65	19.34	18.96			

with increasing temperature, but also with increasing carbon-number. The entire table is well-described by the linear regression model

$$\gamma^d_{0,water} = 51.18 - 2.75 \ln N - 0.082T, \tag{F.6.10}$$

where N is the number of carbons in the oil, and T is the absolute system temperature (the average absolute error is 0.16 dyn/cm, with a maximum absolute error of 0.43 dyn/cm).

The difference between $\gamma^d_{0,water}$ values between n-hexadecane and n-hexane is quite substantial, at about 2.4 dyn/cm up to $30°C$. The average value at $20°C$ is about 20.7 dyn/cm, lower than Fowkes' value by 1.1 dyn/cm. The value derived from Eq. F.6.8 and Demond and Lindner's value for Φ is quite close to that for n-hexane, but overestimates $\gamma^d_{0,water}$ as chain length increases. What this situation indicates is that it is unlikely that a single value for $\gamma^d_{0,water}$ should be applied to branched alkanes, since linear alkanes themselves demonstrate such markedly different values. (Note that this discussion has no bearing on the current method for determining the interfacial tension for linear alkyl tails, since the experimental data is already directly available.)

To proceed, I examined a large dataset for linear and branched alkane surface tensions prepared by Needham and coworkers,[11] which I have reproduced in full in Table F.16 together with a summary of: (i) the number of carbons in the main chain, (ii) the number of carbons in the total chain, and (iii) the number of branches present. In Table F.15, which precedes this full dataset, I provide a subset of molecules containing a maximum of one branch. I have sorted both Tables in order of surface tension.

I prepared Table F.15 in response to a lack of clear or useful trends in the full dataset of Table F.16. That is, in the full set, there is a general increase in the average surface tension for molecules with a given carbon-number, but the spread of values about this mean also increases. For example, the set of alkanes with 7 total carbons spans surface tension values from 18.02 to 20.44 dyn/cm (a 2.42 dyn/cm

span), alkanes with 8 total carbons spans surface tension values from 18.77 to 21.99 dyn/cm (a 3.22 dyn/cm span) and alkanes with 9 total carbons span surface tension values of 20.04 to 23.87 dyn/cm (a 3.83 dyn/cm span). It is clear that there is also significant overlap between these sets. Finally, there is no clear correlation with respect to the number of carbons in the main chain or the number of branches.

However, when one examines the values in Table F.15, where there is a maximum of one branch on the chain, the spread of values is closer to 1 dyn/cm and the surface tensions are clearly grouped by carbon number, with no overlap between groups. This is promising, since my systems are all singly-branched. Based on the lack of experimental data for oil-water interfacial tensions for branched alkanes, and an associated lack of understanding with regards to the sensitivity of $\gamma^d_{0,water}$ to branching, I have decided to proceed by utilizing Eq. F.6.3, with N set equal to the total number of carbons in the tail. Based on the similarity in surface tensions of compounds grouped by carbon-number in Table F.15, and the fact that application of an equation such as Eq. F.6.5 serves to diminish the impact of errors in surface tension (assuming $\gamma^d_{0,water}$ does not vary too much if carbon-number and temperature is fixed), this appears to be the best means for moving forward. However, it is important to keep in mind that the results of Table F.16 indicate that this issue must be revisited if highly-branched surfactant tails are under consideration.

Table F.15: Surface Tension of Oils I[11]

Molecule	γ^{expt*}	$N_{c,main}{}^{\dagger}$	$N_{c,total}{}^{\ddagger}$	$N_b{}^{\S}$
2-methylbutane	15	4	5	1
n-pentane	16	5	5	0
2-methylpentane	17.38	5	6	1
3-methylpentane	18.12	5	6	1
n-hexane	18.42	6	6	0
2-methylhexane	19.29	6	7	1
3-methylhexane	19.79	6	7	1
Continued on Next Page				

Table F.15 – Continued from Previous Page

Molecule	γ^{expt*}	$N_{c,main}{}^{\dagger}$	$N_{c,total}{}^{\ddagger}$	$N_b{}^{\S}$
n-heptane	20.26	7	7	0
3-ethylpentane	20.44	5	7	1
2-methylheptane	20.6	7	8	1
4-methylheptane	21	7	8	1
3-methylheptane	21.17	7	8	1
3-ethylhexane	21.51	6	8	1
n-octane	21.76	8	8	0
2-methyloctane	21.88	8	9	1
3-methyloctane	22.34	8	9	1
4-methyloctane	22.34	8	9	1
3-ethylheptane	22.81	7	9	1
4-ethylheptane	22.81	7	9	1
n-nonane	22.92	9	9	0

*Experimental surface tension in dyn/cm.
†Number of carbons in the main chain.
‡Number of carbons total.
§Number of branches off the main chain.

Table F.16: Surface Tension of Oils II[11]

Molecule	γ^{expt*}	$N_{c,main}{}^{\dagger}$	$N_{c,total}{}^{\ddagger}$	$N_b{}^{\S}$
2-methylbutane	15	4	5	1
n-pentane	16	5	5	0
2,2-dimethylbutane	16.3	4	6	2
2,3-dimethylbutane	17.37	4	6	2
2-methylpentane	17.38	5	6	1
2,2-dimethylpentane	18.02	5	7	2

Continued on Next Page

Table F.16 – Continued from Previous Page

Molecule	γ^{expt*}	$N_{c,main}$ †	$N_{c,total}$ ‡	N_b §
3-methylpentane	18.12	5	6	1
2,4-dimethylpentane	18.15	5	7	2
n-hexane	18.42	6	6	0
2,2,3-trimethylbutane	18.76	4	7	3
2,2,4-trimethylpentane	18.77	5	8	3
2-methylhexane	19.29	6	7	1
3,3-dimethylpentane	19.59	5	7	2
2,2-dimethylhexane	19.6	6	8	2
2,5-dimethylhexane	19.73	6	8	2
3-methylhexane	19.79	6	7	1
2,3-dimethylpentane	19.96	5	7	2
2,2,5-trimethylhexane	20.04	6	9	3
2,4-dimethylhexane	20.05	6	8	2
n-heptane	20.26	7	7	0
2,2,4,4-tetramethylpentane	20.37	5	9	4
3-ethylpentane	20.44	5	7	1
2,2,4-trimethylhexane	20.51	6	9	3
2-methylheptane	20.6	7	8	1
3,3-dimethylhexane	20.63	6	8	2
2,2,3-trimethylpentane	20.67	5	8	3
2,2-dimethylheptane	20.8	7	9	2
2,6-dimethylheptane	20.83	7	9	2
2,3-dimethylhexane	20.99	6	8	2
4-methylheptane	21	7	8	1
2,3,4-trimethylpentane	21.14	5	8	3
3-methylheptane	21.17	7	8	1
2,4,4-trimethylhexane	21.17	6	9	3

Continued on Next Page

Table F.16 – Continued from Previous Page

Molecule	γ^{expt*}	$N_{c,main}{}^{\dagger}$	$N_{c,total}{}^{\ddagger}$	$N_b{}^{\S}$
2,3,5-trimethylhexane	21.27	6	9	3
2,4-dimethylheptane	21.3	7	9	2
2,5-dimethylheptane	21.3	7	9	2
3-ethylhexane	21.51	6	8	1
3-ethyl-2-methylpentane	21.52	5	8	2
2,3,3-trimethylpentane	21.56	5	8	3
3,4-dimethylhexane	21.64	6	8	2
n-octane	21.76	8	8	0
3,5-dimethylheptane	21.77	7	9	2
4-ethyl-2-methylhexane	21.77	6	9	2
2,2,3-trimethylhexane	21.86	6	9	3
2-methyloctane	21.88	8	9	1
2,2,3,4-tetramethylpentane	21.98	5	9	4
3-ethyl-3-methylpentane	21.99	5	8	2
3,3-dimethylheptane	22.01	7	9	2
4,4-dimethylheptane	22.01	7	9	2
3-methyloctane	22.34	8	9	1
4-methyloctane	22.34	8	9	1
2,3-dimethylheptane	22.34	7	9	2
2,2-dimethyl-3-ethylpentane	22.38	5	9	3
2,3,3-trimethylhexane	22.41	6	9	3
3,4-dimethylheptane	22.8	7	9	2
3-ethyl-2-methylhexane	22.8	6	9	2
2,3,4-trimethylhexane	22.8	6	9	3
2,4-dimethyl-3-ethylpentane	22.8	5	9	3
3-ethylheptane	22.81	7	9	1
4-ethylheptane	22.81	7	9	1

Continued on Next Page

Table F.16 – Continued from Previous Page

Molecule	γ^{expt*}	$N_{c,main}$†	$N_{c,total}$‡	N_b§
n-nonane	22.92	9	9	0
3-ethyl-3-methylhexane	23.22	6	9	2
4-ethyl-3-methylhexane	23.27	6	9	2
3,3,4-trimethylhexane	23.27	6	9	3
2,3,3,4-tetramethylpentane	23.31	5	9	4
2,2,3,3-tetramethylpentane	23.38	5	9	4
3,3-diethylpentane	23.75	5	9	2
2,3-dimethyl-3-ethylpentane	23.87	5	9	3

*Experimental surface tension in dyn/cm.
†Number of carbons in the main chain.
‡Number of carbons total.
§Number of branches off the main chain.

F.7 Surfactant Tail Packing Polynomials

In the final section of this Appendix, I provide a listing of packing polynomials obtained using the methodology described in Chapter 4 and the group volumes obtained in Section F.4. For single species, these polynomials are of the form

$$\frac{f_{pack}}{k_B T} = \sum_{n=0}^{N-1} c_n r_{red}^n \tag{F.7.1}$$

where $\{c_n\}$ are the N polynomial coefficients used to fit the f_{pack} data (typically $N = 4$ provides an excellent fit), and r_{red} is the reduced micelle core radius, obtained by dividing the micelle core radius by the maximum chain length of the tail being packed (r_{rel}, obtained through geometrical arguments for the tail in an all-trans configuration). Note that the term "radius" here refers to the radius of a sphere, the cross-sectional radius of a cylinder, and the half-width of a bilayer. The polynomials are intended for use with micelles possessing radii between some $r_{\min}$ and $r_{\max}$. The value of $r_{\max}$ can exceed r_{rel} by a small amount, due the finite width of the concen-

tric layers used during packing. As this width approaches zero, $r_{\max}$ approaches r_{rel} in such cases. In other situations, such as with bilayers, $r_{\max}$ will typically be less than r_{rel} and represents the maximum radius at which the nonlinear solver was able to determine lateral pressures which satisfy the constant density packing constraints (please refer to Chapter 4 for explanations of these terms). The value of $r_{\min}$ is usually determined by the radii selected for packing. The choice of $r_{\min} = 0.4r_{rel}$ appears to be a suitable lower bound, since micelles have not been obtained at such a small radius during CS-PREDICT calculations.

In Table F.17, I provide the packing polynomial coefficients for n-alkanes, singly-branched alkanes, and linear and branched alkyl benzenes for use with Eq. F.7.1. Polynomials for each of the three principal geometries (bilayer, cylinder, and sphere) are provided. These data are for $T = 298.15K$, and were obtained by fitting to the packing results obtained for up to 15 different radii per geometry and molecule, with 500 external orientational samples taken per internal conformation of the chain, a gauche energy of $\sigma = 500\ cal/mol$ applied, and the head group positioned on the interface. All other packing parameters (e.g. RIS weightings and bond lengths, angles, and geometries) are according to the details in Chapter 4.

In order to describe the various molecules compactly in Table F.17, I developed the following systematic formula representations for these molecules.

For the n-alkanes, I use a formula representation of "CN_c", where N_c is the total number of carbons in the chain (e.g., $C10$ for a 10-carbon alkyl chain).

For singly-branched alkanes, I use a formula representation of "$CN_c\,(N_nSN_b)$", where N_c is the length of the main chain, N_n is the node number at which the branch emanates, S indicates the stereochemistry of the chain (where $S = -$ indicates the branch is 120° behind the main chain in terms of dihedral state, and $S = +$ indicates the branch is 120° ahead of the main chain), and N_b is the length of the branch at the specified node. As an example, one valid formula representation is $C10\,(1-5)$, which indicates a $C10$ main chain, with a $C5$ side chain at position 1, connected such that it is 120° behind the main chain in terms of dihedral angle.

This formula can be readily extended to accommodate additional branches on the

main chain, or, through recursion, branches off branches. Mixed stereochemistry can also be specified. In Figure F-3, the branched alkyl tail $C8(1-5(2-1),6+2))$ is presented, indicating a $C8$ main chain attached to the head, with two primary branches at the 1 and 6 positions: the primary branch at position 1 having a length of 5 carbons and possessing a secondary branch at position 2 (with length 1), and the primary branch at position 6 having a length of 2 carbons. The stereochemistry follows the sign convention already mentioned.

For alkyl benzenes, I use a formula representation of "BN_aAN_s" prepended to any valid linear or branched alkyl string, where N_a indicates the number of aromatic carbon groups included in the tail, and N_s indicates the starting position on the ring for the chain specification which follows. For example, $B3A4C10$ indicates a $C10$ linear alkyl tail is attached to a benzene ring at position 4 (the head, say a sulfonate, is at position 1 on the ring), and three of the possible six carbons of the benzene ring are included in the tail. I assume symmetry of the ring when chains are attached to the 4-position, such that valid values for N_a are 0, 1, 3, 5, and 6. Such symmetry has always been observed in the head-tail identifications obtained in my work. Note that packing of a $B0A4C10$ tail is not equivalent to the packing of a $C10$ tail, since the benzene ring is explicitly included in the head region in the $B0A4C10$ case, limiting the available configurations of the tail (head groups are not permitted to enter the core region, and the presence of the benzene ring decreases the size of the set of configurations which satisfy these constraints relative to a small head group, thereby increasing the packing free-energy penalty). In Figure F-3, the branched alkyl benzene tail $B3A4C6(1-2)$ is presented, where a $C6(1-2)$ tail is attached to position 4 of the benzene ring, and three of the benzene ring carbon groups are in the tail.

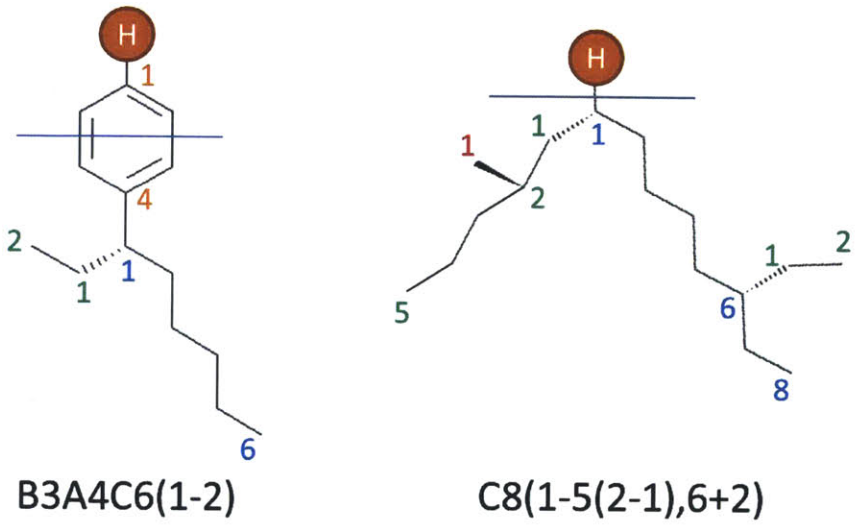


Figure F-3: Examples of formula representations for alkyl benzenes and branched alkanes. The formulas are interpreted as indicated in the Text. Numbering of groups in the benzene ring and along alkane branches is indicated (color code for numbers: brown: benzene ring positions, blue: main chain, green: primary branch, red: secondary branch). A dotted line indicates a bond into the page; a solid wedge indicates a bond out of the page. The circle labeled H indicates the position of the surfactant head, while the horizontal blue line indicates the separation between head and tail regions.

Table F.17: Packing Polynomials

Formula*	**Sh†**	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
Linear Alkanes								
C6	1	3.425	8.372	7.611	-5.6438	33.3479	-53.9343	28.7774
C6	2	3.425	8.372	7.611	4.5315	-7.1857	4.0357	0.1103
C6	3	3.425	8.372	7.611	16.2566	-46.1334	49.7231	-18.1129
C7	1	3.568	9.366	8.920	-1.8019	17.8139	-34.4288	21.5543
C7	2	3.568	9.812	8.920	5.1897	-9.2980	6.0546	-0.3404
C7	3	3.568	9.812	8.920	16.5908	-47.9013	52.4951	-19.4034
C8	1	4.059	10.655	10.147	-4.6043	32.0488	-57.6628	33.9524
C8	2	4.059	11.162	10.147	4.7625	-7.2636	2.9595	1.2274
C8	3	4.059	11.162	10.147	14.5907	-40.0597	42.6370	-15.3627
C9	1	4.579	11.448	11.448	-2.1724	20.5077	-40.4428	26.1404
C9	2	4.579	12.593	11.448	4.3281	-5.2323	-0.0501	2.7448
C9	3	4.579	12.593	11.448	12.8565	-33.0223	33.5017	-11.5058
C10	1	5.074	12.684	12.684	-3.0885	25.8147	-50.1555	32.0174
C10	2	5.074	13.318	12.684	4.9350	-7.4605	2.5421	1.8785
C10	3	5.074	13.953	12.684	13.0166	-33.7442	34.5284	-11.9199
C11	1	5.592	13.280	13.979	-2.7000	24.2921	-48.5820	32.0163
C11	2	5.592	13.979	13.979	5.0608	-8.1237	3.6405	1.3931
C11	3	5.592	15.377	13.979	11.9479	-29.3401	28.7209	-9.4211
C12	1	6.088	14.460	15.221	-3.1095	26.7680	-53.4401	35.3070
C12	2	6.088	15.221	15.221	4.6440	-6.1163	0.4663	3.1054
C12	3	6.088	15.982	15.221	11.9205	-29.8872	30.2667	-10.3728
C13	1	6.605	15.686	16.512	-3.5082	29.2497	-58.3282	38.6101
C13	2	6.605	16.512	16.512	4.9723	-7.3407	1.9130	2.6527
C13	3	6.605	16.512	16.512	12.5447	-33.0003	35.2481	-12.8796
C14	1	7.103	15.982	17.758	-1.9560	21.5505	-46.1830	32.7470
C14	2	7.103	16.870	17.758	5.3312	-9.0539	4.5899	1.3724
C14	3	7.103	17.758	17.758	11.9895	-30.4537	31.5428	-11.1144
C15	1	7.618	17.141	19.046	-2.3036	23.6623	-50.3618	35.6693
C15	2	7.618	18.093	19.046	4.8351	-6.6527	0.8257	3.3681
C15	3	7.618	19.046	19.046	10.9855	-26.1149	25.4973	-8.3530
C16	1	8.118	18.265	20.295	-2.9392	27.3310	-57.1919	39.9968
C16	2	8.118	19.280	20.295	5.1984	-8.0918	2.6145	2.7540
C16	3	8.118	20.295	20.295	11.5662	-28.6191	28.9381	-9.8321
Branched Alkanes								
C6(1-1)	1	3.425	8.372	7.611	-8.4115	47.6695	-75.9277	40.1802
C6(1-1)	2	3.425	8.372	7.611	5.9646	-10.4985	7.4642	-0.9922
C6(1-1)	3	3.425	8.372	7.611	19.4444	-54.4305	58.3283	-21.1225
C6(1-2)	1	3.425	7.991	7.611	-15.9618	84.8273	-134.5453	71.3344
C6(1-2)	2	3.425	8.372	7.611	7.2409	-13.6319	10.4615	-1.6195
C6(1-2)	3	3.425	8.372	7.611	26.8786	-78.9486	86.4938	-31.7270

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C6(1-3)	1	3.425	7.611	7.611	1.3801	6.6437	-22.0299	20.2070
C6(1-3)	2	3.425	8.372	7.611	8.3275	-15.8640	11.6541	-1.3681
C6(1-3)	3	3.425	8.372	7.611	34.1486	-101.0908	110.0608	-40.0815
C6(1-4)	1	3.425	7.611	7.611	2.2897	3.9352	-19.7719	20.0501
C6(1-4)	2	3.425	8.372	7.611	8.8699	-15.7660	9.7793	0.1442
C6(1-4)	3	3.425	8.372	7.611	42.0726	-125.8643	136.9206	-49.7258
C6(1-5)	1	3.425	7.991	7.611	-11.4814	67.9975	-114.3186	64.1666
C6(1-5)	2	3.425	8.372	7.611	10.0349	-18.4851	12.4286	-0.9354
C6(1-5)	3	3.425	8.372	7.611	49.9987	-149.9324	162.6697	-59.0885
C6(2-1)	1	3.425	8.372	7.611	-11.4036	61.6280	-97.6381	50.9603
C6(2-1)	2	3.425	8.372	7.611	4.8183	-6.3147	1.7985	1.4727
C6(2-1)	3	3.425	8.372	7.611	17.2951	-47.7094	50.7554	-18.3147
C6(2-2)	1	3.425	7.991	7.611	-17.3996	92.9031	-149.9115	79.6170
C6(2-2)	2	3.425	8.372	7.611	5.3144	-6.0190	-0.3461	3.0798
C6(2-2)	3	3.425	8.372	7.611	22.1310	-63.1005	68.0232	-24.7599
C6(2-3)	1	3.425	7.991	7.611	-22.7027	120.3645	-195.1467	103.7408
C6(2-3)	2	3.425	8.372	7.611	6.3748	-7.8223	0.2043	3.4833
C6(2-3)	3	3.425	8.372	7.611	34.5304	-105.1561	116.5426	-43.2773
C6(2-4)	1	3.425	7.991	7.611	-12.3664	71.8292	-120.3340	65.9174
C6(2-4)	2	3.425	8.372	7.611	8.4985	-14.5226	8.1295	0.2110
C6(2-4)	3	3.425	8.372	7.611	47.4994	-148.5856	166.0575	-62.0351
C6(3-1)	1	3.425	7.991	7.611	-11.3665	62.4995	-101.5609	53.9265
C6(3-1)	2	3.425	8.372	7.611	4.1565	-3.5966	-2.2249	3.3006
C6(3-1)	3	3.425	8.372	7.611	17.7084	-49.1102	52.1199	-18.7774
C6(3-2)	1	3.425	7.991	7.611	-14.6280	80.4093	-131.8066	70.2039
C6(3-2)	2	3.425	8.372	7.611	2.5142	5.3345	-15.5226	9.5012
C6(3-2)	3	3.425	8.372	7.611	27.8416	-83.2647	91.5784	-33.9086
C6(3-3)	1	3.425	7.991	7.611	-15.0780	81.3607	-130.2901	67.5867
C6(3-3)	2	3.425	8.372	7.611	4.6590	-1.4722	-7.2098	5.9104
C6(3-3)	3	3.425	8.372	7.611	36.9970	-111.6249	122.4967	-45.3218
C6(4-1)	1	3.425	7.991	7.611	-15.1405	79.7217	-125.8626	64.5122
C6(4-1)	2	3.425	8.372	7.611	4.2949	-4.4252	-0.6971	2.4625
C6(4-1)	3	3.425	8.372	7.611	17.0876	-47.4301	50.5982	-18.3163
C6(4-2)	1	3.425	8.372	7.611	-12.9632	68.9264	-107.4813	54.1487
C6(4-2)	2	3.425	8.372	7.611	3.7019	-0.2839	-6.7793	5.0839
C6(4-2)	3	3.425	8.372	7.611	23.5937	-69.0076	75.8016	-28.1744
C6(5-1)	1	3.425	8.372	7.611	-5.2995	31.3307	-49.2239	25.1402
C6(5-1)	2	3.425	8.372	7.611	5.6746	-10.3132	7.6139	-1.4591
C6(5-1)	3	3.425	8.372	7.611	20.8382	-60.4937	66.0005	-24.3984
C7(1-1)	1	3.568	8.920	8.920	-0.1403	11.0715	-24.1750	17.1706
C7(1-1)	2	3.568	9.812	8.920	6.4808	-12.4114	9.5943	-1.6076
C7(1-1)	3	3.568	9.812	8.920	21.2611	-62.8103	70.0689	-26.2451

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C7(1-2)	1	3.568	8.028	8.920	-2.0057	22.3766	-45.1116	31.0680
C7(1-2)	2	3.568	9.812	8.920	7.3559	-14.2372	11.2000	-1.7204
C7(1-2)	3	3.568	9.812	8.920	26.6582	-80.5563	90.6369	-34.0046
C7(1-3)	1	3.568	8.028	8.920	-3.5212	32.4384	-66.3503	46.0910
C7(1-3)	2	3.568	9.812	8.920	8.3891	-16.9800	13.8166	-2.3152
C7(1-3)	3	3.568	9.812	8.920	30.9130	-92.9475	103.4810	-38.4120
C7(1-4)	1	3.568	8.474	8.920	-26.2497	151.6750	-267.0645	155.3691
C7(1-4)	2	3.568	9.366	8.920	9.5655	-20.4906	17.8251	-3.6454
C7(1-4)	3	3.568	9.812	8.920	36.9907	-112.8477	126.1166	-46.8526
C7(1-5)	1	3.568	8.920	8.920	-11.1377	74.2644	-138.7893	86.4725
C7(1-5)	2	3.568	9.366	8.920	10.1273	-20.7106	16.4821	-2.4639
C7(1-5)	3	3.568	9.812	8.920	42.0110	-127.9297	142.1982	-52.6165
C7(1-6)	1	3.568	9.366	8.920	-6.6886	49.9696	-94.1614	59.1600
C7(1-6)	2	3.568	9.812	8.920	10.5599	-20.1832	14.2545	-1.2126
C7(1-6)	3	3.568	9.812	8.920	50.7560	-158.3650	178.3622	-66.8738
C7(2-1)	1	3.568	8.920	8.920	-0.2679	12.0554	-27.2287	19.4066
C7(2-1)	2	3.568	9.812	8.920	5.7890	-9.8301	5.7147	0.2208
C7(2-1)	3	3.568	9.812	8.920	18.2708	-52.1227	56.9554	-21.0294
C7(2-2)	1	3.568	8.028	8.920	-3.0327	28.6976	-58.7703	39.1851
C7(2-2)	2	3.568	9.812	8.920	6.3149	-9.8807	4.1179	1.6549
C7(2-2)	3	3.568	9.812	8.920	24.3635	-73.1616	81.7730	-30.6171
C7(2-3)	1	3.568	8.474	8.920	-16.7906	102.7595	-186.5880	110.5959
C7(2-3)	2	3.568	9.812	8.920	7.3399	-12.0010	5.3590	1.7691
C7(2-3)	3	3.568	9.812	8.920	33.5465	-104.9973	119.3309	-45.2089
C7(2-4)	1	3.568	8.920	8.920	-15.2318	95.1999	-173.9601	103.5924
C7(2-4)	2	3.568	9.812	8.920	8.2330	-13.6944	6.2337	1.9036
C7(2-4)	3	3.568	9.812	8.920	41.6915	-132.3231	150.7993	-57.2150
C7(2-5)	1	3.568	8.920	8.920	-7.1800	53.5816	-103.1438	64.0237
C7(2-5)	2	3.568	9.812	8.920	9.5636	-16.9763	9.3354	0.7954
C7(2-5)	3	3.568	9.812	8.920	50.6534	-162.8966	186.6748	-71.2015
C7(3-1)	1	3.568	9.366	8.920	-3.1836	27.9272	-56.2581	36.2102
C7(3-1)	2	3.568	9.812	8.920	5.1579	-7.4498	2.2361	1.8446
C7(3-1)	3	3.568	9.812	8.920	18.3428	-52.3274	56.7119	-20.7603
C7(3-2)	1	3.568	8.920	8.920	-6.5922	47.5033	-91.3000	56.5478
C7(3-2)	2	3.568	9.812	8.920	4.6704	-3.0838	-5.4759	5.9515
C7(3-2)	3	3.568	9.812	8.920	25.9027	-78.3737	87.5922	-32.8610
C7(3-3)	1	3.568	8.920	8.920	-6.9344	50.2133	-96.6174	59.6919
C7(3-3)	2	3.568	9.812	8.920	5.6319	-4.6934	-4.9742	6.2605
C7(3-3)	3	3.568	9.812	8.920	37.7140	-119.0338	135.2295	-51.3427
C7(3-4)	1	3.568	9.366	8.920	-7.2866	50.7897	-93.7682	55.5167
C7(3-4)	2	3.568	9.812	8.920	7.9302	-12.3344	4.3519	2.3038
C7(3-4)	3	3.568	9.812	8.920	51.6605	-168.5997	194.3117	-74.5278

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C7(4-1)	1	3.568	9.366	8.920	-7.8130	50.1081	-88.8233	50.8755
C7(4-1)	2	3.568	9.812	8.920	5.2418	-7.6960	2.7192	1.5415
C7(4-1)	3	3.568	9.812	8.920	16.8063	-46.5285	49.8305	-18.1084
C7(4-2)	1	3.568	8.920	8.920	-7.5848	51.4600	-94.7057	55.9326
C7(4-2)	2	3.568	9.812	8.920	4.1308	-0.6728	-8.4462	7.0017
C7(4-2)	3	3.568	9.812	8.920	21.1680	-59.9477	64.6428	-23.6162
C7(4-3)	1	3.568	8.920	8.920	-9.1041	58.6593	-104.1046	59.1747
C7(4-3)	2	3.568	9.812	8.920	5.7575	-5.0850	-3.8332	5.2325
C7(4-3)	3	3.568	9.812	8.920	29.5534	-87.1554	95.5409	-35.4259
C7(5-1)	1	3.568	9.366	8.920	-10.6078	61.8792	-104.0286	56.3711
C7(5-1)	2	3.568	9.812	8.920	5.0728	-7.3860	2.6737	1.3151
C7(5-1)	3	3.568	9.812	8.920	17.4399	-49.0129	52.9811	-19.4722
C7(5-2)	1	3.568	9.366	8.920	-7.1221	44.8261	-76.2347	41.4794
C7(5-2)	2	3.568	9.812	8.920	5.4363	-7.0997	1.3649	2.0398
C7(5-2)	3	3.568	9.812	8.920	23.1169	-68.5087	76.3688	-28.8080
C7(6-1)	1	3.568	9.366	8.920	-4.4667	30.8485	-53.3985	29.6562
C7(6-1)	2	3.568	9.812	8.920	6.5513	-13.5106	11.1767	-2.5956
C7(6-1)	3	3.568	9.812	8.920	22.6160	-68.9017	77.8550	-29.5662
C8(1-1)	1	4.059	9.640	10.147	-0.3773	12.2846	-26.2251	18.5353
C8(1-1)	2	4.059	11.162	10.147	5.6828	-8.9884	4.7463	0.6849
C8(1-1)	3	4.059	11.162	10.147	17.9946	-50.3580	54.6174	-19.9872
C8(1-2)	1	4.059	9.640	10.147	-9.4600	60.9523	-109.1419	65.6498
C8(1-2)	2	4.059	10.655	10.147	7.0217	-13.5991	11.1854	-1.9627
C8(1-2)	3	4.059	11.162	10.147	21.6374	-61.9222	67.9339	-24.9406
C8(1-3)	1	4.059	8.118	10.147	-2.6083	28.2418	-60.8626	44.9671
C8(1-3)	2	4.059	10.655	10.147	7.6914	-15.0739	12.3207	-1.9826
C8(1-3)	3	4.059	11.162	10.147	24.6065	-70.4515	76.8122	-28.0001
C8(1-4)	1	4.059	9.133	10.147	3.3237	-2.1493	-10.6266	19.0006
C8(1-4)	2	4.059	10.655	10.147	8.2501	-15.9561	12.5632	-1.4998
C8(1-4)	3	4.059	11.162	10.147	27.2717	-78.1008	84.9376	-30.7944
C8(1-5)	1	4.059	9.640	10.147	-21.1879	126.0222	-225.6997	135.1910
C8(1-5)	2	4.059	10.655	10.147	8.8010	-16.5129	11.8862	-0.5596
C8(1-5)	3	4.059	11.162	10.147	30.7293	-88.2752	95.6837	-34.5726
C8(1-6)	1	4.059	9.640	10.147	-6.9600	52.7907	-103.3040	68.5130
C8(1-6)	2	4.059	10.655	10.147	9.2002	-16.1690	9.7779	0.9454
C8(1-6)	3	4.059	10.655	10.147	38.6083	-118.6779	135.7921	-52.0336
C8(1-7)	1	4.059	10.147	10.147	-9.6186	65.8242	-121.7496	75.3277
C8(1-7)	2	4.059	10.655	10.147	10.3035	-19.2856	13.4854	-0.8704
C8(1-7)	3	4.059	10.655	10.147	44.4462	-138.8820	160.3400	-62.1016
C8(2-1)	1	4.059	9.640	10.147	-1.3766	17.4573	-35.6839	23.9163
C8(2-1)	2	4.059	11.162	10.147	5.4738	-8.4273	3.5938	1.3432
C8(2-1)	3	4.059	11.162	10.147	16.6949	-46.0170	49.1985	-17.7917

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C8(2-2)	1	4.059	9.640	10.147	-12.1297	75.1309	-135.1040	80.1221
C8(2-2)	2	4.059	10.655	10.147	6.7666	-12.5772	8.6055	-0.5178
C8(2-2)	3	4.059	11.162	10.147	20.5273	-58.7688	64.0517	-23.4597
C8(2-3)	1	4.059	9.133	10.147	-5.3329	43.8101	-90.1965	61.0655
C8(2-3)	2	4.059	10.655	10.147	7.6279	-14.4327	9.8505	-0.4980
C8(2-3)	3	4.059	11.162	10.147	25.4813	-75.0902	82.7942	-30.5671
C8(2-4)	1	4.059	9.133	10.147	-2.5704	30.7749	-70.2380	51.4951
C8(2-4)	2	4.059	10.655	10.147	8.0021	-14.1329	8.0181	0.9250
C8(2-4)	3	4.059	11.162	10.147	29.3761	-87.0227	95.7618	-35.2469
C8(2-5)	1	4.059	9.640	10.147	-8.9610	64.3607	-125.8772	80.7298
C8(2-5)	2	4.059	10.655	10.147	8.6053	-14.5582	6.9499	2.0146
C8(2-5)	3	4.059	11.162	10.147	33.4168	-99.1987	108.9229	-40.0259
C8(2-6)	1	4.059	10.147	10.147	-8.3865	60.2587	-115.1367	71.8115
C8(2-6)	2	4.059	10.655	10.147	10.1193	-19.3372	12.8577	-0.6590
C8(2-6)	3	4.059	11.162	10.147	39.5125	-119.7640	132.9781	-49.4271
C8(3-1)	1	4.059	9.640	10.147	-2.4143	23.2210	-47.3952	31.3005
C8(3-1)	2	4.059	11.162	10.147	5.0691	-7.0491	1.5461	2.3041
C8(3-1)	3	4.059	11.162	10.147	15.8770	-42.8880	45.0053	-16.0142
C8(3-2)	1	4.059	10.655	10.147	-12.1744	77.3629	-143.4384	86.6121
C8(3-2)	2	4.059	11.162	10.147	5.2449	-5.8606	-1.5283	4.3084
C8(3-2)	3	4.059	11.162	10.147	20.3307	-57.4949	61.8794	-22.4823
C8(3-3)	1	4.059	9.640	10.147	-9.6003	65.0213	-124.0737	77.0066
C8(3-3)	2	4.059	10.655	10.147	6.0678	-7.2988	-1.0480	4.6435
C8(3-3)	3	4.059	11.162	10.147	28.0175	-83.4364	91.9613	-34.0431
C8(3-4)	1	4.059	10.655	10.147	-10.8080	71.3065	-133.3214	80.9046
C8(3-4)	2	4.059	10.655	10.147	6.8621	-8.6741	-0.5354	4.8597
C8(3-4)	3	4.059	11.162	10.147	36.1248	-111.7345	125.4846	-47.1285
C8(3-5)	1	4.059	10.655	10.147	-16.1453	96.3564	-168.7414	95.6969
C8(3-5)	2	4.059	10.655	10.147	8.9840	-16.1184	9.1970	0.4287
C8(3-5)	3	4.059	11.162	10.147	44.1619	-139.6954	158.8356	-60.3241
C8(4-1)	1	4.059	10.655	10.147	-4.8610	36.4084	-69.4514	42.9541
C8(4-1)	2	4.059	11.162	10.147	4.8344	-5.8617	-0.1169	3.0686
C8(4-1)	3	4.059	11.162	10.147	14.2375	-36.6774	37.4348	-12.9875
C8(4-2)	1	4.059	10.147	10.147	-13.8987	84.9563	-152.6469	89.0095
C8(4-2)	2	4.059	11.162	10.147	4.4026	-1.8331	-7.3882	7.0184
C8(4-2)	3	4.059	11.162	10.147	18.1066	-48.7699	50.8018	-17.8880
C8(4-3)	1	4.059	10.147	10.147	-12.5573	79.2993	-144.4993	85.1646
C8(4-3)	2	4.059	10.655	10.147	4.6755	-0.5126	-11.0186	9.2193
C8(4-3)	3	4.059	11.162	10.147	25.5249	-73.4521	79.2698	-28.8395
C8(4-4)	1	4.059	10.655	10.147	-17.9358	104.4563	-179.9780	100.0422
C8(4-4)	2	4.059	10.655	10.147	6.7566	-7.9355	-1.1653	4.6793
C8(4-4)	3	4.059	11.162	10.147	33.2031	-100.2050	111.2067	-41.4922

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C8(5-1)	1	4.059	10.147	10.147	-6.9230	45.3292	-81.2527	47.2093
C8(5-1)	2	4.059	11.162	10.147	4.7424	-5.5720	-0.4475	3.0989
C8(5-1)	3	4.059	11.162	10.147	14.4916	-37.3057	37.9599	-13.1798
C8(5-2)	1	4.059	10.147	10.147	-9.5737	59.8970	-105.8725	60.3496
C8(5-2)	2	4.059	11.162	10.147	4.3347	-1.7378	-7.2258	6.6234
C8(5-2)	3	4.059	11.162	10.147	16.9187	-43.3753	43.5670	-14.9375
C8(5-3)	1	4.059	10.655	10.147	-15.4878	88.8921	-149.9233	81.2437
C8(5-3)	2	4.059	11.162	10.147	5.2167	-3.2965	-6.2635	6.3912
C8(5-3)	3	4.059	11.162	10.147	22.4366	-60.6709	63.0128	-22.3663
C8(6-1)	1	4.059	10.655	10.147	-5.9223	39.1618	-68.9828	39.3233
C8(6-1)	2	4.059	11.162	10.147	4.3412	-4.0500	-2.2371	3.7078
C8(6-1)	3	4.059	11.162	10.147	15.6878	-42.8125	45.6346	-16.5432
C8(6-2)	1	4.059	10.655	10.147	-4.5106	32.2223	-57.0419	32.3663
C8(6-2)	2	4.059	11.162	10.147	4.8896	-4.5424	-2.5753	4.0469
C8(6-2)	3	4.059	11.162	10.147	20.1524	-57.8855	63.5495	-23.6565
C8(7-1)	1	4.059	10.655	10.147	-1.5369	16.5480	-31.3977	19.0726
C8(7-1)	2	4.059	11.162	10.147	5.7422	-10.0564	6.4003	-0.4338
C8(7-1)	3	4.059	11.162	10.147	17.6019	-49.6411	53.9898	-19.9824
C9(1-1)	1	4.579	11.448	11.448	-3.7621	29.9919	-55.9650	34.9565
C9(1-1)	2	4.579	12.020	11.448	5.2322	-7.2202	2.4596	1.7548
C9(1-1)	3	4.579	12.593	11.448	15.6121	-41.0641	42.7797	-15.0532
C9(1-2)	1	4.579	9.731	11.448	-4.7996	37.9730	-73.0539	48.1538
C9(1-2)	2	4.579	12.020	11.448	5.9608	-8.9675	4.6483	1.1207
C9(1-2)	3	4.579	12.020	11.448	19.3419	-54.5815	60.5499	-22.6517
C9(1-3)	1	4.579	9.731	11.448	-11.2334	74.9440	-142.7621	92.4015
C9(1-3)	2	4.579	12.020	11.448	6.4448	-9.8128	5.0068	1.4576
C9(1-3)	3	4.579	12.020	11.448	21.2228	-59.5629	65.4782	-24.2322
C9(1-4)	1	4.579	9.731	11.448	-14.0941	91.3635	-173.3524	112.0778
C9(1-4)	2	4.579	12.020	11.448	6.8191	-10.0278	4.4249	2.3100
C9(1-4)	3	4.579	12.020	11.448	22.8808	-63.6944	69.4444	-25.4012
C9(1-5)	1	4.579	9.731	11.448	-10.0975	70.4555	-137.8264	92.7922
C9(1-5)	2	4.579	11.448	11.448	7.3638	-10.9131	4.5263	2.8429
C9(1-5)	3	4.579	12.020	11.448	25.0031	-69.3725	75.0712	-27.2338
C9(1-6)	1	4.579	10.875	11.448	-18.8807	116.0082	-212.7038	131.7804
C9(1-6)	2	4.579	11.448	11.448	7.5481	-9.9984	1.8543	4.6398
C9(1-6)	3	4.579	12.020	11.448	27.9859	-79.0156	86.2761	-31.5623
C9(1-7)	1	4.579	10.875	11.448	-9.2436	65.8043	-127.0425	83.5015
C9(1-7)	2	4.579	11.448	11.448	7.8514	-9.4453	-0.3077	5.9915
C9(1-7)	3	4.579	12.020	11.448	31.4138	-90.0186	98.9588	-36.5326
C9(1-8)	1	4.579	11.448	11.448	-9.8843	68.1405	-127.1921	80.1178
C9(1-8)	2	4.579	12.020	11.448	8.5738	-11.0145	1.2138	5.1784
C9(1-8)	3	4.579	12.020	11.448	35.9512	-105.7716	118.1318	-44.4032

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C9(2-1)	1	4.579	11.448	11.448	-5.6418	39.7739	-73.2542	44.6132
C9(2-1)	2	4.579	12.020	11.448	5.1852	-7.3370	2.2457	1.9776
C9(2-1)	3	4.579	12.593	11.448	14.5789	-37.7993	38.7958	-13.4763
C9(2-2)	1	4.579	9.731	11.448	-5.4768	41.8410	-82.1466	53.8273
C9(2-2)	2	4.579	12.020	11.448	5.4598	-6.9324	0.5880	3.3020
C9(2-2)	3	4.579	12.020	11.448	17.9894	-50.3935	55.3164	-20.6048
C9(2-3)	1	4.579	9.731	11.448	-13.9877	90.5238	-172.3847	108.8970
C9(2-3)	2	4.579	11.448	11.448	6.8205	-11.5910	6.5222	0.9068
C9(2-3)	3	4.579	12.020	11.448	20.8879	-59.1735	64.9667	-24.1433
C9(2-4)	1	4.579	10.303	11.448	-41.7997	239.8375	-431.0182	254.2237
C9(2-4)	2	4.579	11.448	11.448	7.2319	-11.7443	5.5233	1.9443
C9(2-4)	3	4.579	12.020	11.448	23.2125	-65.7775	71.8293	-26.4873
C9(2-5)	1	4.579	10.875	11.448	-25.5103	152.2681	-277.9347	167.4063
C9(2-5)	2	4.579	11.448	11.448	7.3843	-10.4588	2.1580	4.1368
C9(2-5)	3	4.579	12.020	11.448	25.7019	-72.6627	78.8827	-28.9138
C9(2-6)	1	4.579	10.875	11.448	-11.3233	77.9153	-151.0511	96.8187
C9(2-6)	2	4.579	11.448	11.448	7.6123	-9.4916	-0.6664	5.9153
C9(2-6)	3	4.579	12.020	11.448	30.0028	-87.2776	96.3309	-35.8607
C9(2-7)	1	4.579	11.448	11.448	-13.0880	85.4744	-158.7471	97.0136
C9(2-7)	2	4.579	11.448	11.448	8.8875	-13.7219	4.9964	3.1057
C9(2-7)	3	4.579	12.020	11.448	34.7635	-104.0088	117.0295	-44.4539
C9(3-1)	1	4.579	10.875	11.448	-2.8660	26.1050	-53.4574	35.7670
C9(3-1)	2	4.579	12.020	11.448	4.7604	-5.7237	-0.3976	3.3444
C9(3-1)	3	4.579	12.593	11.448	14.0257	-35.6707	35.8210	-12.1837
C9(3-2)	1	4.579	10.875	11.448	-8.0361	55.9569	-108.5434	69.2350
C9(3-2)	2	4.579	12.020	11.448	4.7609	-3.7570	-4.6771	5.9842
C9(3-2)	3	4.579	12.593	11.448	16.9511	-44.6544	45.7424	-15.8155
C9(3-3)	1	4.579	10.303	11.448	-10.3892	70.6147	-137.2306	87.4754
C9(3-3)	2	4.579	12.020	11.448	5.2688	-3.9706	-5.8657	7.1119
C9(3-3)	3	4.579	12.593	11.448	22.0451	-61.3850	64.9106	-23.1052
C9(3-4)	1	4.579	10.875	11.448	-8.0143	58.9490	-117.9320	77.0422
C9(3-4)	2	4.579	12.020	11.448	5.7543	-4.2579	-6.8252	8.0899
C9(3-4)	3	4.579	12.593	11.448	26.6871	-76.9255	82.9115	-29.9822
C9(3-5)	1	4.579	11.448	11.448	-20.7412	123.3911	-221.2636	129.6576
C9(3-5)	2	4.579	12.020	11.448	6.1934	-4.3046	-7.9943	8.9852
C9(3-5)	3	4.579	12.020	11.448	32.8681	-100.2042	113.4745	-43.3338
C9(3-6)	1	4.579	11.448	11.448	-8.0321	58.1416	-112.0765	70.0666
C9(3-6)	2	4.579	12.020	11.448	7.4857	-8.3756	-2.8231	6.5434
C9(3-6)	3	4.579	12.593	11.448	36.1042	-107.9326	118.6809	-43.7935
C9(4-1)	1	4.579	10.875	11.448	-2.0103	21.9817	-46.4343	31.7186
C9(4-1)	2	4.579	12.020	11.448	4.3455	-3.6490	-3.3571	4.6971
C9(4-1)	3	4.579	12.593	11.448	12.4734	-29.4438	27.9381	-8.9432

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C9(4-2)	1	4.579	11.448	11.448	-15.8020	96.2454	-174.9690	103.9753
C9(4-2)	2	4.579	12.020	11.448	3.9827	0.1790	-10.6988	8.9188
C9(4-2)	3	4.579	12.593	11.448	14.6899	-35.5052	33.9838	-10.9103
C9(4-3)	1	4.579	10.875	11.448	-24.7987	145.6134	-261.3806	152.7196
C9(4-3)	2	4.579	12.020	11.448	4.3122	0.9249	-13.3446	10.6945
C9(4-3)	3	4.579	12.593	11.448	19.8134	-52.1502	52.9433	-18.1184
C9(4-4)	1	4.579	11.448	11.448	-8.7838	63.3153	-124.3909	78.8741
C9(4-4)	2	4.579	12.020	11.448	4.6305	1.4630	-15.4076	12.0181
C9(4-4)	3	4.579	12.593	11.448	25.0015	-69.7984	73.7762	-26.2878
C9(4-5)	1	4.579	11.448	11.448	-11.0922	73.7519	-137.3600	82.6207
C9(4-5)	2	4.579	12.020	11.448	5.8940	-2.4258	-10.5265	9.7093
C9(4-5)	3	4.579	12.593	11.448	29.8436	-86.0044	92.8729	-33.8596
C9(5-1)	1	4.579	11.448	11.448	-3.7342	30.6102	-60.4327	38.6801
C9(5-1)	2	4.579	12.020	11.448	4.2923	-3.5263	-3.6169	4.8210
C9(5-1)	3	4.579	12.593	11.448	12.4308	-29.0143	27.0141	-8.4589
C9(5-2)	1	4.579	11.448	11.448	-8.1503	55.7900	-105.6015	64.7275
C9(5-2)	2	4.579	12.020	11.448	3.8303	0.6572	-11.2015	8.9488
C9(5-2)	3	4.579	12.593	11.448	14.5725	-34.4251	32.0241	-9.9839
C9(5-3)	1	4.579	11.448	11.448	-12.2271	77.2994	-141.0863	83.2872
C9(5-3)	2	4.579	12.020	11.448	3.8198	2.9651	-15.9731	11.5656
C9(5-3)	3	4.579	12.593	11.448	19.0539	-48.2135	47.2066	-15.6339
C9(5-4)	1	4.579	11.448	11.448	-8.7070	58.6282	-107.8587	63.8059
C9(5-4)	2	4.579	12.020	11.448	4.7645	0.3909	-12.9182	10.1234
C9(5-4)	3	4.579	12.593	11.448	24.5549	-67.2360	70.0327	-24.7658
C9(6-1)	1	4.579	11.448	11.448	-6.1748	42.4182	-78.0368	46.6202
C9(6-1)	2	4.579	12.020	11.448	4.2100	-3.3384	-3.4385	4.4817
C9(6-1)	3	4.579	12.593	11.448	11.9703	-27.5708	25.6823	-8.1007
C9(6-2)	1	4.579	10.875	11.448	-11.1837	69.1241	-122.8923	70.6246
C9(6-2)	2	4.579	12.020	11.448	3.7464	0.7898	-10.8054	8.3602
C9(6-2)	3	4.579	12.593	11.448	14.3857	-34.3928	32.7529	-10.5727
C9(6-3)	1	4.579	11.448	11.448	-12.6468	76.5436	-133.6072	75.0428
C9(6-3)	2	4.579	12.020	11.448	4.7282	-1.6215	-8.2425	7.2462
C9(6-3)	3	4.579	12.593	11.448	19.7646	-52.0617	53.2076	-18.5365
C9(7-1)	1	4.579	11.448	11.448	-3.3649	26.8172	-50.2444	30.5246
C9(7-1)	2	4.579	12.020	11.448	4.3679	-4.2597	-1.8349	3.5298
C9(7-1)	3	4.579	12.593	11.448	13.0602	-32.0933	31.6451	-10.6638
C9(7-2)	1	4.579	11.448	11.448	-2.5602	22.9781	-43.3594	26.1659
C9(7-2)	2	4.579	12.020	11.448	4.9530	-5.2862	-1.1471	3.3128
C9(7-2)	3	4.579	12.593	11.448	15.6141	-39.9809	40.6540	-14.1863
C9(8-1)	1	4.579	11.448	11.448	-1.3814	16.3669	-32.1650	20.2820
C9(8-1)	2	4.579	12.593	11.448	5.0282	-6.8983	1.8355	1.7750
C9(8-1)	3	4.579	12.593	11.448	15.3373	-40.9263	42.8360	-15.2817

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C10(1-1)	1	5.074	11.416	12.684	0.9426	6.7176	-18.9644	16.2600
C10(1-1)	2	5.074	13.318	12.684	6.0746	-10.4863	6.4195	0.3158
C10(1-1)	3	5.074	13.953	12.684	15.4312	-40.7121	42.6289	-15.0367
C10(1-2)	1	5.074	11.416	12.684	-5.8886	46.5976	-92.8978	62.4816
C10(1-2)	2	5.074	12.684	12.684	7.5495	-15.9810	14.5257	-3.2741
C10(1-2)	3	5.074	13.953	12.684	17.2617	-45.9373	48.5746	-17.1686
C10(1-3)	1	5.074	9.513	12.684	-5.0737	45.2849	-97.5991	71.4345
C10(1-3)	2	5.074	12.684	12.684	8.0923	-17.2080	15.5219	-3.2648
C10(1-3)	3	5.074	13.318	12.684	20.2433	-56.8294	62.9157	-23.4225
C10(1-4)	1	5.074	10.782	12.684	0.1584	19.0690	-56.0027	52.0922
C10(1-4)	2	5.074	12.684	12.684	8.7003	-18.5853	16.6828	-3.2143
C10(1-4)	3	5.074	13.318	12.684	21.5297	-59.9975	66.0902	-24.4079
C10(1-5)	1	5.074	10.782	12.684	-2.7350	34.8396	-83.7029	68.3107
C10(1-5)	2	5.074	12.684	12.684	9.2214	-19.3813	16.5381	-2.4552
C10(1-5)	3	5.074	12.684	12.684	24.6282	-71.2641	81.0390	-31.0473
C10(1-6)	1	5.074	11.416	12.684	-25.2182	153.9604	-286.2562	179.0405
C10(1-6)	2	5.074	12.050	12.684	11.2078	-27.5689	28.4062	-8.0313
C10(1-6)	3	5.074	12.684	12.684	26.3462	-76.2579	86.5773	-33.0925
C10(1-7)	1	5.074	11.416	12.684	-6.9356	56.9179	-118.6916	84.4739
C10(1-7)	2	5.074	12.684	12.684	10.1961	-20.5643	15.7527	-1.0766
C10(1-7)	3	5.074	12.684	12.684	28.6653	-83.3629	94.6462	-36.2346
C10(1-8)	1	5.074	12.050	12.684	-8.9363	66.8266	-132.6371	89.1829
C10(1-8)	2	5.074	12.684	12.684	10.6655	-20.9735	15.1564	-0.5290
C10(1-8)	3	5.074	12.684	12.684	32.0542	-95.1692	109.2505	-42.3298
C10(1-9)	1	5.074	12.050	12.684	-3.8704	40.1002	-85.3839	60.8636
C10(1-9)	2	5.074	12.684	12.684	11.6659	-24.1511	19.3780	-2.7720
C10(1-9)	3	5.074	12.684	12.684	35.5638	-107.4140	124.6251	-48.9304
C10(2-1)	1	5.074	11.416	12.684	-0.1934	12.5982	-29.7281	22.4328
C10(2-1)	2	5.074	12.684	12.684	6.1449	-11.5838	8.2689	-0.7027
C10(2-1)	3	5.074	13.953	12.684	14.3671	-37.1273	38.0304	-13.1336
C10(2-2)	1	5.074	11.416	12.684	-9.1390	63.3609	-122.8148	78.9622
C10(2-2)	2	5.074	12.684	12.684	6.6756	-12.4211	8.4477	-0.2287
C10(2-2)	3	5.074	13.318	12.684	17.0283	-46.8409	50.8782	-18.7292
C10(2-3)	1	5.074	10.782	12.684	-9.1379	67.2526	-137.5448	93.4953
C10(2-3)	2	5.074	12.684	12.684	7.5848	-14.9739	11.0873	-0.9309
C10(2-3)	3	5.074	13.318	12.684	19.9357	-56.2587	61.8005	-22.9197
C10(2-4)	1	5.074	10.782	12.684	-6.2425	53.0663	-115.5147	83.4256
C10(2-4)	2	5.074	12.684	12.684	8.3889	-17.0849	13.0775	-1.3130
C10(2-4)	3	5.074	13.318	12.684	21.9580	-62.2024	68.2744	-25.2353
C10(2-5)	1	5.074	11.416	12.684	-15.8872	105.0628	-205.3707	133.6285
C10(2-5)	2	5.074	12.684	12.684	9.0975	-18.5240	13.8048	-1.0352
C10(2-5)	3	5.074	12.684	12.684	24.8769	-72.6573	82.2592	-31.5732

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C10(2-6)	1	5.074	11.416	12.684	-5.6730	50.6795	-110.5903	79.6079
C10(2-6)	2	5.074	12.684	12.684	9.1196	-16.8148	9.8694	1.4126
C10(2-6)	3	5.074	12.684	12.684	26.9828	-78.8535	89.0355	-34.0741
C10(2-7)	1	5.074	12.050	12.684	-15.5388	101.5350	-193.1724	121.6783
C10(2-7)	2	5.074	12.050	12.684	10.5217	-21.9795	17.0672	-2.0673
C10(2-7)	3	5.074	12.684	12.684	29.7186	-87.8810	99.9823	-38.5980
C10(2-8)	1	5.074	12.050	12.684	-7.5130	59.0477	-118.8265	78.5542
C10(2-8)	2	5.074	12.684	12.684	10.5880	-20.2944	13.2765	-0.0296
C10(2-8)	3	5.074	12.684	12.684	33.9866	-103.2448	119.4868	-46.9442
C10(3-1)	1	5.074	11.416	12.684	-1.5238	19.3615	-42.5518	30.3814
C10(3-1)	2	5.074	12.684	12.684	5.9781	-11.1014	7.2188	-0.0582
C10(3-1)	3	5.074	13.953	12.684	13.9469	-35.5476	35.7796	-12.1468
C10(3-2)	1	5.074	12.050	12.684	-17.8289	107.2529	-195.1642	116.9250
C10(3-2)	2	5.074	12.684	12.684	6.2549	-10.5054	4.9811	1.6450
C10(3-2)	3	5.074	13.318	12.684	16.6087	-44.7563	47.4626	-17.0957
C10(3-3)	1	5.074	11.416	12.684	-12.6191	83.1865	-160.6804	102.4162
C10(3-3)	2	5.074	12.684	12.684	7.1468	-12.7310	6.8796	1.3128
C10(3-3)	3	5.074	13.318	12.684	20.6907	-58.5567	63.9369	-23.6445
C10(3-4)	1	5.074	12.050	12.684	-11.8658	80.1617	-156.4910	100.7298
C10(3-4)	2	5.074	12.684	12.684	7.9379	-14.5785	8.3117	1.1604
C10(3-4)	3	5.074	13.318	12.684	23.9196	-69.1597	76.3069	-28.4389
C10(3-5)	1	5.074	12.050	12.684	-9.2842	67.2059	-134.6841	88.5761
C10(3-5)	2	5.074	12.684	12.684	8.2589	-14.2013	6.4150	2.5573
C10(3-5)	3	5.074	12.684	12.684	28.3126	-85.9066	99.1209	-38.8558
C10(3-6)	1	5.074	12.050	12.684	-7.7294	59.1484	-119.8199	79.1198
C10(3-6)	2	5.074	12.684	12.684	8.6184	-14.0494	5.0121	3.5588
C10(3-6)	3	5.074	12.684	12.684	31.4901	-96.3980	111.6048	-43.8884
C10(3-7)	1	5.074	12.050	12.684	-5.6165	47.4391	-97.2948	64.4315
C10(3-7)	2	5.074	12.684	12.684	9.8418	-18.2108	10.7102	0.6994
C10(3-7)	3	5.074	13.318	12.684	32.8461	-97.5803	109.2061	-41.3739
C10(4-1)	1	5.074	12.684	12.684	-18.0817	103.9784	-180.4943	102.4569
C10(4-1)	2	5.074	12.684	12.684	6.0350	-11.0603	7.0869	0.0414
C10(4-1)	3	5.074	13.953	12.684	13.5176	-33.8167	33.6656	-11.2831
C10(4-2)	1	5.074	12.050	12.684	-24.5379	142.6397	-254.7201	148.8356
C10(4-2)	2	5.074	12.684	12.684	6.2222	-9.8826	3.6545	2.4536
C10(4-2)	3	5.074	13.318	12.684	16.0556	-42.2160	43.9619	-15.5376
C10(4-3)	1	5.074	12.050	12.684	-17.4508	108.5767	-202.3524	123.5448
C10(4-3)	2	5.074	12.684	12.684	6.6331	-9.6156	1.6064	4.0654
C10(4-3)	3	5.074	13.318	12.684	19.9314	-54.7215	58.3462	-21.0924
C10(4-4)	1	5.074	12.050	12.684	-11.5952	79.1939	-154.1900	97.9359
C10(4-4)	2	5.074	12.684	12.684	6.9681	-9.3234	-0.1079	5.3265
C10(4-4)	3	5.074	13.318	12.684	23.6057	-67.1941	73.3154	-27.0842

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C10(4-5)	1	5.074	12.050	12.684	-10.2861	72.7071	-142.6042	90.7220
C10(4-5)	2	5.074	12.684	12.684	7.5242	-9.9943	-0.3128	5.7302
C10(4-5)	3	5.074	13.318	12.684	26.9868	-78.1776	86.1803	-32.1985
C10(4-6)	1	5.074	12.050	12.684	-7.4934	57.3550	-113.8044	72.6480
C10(4-6)	2	5.074	12.684	12.684	8.9671	-15.2121	7.0106	2.0834
C10(4-6)	3	5.074	13.318	12.684	31.5266	-94.1528	105.8675	-40.3226
C10(5-1)	1	5.074	12.050	12.684	-6.1340	43.3952	-82.4362	51.2621
C10(5-1)	2	5.074	13.318	12.684	5.4297	-8.0509	2.1005	2.6099
C10(5-1)	3	5.074	13.953	12.684	14.1303	-36.0862	36.3301	-12.3315
C10(5-2)	1	5.074	12.050	12.684	-7.8021	55.2150	-107.4708	68.1017
C10(5-2)	2	5.074	12.684	12.684	6.1942	-9.6794	3.0870	2.7572
C10(5-2)	3	5.074	13.953	12.684	14.8827	-36.2914	34.8157	-11.1553
C10(5-3)	1	5.074	12.050	12.684	-10.0930	68.9283	-132.8448	83.0580
C10(5-3)	2	5.074	12.684	12.684	6.0109	-6.6326	-2.9713	6.1645
C10(5-3)	3	5.074	13.953	12.684	17.8505	-44.6024	43.2960	-14.0838
C10(5-4)	1	5.074	12.050	12.684	-9.6983	67.1051	-128.9284	80.0001
C10(5-4)	2	5.074	12.684	12.684	6.1310	-5.4572	-5.6804	7.6648
C10(5-4)	3	5.074	13.318	12.684	22.4536	-61.4675	65.0574	-23.4808
C10(5-5)	1	5.074	12.050	12.684	-7.8828	57.0796	-109.7189	67.6764
C10(5-5)	2	5.074	12.684	12.684	7.5030	-10.3220	1.1323	4.2512
C10(5-5)	3	5.074	13.318	12.684	26.6954	-76.3354	83.5587	-31.2224
C10(6-1)	1	5.074	12.050	12.684	-1.8708	21.5309	-45.9467	31.3263
C10(6-1)	2	5.074	12.684	12.684	5.8055	-9.9536	5.2785	0.8889
C10(6-1)	3	5.074	13.953	12.684	13.2643	-32.8166	32.3121	-10.7282
C10(6-2)	1	5.074	12.050	12.684	-3.4669	32.0188	-66.7409	44.3685
C10(6-2)	2	5.074	12.684	12.684	5.2247	-5.2775	-3.1819	5.5300
C10(6-2)	3	5.074	13.953	12.684	14.0457	-33.4105	31.5480	-9.9534
C10(6-3)	1	5.074	12.050	12.684	-7.7161	55.1086	-105.9618	65.5136
C10(6-3)	2	5.074	12.684	12.684	5.3427	-3.6665	-6.9752	7.7390
C10(6-3)	3	5.074	13.318	12.684	18.9999	-50.9053	53.4468	-19.1659
C10(6-4)	1	5.074	12.050	12.684	-6.9076	50.4376	-96.0173	58.4642
C10(6-4)	2	5.074	12.684	12.684	6.3675	-7.0379	-2.3137	5.3629
C10(6-4)	3	5.074	13.318	12.684	23.5444	-66.8836	73.1330	-27.2866
C10(7-1)	1	5.074	12.684	12.684	-7.1469	47.3534	-86.1224	51.0059
C10(7-1)	2	5.074	12.684	12.684	5.6684	-9.5888	5.1145	0.7328
C10(7-1)	3	5.074	13.953	12.684	13.2135	-32.4676	31.8515	-10.5901
C10(7-2)	1	5.074	12.050	12.684	-6.8619	47.1108	-86.9968	51.8776
C10(7-2)	2	5.074	12.684	12.684	5.4198	-6.6196	-0.4871	3.7586
C10(7-2)	3	5.074	13.318	12.684	15.5029	-40.0515	41.4157	-14.7376
C10(7-3)	1	5.074	12.050	12.684	-4.8228	36.7835	-69.0469	41.4249
C10(7-3)	2	5.074	12.684	12.684	6.6892	-10.6618	4.6907	1.3481
C10(7-3)	3	5.074	13.953	12.684	17.7344	-45.0810	45.1179	-15.4301

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
C10(8-1)	1	5.074	12.684	12.684	-4.1291	31.6568	-59.5976	36.3720
C10(8-1)	2	5.074	12.684	12.684	5.6606	-9.8392	5.8893	0.1491
C10(8-1)	3	5.074	13.953	12.684	13.6935	-34.9829	35.5805	-12.2989
C10(8-2)	1	5.074	12.050	12.684	-2.3900	22.8871	-44.4238	27.5027
C10(8-2)	2	5.074	12.684	12.684	6.4560	-12.1401	8.7724	-1.2415
C10(8-2)	3	5.074	13.318	12.684	17.1291	-47.7139	52.3663	-19.6927
C10(9-1)	1	5.074	12.684	12.684	-1.4458	17.1399	-34.2725	22.0474
C10(9-1)	2	5.074	13.318	12.684	5.7214	-9.7112	5.3163	0.4969
C10(9-1)	3	5.074	13.953	12.684	15.0544	-39.9684	41.7586	-14.8631
Linear Alkyl Benzenes								
B6A4C8	1	5.565	13.217	13.912	-6.7825	47.9076	-90.3069	55.1941
B6A4C8	2	5.565	14.608	13.912	5.6112	-8.5240	3.3193	1.6089
B6A4C8	3	5.565	14.608	13.912	15.0169	-41.8004	46.4674	-17.6117
B6A4C10	1	6.567	14.775	16.417	-3.1283	30.0616	-62.8663	42.3667
B6A4C10	2	6.567	16.417	16.417	6.1162	-11.0326	7.1403	-0.0874
B6A4C10	3	6.567	17.238	16.417	12.7948	-32.6364	34.3791	-12.4357
B6A4C12	1	7.572	17.037	18.930	-10.0664	67.0609	-126.8753	78.7360
B6A4C12	2	7.572	18.930	18.930	4.9124	-5.3089	-1.5776	4.3271
B6A4C12	3	7.572	18.930	18.930	11.0980	-25.4441	24.7702	-8.3021
B5A4C8	1	5.020	10.666	12.549	1.0504	9.6495	-25.0676	19.2667
B5A4C8	2	5.020	13.804	12.549	7.5588	-13.7488	10.1786	-1.5986
B5A4C8	3	5.020	13.804	12.549	18.3858	-50.2825	54.7296	-20.2438
B5A4C10	1	6.028	12.056	15.069	0.5966	12.2711	-30.4229	23.2881
B5A4C10	2	6.028	15.823	15.069	6.6531	-10.1418	5.2111	0.7895
B5A4C10	3	6.028	16.576	15.069	15.9368	-40.7308	42.3823	-14.9855
B5A4C12	1	7.038	14.955	17.595	0.5549	12.9835	-32.6524	25.4876
B5A4C12	2	7.038	17.595	17.595	7.0131	-11.7184	7.2311	0.1238
B5A4C12	3	7.038	18.474	17.595	15.6021	-40.3150	42.8876	-15.5667
B3A4C8	1	4.796	10.192	11.990	0.0801	18.9454	-41.1809	27.9477
B3A4C8	2	4.796	13.189	11.990	7.2727	-10.3186	6.2263	-0.0941
B3A4C8	3	4.796	13.189	11.990	18.6286	-49.3741	53.9695	-19.9583
B3A4C10	1	5.803	11.605	14.506	3.0911	3.5336	-16.0003	15.1469
B3A4C10	2	5.803	15.232	14.506	7.1754	-9.8195	5.2557	0.6312
B3A4C10	3	5.803	15.957	14.506	15.9922	-39.1548	41.0383	-14.5545
B3A4C12	1	6.811	13.623	17.029	2.8464	4.8165	-18.2276	16.7733
B3A4C12	2	6.811	17.029	17.029	7.1346	-9.5767	4.7775	1.0758
B3A4C12	3	6.811	17.880	17.029	15.5749	-38.2833	41.0272	-14.9593
B1A4C8	1	4.315	10.786	10.786	-2.1018	28.8612	-53.4154	31.8783
B1A4C8	2	4.315	11.865	10.786	5.7838	-5.2530	0.7257	2.0348
B1A4C8	3	4.315	11.865	10.786	16.0212	-41.5190	46.1136	-17.2170
B1A4C10	1	5.324	13.309	13.309	-1.4457	26.4782	-51.3224	32.2919
B1A4C10	2	5.324	13.974	13.309	5.9046	-5.0291	-0.2465	2.8583

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
B1A4C10	3	5.324	13.974	13.309	14.7498	-36.4832	40.1746	-14.9719
B1A4C12	1	6.334	15.044	15.836	-0.7440	24.5090	-50.6560	33.7889
B1A4C12	2	6.334	15.836	15.836	6.9299	-8.8929	4.7248	0.9067
B1A4C12	3	6.334	16.628	15.836	14.7378	-35.9148	39.2979	-14.5564
B0A4C8	1	4.052	10.637	10.131	-2.7825	28.4399	-51.8366	30.2374
B0A4C8	2	4.052	11.144	10.131	5.5097	-7.5880	4.3982	0.1274
B0A4C8	3	4.052	11.144	10.131	16.1889	-45.2842	50.9926	-19.2791
B0A4C10	1	5.067	12.668	12.668	-1.3601	22.5295	-44.6144	28.1946
B0A4C10	2	5.067	13.301	12.668	5.8399	-7.8268	3.6172	1.0040
B0A4C10	3	5.067	13.934	12.668	14.6396	-38.2005	41.4414	-15.1685
B0A4C12	1	6.082	14.444	15.205	-1.7432	25.2255	-50.6114	32.7993
B0A4C12	2	6.082	15.205	15.205	5.5551	-6.4740	1.7126	2.0245
B0A4C12	3	6.082	15.965	15.205	13.2911	-33.0495	35.4927	-12.9641
Branched Alkyl Benzenes								
B6A4C5(1-3)	1	4.168	10.420	10.420	-12.2162	77.6453	-139.0037	79.8127
B6A4C5(1-3)	2	4.168	11.462	10.420	2.5374	9.9876	-24.4963	14.4389
B6A4C5(1-3)	3	4.168	11.462	10.420	44.6021	-144.8538	169.4252	-65.9421
B6A4C6(1-2)	1	4.569	10.851	11.422	0.8592	11.5144	-33.7676	26.9213
B6A4C6(1-2)	2	4.569	12.564	11.422	3.7917	2.2780	-12.5915	8.8433
B6A4C6(1-2)	3	4.569	12.564	11.422	19.3534	-50.6509	52.8052	-18.8296
B6A4C7(1-1)	1	5.142	11.569	12.854	-17.1671	105.0969	-191.8381	113.5588
B6A4C7(1-1)	2	5.142	14.140	12.854	4.9161	-3.8152	-4.1084	5.2102
B6A4C7(1-1)	3	5.142	14.140	12.854	16.2271	-40.9388	41.2796	-14.1906
B6A4C6(1-4)	1	4.569	11.993	11.422	-21.1087	121.6033	-207.6173	114.2757
B6A4C6(1-4)	2	4.569	12.564	11.422	3.8940	5.7009	-19.4926	12.5051
B6A4C6(1-4)	3	4.569	12.564	11.422	38.3740	-119.0321	136.1444	-52.2129
B6A4C7(1-3)	1	5.142	12.212	12.854	-17.0674	107.8831	-200.4260	120.6160
B6A4C7(1-3)	2	5.142	13.497	12.854	5.0215	-0.6302	-10.9333	9.2477
B6A4C7(1-3)	3	5.142	14.140	12.854	19.3909	-47.5160	46.5698	-15.5828
B6A4C8(1-2)	1	5.565	12.521	13.912	0.1478	16.0418	-44.0177	35.2190
B6A4C8(1-2)	2	5.565	14.608	13.912	5.3084	-3.4541	-6.3343	7.0124
B6A4C8(1-2)	3	5.565	15.304	13.912	15.8898	-37.9771	36.8284	-12.1565
B6A4C9(1-1)	1	6.129	13.790	15.322	0.7494	10.7896	-32.1974	27.0775
B6A4C9(1-1)	2	6.129	16.088	15.322	5.5880	-6.2717	-1.6055	4.6723
B6A4C9(1-1)	3	6.129	16.088	15.322	14.3053	-34.3366	33.8124	-11.4645
B6A4C7(1-5)	1	5.142	12.854	12.854	-16.5343	104.4118	-190.5716	112.2032
B6A4C7(1-5)	2	5.142	13.497	12.854	5.8828	-1.1911	-12.0402	10.2497
B6A4C7(1-5)	3	5.142	14.140	12.854	28.1755	-78.2574	83.8612	-30.5355
B6A4C8(1-4)	1	5.565	13.217	13.912	-9.0492	66.8122	-132.2512	84.6281
B6A4C8(1-4)	2	5.565	14.608	13.912	6.2832	-4.2537	-7.5467	8.4516
B6A4C8(1-4)	3	5.565	14.608	13.912	22.1716	-58.5594	61.6604	-22.2949
B6A4C9(1-3)	1	6.129	13.024	15.322	-7.8170	61.0609	-126.2787	84.8468

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
B6A4C9(1-3)	2	6.129	15.322	15.322	7.1718	-10.3358	2.4626	3.5502
B6A4C9(1-3)	3	6.129	16.088	15.322	16.7357	-39.0713	37.1723	-12.0766
B6A4C10(1-2)	1	6.567	14.775	16.417	-27.4799	163.0705	-297.6183	177.8354
B6A4C10(1-2)	2	6.567	16.417	16.417	6.6222	-9.8811	2.9090	3.0400
B6A4C10(1-2)	3	6.567	17.238	16.417	14.9792	-36.0935	35.7718	-12.1291
B6A4C11(1-1)	1	7.124	16.028	17.809	-7.8199	56.7670	-112.3803	73.0516
B6A4C11(1-1)	2	7.124	17.809	17.809	5.7278	-7.2840	-0.0768	4.1702
B6A4C11(1-1)	3	7.124	18.700	17.809	11.9404	-25.3375	22.0792	-6.3450
B5A4C5(1-3)	1	3.591	8.976	8.976	-15.8541	94.6520	-159.2692	86.6141
B5A4C5(1-3)	2	3.591	9.874	8.976	1.4936	15.7501	-27.7463	13.2756
B5A4C5(1-3)	3	3.591	9.874	8.976	44.9093	-140.9214	166.9253	-66.7226
B5A4C6(1-2)	1	4.015	9.535	10.036	2.3528	5.2531	-20.2898	18.0612
B5A4C6(1-2)	2	4.015	11.040	10.036	4.3040	3.7639	-14.1268	8.7229
B5A4C6(1-2)	3	4.015	11.040	10.036	32.0506	-95.7587	109.8566	-42.7100
B5A4C7(1-1)	1	4.576	9.724	11.440	3.7667	-5.2089	1.7175	4.0312
B5A4C7(1-1)	2	4.576	12.585	11.440	6.4461	-7.3109	0.9156	2.5466
B5A4C7(1-1)	3	4.576	12.585	11.440	14.7210	-31.5468	28.9188	-9.2198
B5A4C6(1-4)	1	4.015	10.538	10.036	-15.3155	94.0955	-160.8913	88.6949
B5A4C6(1-4)	2	4.015	11.040	10.036	3.7009	9.5427	-22.8647	12.4862
B5A4C6(1-4)	3	4.015	11.040	10.036	48.3332	-149.1984	171.7361	-66.8316
B5A4C7(1-3)	1	4.576	10.296	11.440	-13.5080	91.1369	-169.7686	102.5007
B5A4C7(1-3)	2	4.576	12.585	11.440	3.9516	8.3955	-23.5279	14.0520
B5A4C7(1-3)	3	4.576	12.585	11.440	20.8524	-45.2226	40.4989	-12.7714
B5A4C8(1-2)	1	5.020	11.294	12.549	1.2327	14.0629	-40.9715	33.6000
B5A4C8(1-2)	2	5.020	13.176	12.549	8.5568	-13.3900	6.9144	0.7515
B5A4C8(1-2)	3	5.020	13.176	12.549	21.4995	-54.7480	56.9488	-20.6313
B5A4C9(1-1)	1	5.572	11.841	13.931	-8.3883	62.1317	-119.2476	74.9209
B5A4C9(1-1)	2	5.572	14.627	13.931	7.2360	-10.6503	4.8440	1.2873
B5A4C9(1-1)	3	5.572	15.324	13.931	15.9425	-37.1121	36.0789	-12.0727
B5A4C7(1-5)	1	4.576	11.440	11.440	-12.9456	87.1451	-158.9581	93.6127
B5A4C7(1-5)	2	4.576	12.585	11.440	4.5559	8.5343	-25.1428	15.1371
B5A4C7(1-5)	3	4.576	12.585	11.440	35.3439	-96.1183	101.2087	-36.6517
B5A4C8(1-4)	1	5.020	11.921	12.549	-8.3690	66.9379	-132.8122	84.6904
B5A4C8(1-4)	2	5.020	13.176	12.549	8.2240	-8.1394	-2.8782	5.8887
B5A4C8(1-4)	3	5.020	13.804	12.549	29.0639	-75.9400	77.7383	-27.3697
B5A4C9(1-3)	1	5.572	11.841	13.931	-13.3188	94.5754	-185.7389	118.8875
B5A4C9(1-3)	2	5.572	14.627	13.931	8.8465	-12.8611	4.2269	2.8638
B5A4C9(1-3)	3	5.572	15.324	13.931	21.4966	-50.5419	47.4404	-15.2146
B5A4C10(1-2)	1	6.028	12.809	15.069	3.8046	0.0844	-18.0516	23.0897
B5A4C10(1-2)	2	6.028	15.069	15.069	8.0467	-12.4936	6.1664	1.2887
B5A4C10(1-2)	3	6.028	15.823	15.069	19.3611	-48.1577	48.9312	-17.1289
B5A4C11(1-1)	1	6.574	13.970	16.435	1.1129	10.0793	-26.7466	22.0041

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
B5A4C11(1-1)	2	6.574	16.435	16.435	6.9952	-9.8070	3.3866	2.2952
B5A4C11(1-1)	3	6.574	17.257	16.435	16.2034	-39.3129	39.5554	-13.6325
B3A4C5(1-3)	1	3.278	8.196	8.196	-12.3581	79.7483	-132.2474	70.5993
B3A4C5(1-3)	2	3.278	9.016	8.196	2.6946	13.7201	-24.1513	11.2295
B3A4C5(1-3)	3	3.278	9.016	8.196	48.3328	-151.0491	182.5575	-74.5331
B3A4C6(1-2)	1	3.684	9.669	9.209	-6.0398	51.9735	-94.6298	54.8376
B3A4C6(1-2)	2	3.684	10.130	9.209	5.9088	0.9244	-10.3304	6.8593
B3A4C6(1-2)	3	3.684	10.130	9.209	21.2049	-45.0909	42.9613	-14.8585
B3A4C7(1-1)	1	4.249	10.622	10.622	0.5096	17.4396	-38.3742	26.0061
B3A4C7(1-1)	2	4.249	11.684	10.622	7.9990	-10.0314	4.1276	1.1510
B3A4C7(1-1)	3	4.249	11.684	10.622	18.9294	-40.8505	37.3140	-11.8001
B3A4C6(1-4)	1	3.684	9.669	9.209	-5.7586	51.6945	-94.4053	54.7849
B3A4C6(1-4)	2	3.684	10.130	9.209	5.0925	6.6226	-18.2469	10.1191
B3A4C6(1-4)	3	3.684	10.130	9.209	50.3553	-152.5102	175.2857	-68.1190
B3A4C7(1-3)	1	4.249	10.090	10.622	-10.2474	77.3437	-143.6640	85.7573
B3A4C7(1-3)	2	4.249	11.684	10.622	8.1403	-6.2487	-3.4003	5.0195
B3A4C7(1-3)	3	4.249	11.684	10.622	31.5611	-82.8345	87.0329	-31.5796
B3A4C8(1-2)	1	4.680	11.115	11.700	-4.9545	47.6307	-90.9983	55.9689
B3A4C8(1-2)	2	4.680	12.870	11.700	11.2364	-21.3135	17.5571	-3.9438
B3A4C8(1-2)	3	4.680	12.870	11.700	27.6367	-72.5781	76.3925	-27.4999
B3A4C9(1-1)	1	5.236	11.780	13.089	0.2882	18.2728	-39.7733	27.4188
B3A4C9(1-1)	2	5.236	13.744	13.089	9.0449	-15.6603	12.4867	-2.5107
B3A4C9(1-1)	3	5.236	14.398	13.089	21.4805	-54.5129	57.0413	-20.3890
B3A4C7(1-5)	1	4.249	10.622	10.622	-7.6338	64.7411	-122.6937	74.3393
B3A4C7(1-5)	2	4.249	11.684	10.622	7.6691	-1.9439	-10.2308	8.2459
B3A4C7(1-5)	3	4.249	11.684	10.622	46.1792	-135.7261	151.3696	-57.1895
B3A4C8(1-4)	1	4.680	11.115	11.700	-1.1027	32.3612	-72.7874	50.6645
B3A4C8(1-4)	2	4.680	12.870	11.700	10.2515	-14.0149	5.8844	1.6396
B3A4C8(1-4)	3	4.680	12.870	11.700	31.2833	-81.2871	84.6531	-30.3665
B3A4C9(1-3)	1	5.236	11.780	13.089	-13.4934	96.3564	-181.5386	111.3934
B3A4C9(1-3)	2	5.236	14.398	13.089	10.5258	-17.5585	11.6771	-0.9260
B3A4C9(1-3)	3	5.236	14.398	13.089	25.6265	-63.5145	63.7953	-21.9627
B3A4C10(1-2)	1	5.683	12.786	14.207	-9.1856	69.1320	-127.0558	76.6129
B3A4C10(1-2)	2	5.683	14.917	14.207	8.5308	-11.5777	5.6826	1.0046
B3A4C10(1-2)	3	5.683	15.627	14.207	20.7101	-48.3879	47.3221	-15.8260
B3A4C11(1-1)	1	6.232	13.242	15.579	4.3006	-2.2312	-5.8996	9.9794
B3A4C11(1-1)	2	6.232	16.358	15.579	8.1823	-12.1574	7.5880	-0.0679
B3A4C11(1-1)	3	6.232	16.358	15.579	18.5297	-45.4114	47.8051	-17.3465
B1A4C5(1-3)	1	3.093	7.217	6.874	-1.5502	31.1085	-61.1410	37.5861
B1A4C5(1-3)	2	3.093	7.561	6.874	10.0541	-18.0399	17.0758	-5.4326
B1A4C5(1-3)	3	3.093	7.561	6.874	56.4203	-184.8620	217.6114	-84.6639
B1A4C6(1-2)	1	3.182	8.352	7.955	-3.0005	37.9119	-71.8949	43.2587

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
B1A4C6(1-2)	2	3.182	8.750	7.955	10.7278	-20.9133	20.0570	-6.2503
B1A4C6(1-2)	3	3.182	8.750	7.955	30.9366	-90.9472	104.6702	-40.4831
B1A4C7(1-1)	1	3.731	8.861	9.328	1.7750	10.2372	-22.7649	15.5184
B1A4C7(1-1)	2	3.731	10.260	9.328	9.6392	-17.8861	16.2780	-4.5839
B1A4C7(1-1)	3	3.731	10.260	9.328	21.3363	-54.5815	58.5775	-21.5513
B1A4C6(1-4)	1	3.182	8.352	7.955	0.5779	22.8207	-51.5608	35.1635
B1A4C6(1-4)	2	3.182	8.750	7.955	11.4041	-20.6892	17.7022	-4.4958
B1A4C6(1-4)	3	3.182	8.750	7.955	52.0644	-164.5688	190.7415	-73.5752
B1A4C7(1-3)	1	3.731	8.861	9.328	2.0671	14.9818	-41.3702	32.7330
B1A4C7(1-3)	2	3.731	10.260	9.328	9.9658	-16.9576	13.5479	-2.7071
B1A4C7(1-3)	3	3.731	10.260	9.328	29.9285	-85.1475	95.1771	-35.7555
B1A4C8(1-2)	1	4.185	9.939	10.462	-4.9255	47.9331	-90.4527	55.9262
B1A4C8(1-2)	2	4.185	10.986	10.462	8.5461	-12.5373	9.1315	-1.3519
B1A4C8(1-2)	3	4.185	11.509	10.462	21.2483	-53.3770	56.7765	-20.5974
B1A4C9(1-1)	1	4.727	10.635	11.816	4.7948	-4.5300	0.4713	4.1712
B1A4C9(1-1)	2	4.727	12.998	11.816	7.6753	-10.0475	5.7878	0.1455
B1A4C9(1-1)	3	4.727	12.998	11.816	21.4224	-56.7964	62.3847	-23.2235
B1A4C7(1-5)	1	3.731	9.328	9.328	-2.3883	39.1250	-81.7106	54.4918
B1A4C7(1-5)	2	3.731	10.260	9.328	10.6417	-15.9143	9.3869	0.1720
B1A4C7(1-5)	3	3.731	10.260	9.328	42.0709	-124.4259	139.3475	-52.2868
B1A4C8(1-4)	1	4.185	9.939	10.462	-0.7255	29.4446	-65.7193	47.4915
B1A4C8(1-4)	2	4.185	11.509	10.462	8.5361	-9.4150	2.3698	2.7500
B1A4C8(1-4)	3	4.185	11.509	10.462	25.2251	-64.7146	68.5846	-24.6815
B1A4C9(1-3)	1	4.727	10.635	11.816	2.6722	11.9035	-37.9002	33.7600
B1A4C9(1-3)	2	4.727	12.998	11.816	9.1498	-13.6361	8.7035	-0.1393
B1A4C9(1-3)	3	4.727	12.998	11.816	22.6821	-58.8438	63.4398	-23.1159
B1A4C10(1-2)	1	5.193	11.684	12.982	-2.4173	37.1889	-76.9739	52.0734
B1A4C10(1-2)	2	5.193	13.631	12.982	8.9517	-14.7414	12.2469	-2.4887
B1A4C10(1-2)	3	5.193	14.280	12.982	18.2503	-43.5309	45.5099	-16.1754
B1A4C11(1-1)	1	5.729	12.174	14.322	3.3268	2.5149	-10.4674	10.1764
B1A4C11(1-1)	2	5.729	15.038	14.322	6.6313	-5.5798	-0.4837	3.1751
B1A4C11(1-1)	3	5.729	15.754	14.322	15.5356	-34.1644	33.7852	-11.3796
B0A4C5(1-3)	1	3.191	6.701	6.382	-5.1231	44.0005	-80.8977	47.7736
B0A4C5(1-3)	2	3.191	7.020	6.382	10.4810	-22.7844	21.4147	-6.1271
B0A4C5(1-3)	3	3.191	7.020	6.382	39.6811	-118.3020	129.3601	-47.3034
B0A4C6(1-2)	1	3.417	7.974	7.594	-14.0935	81.1913	-130.5583	69.2818
B0A4C6(1-2)	2	3.417	8.353	7.594	7.6570	-14.0183	12.0133	-2.7088
B0A4C6(1-2)	3	3.417	8.353	7.594	26.5484	-77.9431	87.0650	-32.4821
B0A4C7(1-1)	1	3.561	8.903	8.903	1.1509	9.4134	-21.6858	15.3445
B0A4C7(1-1)	2	3.561	9.793	8.903	7.2330	-13.7004	12.4586	-3.3139
B0A4C7(1-1)	3	3.561	9.793	8.903	22.0144	-64.6641	73.3932	-27.8231
B0A4C6(1-4)	1	3.417	7.594	7.594	3.0027	5.7625	-23.8784	21.7441

Continued on Next Page

Table F.17 – Continued from Previous Page

Formula*	Sh†	r_{min}‡	r_{max}‡	r_{rel}‡	c_0§	c_1§	c_2§	c_3§
B0A4C6(1-4)	2	3.417	8.353	7.594	9.3839	-16.9524	12.6938	-1.6868
B0A4C6(1-4)	3	3.417	8.353	7.594	41.9802	-125.9784	138.9203	-51.0729
B0A4C7(1-3)	1	3.561	8.012	8.903	-2.1322	30.8798	-65.5104	45.4035
B0A4C7(1-3)	2	3.561	9.793	8.903	8.3618	-15.4301	12.8275	-2.4354
B0A4C7(1-3)	3	3.561	9.793	8.903	30.4011	-90.5100	101.6165	-38.0101
B0A4C8(1-2)	1	4.052	9.624	10.131	-10.0033	67.7167	-119.6488	69.9941
B0A4C8(1-2)	2	4.052	11.144	10.131	6.1670	-7.5125	3.1882	1.3296
B0A4C8(1-2)	3	4.052	11.144	10.131	21.6243	-61.0093	67.9582	-25.3115
B0A4C9(1-1)	1	4.572	11.430	11.430	-3.7476	33.8420	-60.9313	36.1986
B0A4C9(1-1)	2	4.572	12.574	11.430	5.2919	-5.0651	0.4929	2.1187
B0A4C9(1-1)	3	4.572	12.574	11.430	16.2353	-42.1263	45.2888	-16.4396
B0A4C7(1-5)	1	3.561	8.903	8.903	-8.4406	65.7565	-126.0061	79.3491
B0A4C7(1-5)	2	3.561	9.793	8.903	9.2768	-15.2457	9.7055	0.1156
B0A4C7(1-5)	3	3.561	9.793	8.903	42.3149	-129.1790	145.6360	-54.6271
B0A4C8(1-4)	1	4.052	9.118	10.131	3.4739	1.7329	-17.4987	21.8082
B0A4C8(1-4)	2	4.052	10.637	10.131	7.9797	-13.1320	9.7736	-0.8435
B0A4C8(1-4)	3	4.052	11.144	10.131	26.8910	-75.8740	83.3307	-30.5265
B0A4C9(1-3)	1	4.572	9.716	11.430	-13.3886	90.7042	-169.5779	105.9729
B0A4C9(1-3)	2	4.572	12.002	11.430	5.8642	-5.1812	-0.5760	3.4104
B0A4C9(1-3)	3	4.572	12.574	11.430	19.9270	-52.3106	55.7215	-19.8653
B0A4C10(1-2)	1	5.067	11.401	12.668	-6.8381	56.4697	-109.8387	70.6639
B0A4C10(1-2)	2	5.067	13.301	12.668	7.1882	-11.3145	7.9201	-0.4015
B0A4C10(1-2)	3	5.067	13.934	12.668	17.6021	-45.9518	49.6880	-17.9996
B0A4C11(1-1)	1	5.585	13.264	13.962	-1.9807	26.9471	-52.8076	33.9787
B0A4C11(1-1)	2	5.585	14.660	13.962	5.9679	-6.8257	1.6494	2.2732
B0A4C11(1-1)	3	5.585	15.358	13.962	14.5554	-35.3580	36.5406	-12.6927

*Formula representation, as described in the Text.

†Shape: 1 = Bilayer, 2 = Cylinder, 3 = Sphere.

‡Radii in units of Å, as described in the Text.

§Packing polynomial coefficients, as described in the Text. Units of k_BT.

Bibliography

[1] Seminario, J. M. *International Journal of Quantum Chemistry* **1996,** *60,* 1271–1277.

[2] Halgren, T. A. *J. Am. Chem. Soc.* **1990,** *112,* 4710–4723.

[3] Haynes, W. M. *CRC Handbook of Chemistry and Physics;* CRC Press: Boca Raton, 92nd ed.; 2011.

[4] Flory, P. J.; Orwoll, R. A.; Vrij, A. *J. Am. Chem. Soc.* **1964,** *86,* 3507–3514.

[5] Klopman, G.; Wang, S.; Balthasar, D. M. *J. Chem. Inf. Comput. Sci.* **1992,** *32,* 474–482.

[6] Kühne, R.; Ebert, R.-U.; Kleint, F.; Schmidt, G.; Schüürmann, G. *Chemosphere* **1995,** *30,* 2061–2077.

[7] Vargaftik, N. B.; Volkov, B. N.; Voljak, L. D. *Journal of Physical and Chemical Reference Data* **1983,** *12,* 817–820.

[8] Rossini, F. D.; Pitzer, K. S.; Arnett, R. L.; Braun, R. M.; Pimentel, G. C. *Selected Values of Physical and Thermodynamic Properties of Hydrocarbons and Related Compounds (API Research Project 44);* Carnegie Press: Pittsburgh, 1953.

[9] Zeppieri, S.; Rodríguez, J.; López de Ramos, A. L. *J. Chem. Eng. Data* **2001,** *46,* 1086–1088.

[10] Aveyard, R.; Haydon, D. A. *Transactions of the Faraday Society* **1965,** *61,* 2255–2261.

[11] Needham, D. E.; Wei, I. C.; Seybold, P. G. *J. Am. Chem. Soc.* **1988,** *110,* 4186–4194.

[12] Rosen, M. J. *Surfactants and Interfacial Phenomena;* John Wiley & Sons: Hoboken, 3rd ed.; 2004.

[13] Holmberg, K.; Jönsson, B.; Kronberg, B.; Lindman, B. *Surfactants and Polymers in Aqueous Solution;* John Wiley & Sons: Chichester, West Sussex, 2nd ed.; 2002.

[14] Lu, T.; Han, F.; Mao, G.; Lin, G.; Huang, J.; Huang, X.; Wang, Y.; Fu, H. *Langmuir* **2007,** *23,* 2932–2936.

[15] Moroi, Y.; Murata, Y.; Fukuda, Y.; Kido, Y.; Seto, W.; Tanaka, M. *The Journal of Physical Chemistry* **1992,** *96,* 8610–8613.

[16] Rosen, M.; Tracy, D. *Journal of Surfactants and Detergents* **1998,** *1,* 547–554.

[17] Zana, R. *Journal of Colloid and Interface Science* **2002,** *248,* 203–220.

[18] Porter, M. R. *Handbook of Surfactants;* Chapman & Hall: New York, 2nd ed.; 1994.

[19] Riess, G. *Progress in Polymer Science* **2003,** *28,* 1107–1170.

[20] Stache, H. W. *Anionic Surfactants;* Marcel Dekker: New York, 1996.

[21] Cross, J.; Singer, E. J. *Cationic Surfactants;* Marcel Dekker: New York, 1994.

[22] Lomax, E. G. *Amphoteric Surfactants;* Marcel Dekker: New York, 1996.

[23] Oldenhove de Guertechin, L. Surfactants: Classification. In *Handbook of Detergents. Part A. Properties*; CRC Press: Boca Raton, 1999.

[24] Maeda, H. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **1996,** *109,* 263–271.

[25] Maeda, H.; Kakehashi, R. *Advances in Colloid and Interface Science* **2000,** *88,* 275–293.

[26] Varadaraj, R.; Bock, J.; Zushma, S.; Brons, N. *Langmuir* **1992,** *8,* 14–17.

[27] Doe, P.; El-Emary, M.; Wade, W.; Schechter, R. *Journal of the American Oil Chemists' Society* **1977,** *54,* 570–577.

[28] Tan, X.-L.; Zhang, L.; Zhao, S.; Li, W.; Ye, J.-P.; Yu, J.-Y.; An, J.-Y. *Langmuir* **2004,** *20,* 7010–7014.

[29] Kissa, E. *Fluorinated Surfactants and Repellents;* Marcel Dekker: New York, 2nd ed.; 2001.

[30] Hill, R. M. Siloxane surfactants. In *Silicone Surfactants*; Marcel Dekker: New York, 1999.

[31] Guo, W.; Li, Z.; Fung, B. M.; O'Rear, E. A.; Harwell, J. H. *The Journal of Physical Chemistry* **1992,** *96,* 6738–6742.

[32] Chandler, D. *Nature* **2005,** *437,* 640–647.

[33] Tanford, C. *Science* **1978,** *200,* 1012–1018.

[34] Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes;* John Wiley & Sons: New York, 1973.

[35] Israelachvili, J. N. *Intermolecular and Surface Forces;* Academic Press: San Diego, 3rd ed.; 2011.

[36] Myers, D. *Surfactant Science and Technology;* John Wiley & Sons: Hoboken, 2006.

[37] Noïk, C.; Bavière, M.; Defives, D. *Journal of Colloid and Interface Science* **1987,** *115,* 36–45.

[38] Stellner, K.; Scamehorn, J. *Journal of the American Oil Chemists' Society* **1986,** *63,* 566–574.

[39] Shinoda, K.; Hato, M.; Hayashi, T. *The Journal of Physical Chemistry* **1972,** *76,* 909–914.

[40] Gu, T.; Sjöblom, J. *Colloids and Surfaces* **1992,** *64,* 39–46.

[41] Lockwood, N. A.; de Pablo, J. J.; Abbott, N. L. *Langmuir* **2005,** *21,* 6805–6814.

[42] Prosser, A. J.; Franses, E. I. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2001,** *178,* 1–40.

[43] Manglik, R. M.; Wasekar, V. M.; Zhang, J. *Experimental Thermal and Fluid Science* **2001,** *25,* 55–64.

[44] Smit, B.; Schlijper, A. G.; Rupert, L. A. M.; Van Os, N. M. *The Journal of Physical Chemistry* **1990,** *94,* 6933–6935.

[45] Shah, D. O. *Surface Phenomena in Enhanced Oil Recovery;* Plenum Press: New York, 1981.

[46] Marmur, A.; Lelah, M. D. *Chemical Engineering Communications* **1981,** *13,* 133–143.

[47] Daniel, R. C.; Berg, J. C. *Advances in Colloid and Interface Science* **2006,** *123-126,* 439–469.

[48] Imae, T.; Abe, A.; Ikeda, S. *J. Phys. Chem.* **1988,** *92,* 1548–1553.

[49] Christian, S. D.; Scamehorn, J. F. *Solubilization in Surfactant Aggregates;* CRC Press: Boca Raton, 1995.

[50] Tadros, T. F. *Applied Surfactants;* Wiley VCH: Weinheim, 2005.

[51] Cahn, A.; Lai, K.-Y. Liquid detergents: An overview. In *Liquid Detergents,* 2nd ed.; CRC Press: Boca Raton, FL, 2006.

[52] Esumi, K. Dispersion of particles by surfactants. In *Structure-Performance Relationships in Surfactants*, 2nd ed.; Marcel Dekker: New York, 2003.

[53] Bos, M. A.; van Vliet, T. *Advances in Colloid and Interface Science* **2001,** *91,* 437–471.

[54] Holmberg, K.; Jönsson, B.; Kronberg, B.; Lindman, B. *Surfactants and Polymers in Aqueous Solution;* John Wiley & Sons: Chichester, West Sussex, UK, 2003.

[55] Schulte, H.-G.; Höfer, R. Uses of anti-foaming agents in paints and surface coatings. In *Surfactants in Polymers, Coatings, Inks and Adhesives*; CRC Press: Boca Raton, 2003.

[56] Ludlum, D. B. *J. Phys. Chem.* **1956,** *60,* 1240–1244.

[57] van Os, N.; Daane, G.; Haandrikman, G. *Journal of Colloid and Interface Science* **1991,** *141,* 199–217.

[58] Binana-Limbelé, W.; van Os, N.; Rupert, L.; Zana, R. *Journal of Colloid and Interface Science* **1991,** *141,* 157–167.

[59] Granet, R.; Piekarski, S. *Colloids and Surfaces* **1988,** *33,* 321–336.

[60] Nusselder, J. J. H.; Engberts, J. B. F. N. *J. Org. Chem.* **1991,** *56,* 5522–5527.

[61] Nusselder, J. J. H.; Engberts, J. B. F. N. *Langmuir* **1991,** *7,* 2089–2096.

[62] Varadaraj, R.; Bock, J.; Valint, P.; Zushma, S.; Thomas, R. *J. Phys. Chem.* **1991,** *95,* 1671–1676.

[63] Varadaraj, R.; Bock, J.; Zushma, S.; Brons, N.; Colletti, T. *Journal of Colloid and Interface Science* **1991,** *147,* 387–395.

[64] Abe, M.; Scamehorn, J. F. *Mixed Surfactant Systems;* Marcel Dekker: New York, 2nd ed.; 2007.

[65] Esumi, K.; Ueno, M. *Structure-Performance Relationships in Surfactants;* Marcel Dekker: New York, 1997.

[66] Clint, J. H. *J. Chem. Soc., Faraday Trans. 1* **1975,** *71,* 1327–1334.

[67] Holland, P. M.; Rubingh, D. N. *The Journal of Physical Chemistry* **1983,** *87,* 1984–1990.

[68] Huibers, P. D. T.; Lobanov, V. S.; Katritzky, A. R.; Shah, D. O.; Karelson, M. *Langmuir* **1996,** *12,* 1462–1470.

[69] Katritzky, A. R.; Pacureanu, L.; Dobchev, D.; Karelson, M. *Journal of Chemical Information and Modeling* **2007,** *47,* 782–793.

[70] Roberts, D. W. *Langmuir* **2001,** *18,* 345–352.

[71] Al-Anber, Z. A.; Bonet Avalos, J.; Mackie, A. D. *The Journal of Chemical Physics* **2005,** *122,* 104910–104910–8.

[72] Israelachvili, J. N.; Mitchell, D. J.; Ninham, B. W. *J. Chem. Soc., Faraday Trans. 2* **1976,** *72,* 1525–1568.

[73] Bohmer, M. R.; Koopal, L. K. *Langmuir* **1990,** *6,* 1478–1484.

[74] Hurter, P. N.; Scheutjens, J. M. H. M.; Hatton, T. A. *Macromolecules* **1993,** *26,* 5592–5601.

[75] Jódar-Reyes, A. B.; Ortega-Vinuesa, J. L.; Martín-Rodríguez, A.; Leermakers, F. A. M. *Langmuir* **2002,** *18,* 8706–8713.

[76] Scheutjens, J. M. H. M.; Fleer, G. J. *The Journal of Physical Chemistry* **1979,** *83,* 1619–1635.

[77] Flory, P. J. *Principles of Polymer Chemistry;* Cornell University Press: Ithaca, 1953.

[78] Hill, T. L. *Thermodynamics of Small Systems: Part I;* W. A. Benjamin: New York, 1963.

[79] Hill, T. L. *Thermodynamics of Small Systems: Part II;* W. A. Benjamin: New York, 1963.

[80] Nagarajan, R.; Ruckenstein, E. *Journal of Colloid and Interface Science* **1979,** *71,* 580–604.

[81] Puvvada, S.; Blankschtein, D. *The Journal of Chemical Physics* **1990,** *92,* 3710–3724.

[82] Ben-Shaul, A.; Szleifer, I.; Gelbart, W. M. *Proceedings of the National Academy of Sciences* **1984,** *81,* 4601–4605.

[83] Puvvada, S.; Blankschtein, D. *J. Phys. Chem.* **1992,** *96,* 5567–5579.

[84] Puvvada, S.; Blankschtein, D. *J. Phys. Chem.* **1992,** *96,* 5579–5592.

[85] Shiloach, A.; Blankschtein, D. *Langmuir* **1998,** *14,* 7166–7182.

[86] Yuet, P. K. *Langmuir* **1996,** *12,* 3802–3818.

[87] Goldsipe, A.; Blankschtein, D. *Langmuir* **2007,** *23,* 5953–5962.

[88] Srinivasan, V.; Blankschtein, D. *Langmuir* **2003,** *19,* 9932–9945.

[89] Srinivasan, V.; Blankschtein, D. *Langmuir* **2003,** *19,* 9946–9961.

[90] Goldsipe, A.; Blankschtein, D. *Langmuir* **2005,** *21,* 9850–9865.

[91] Shelley, J. C.; Shelley, M. Y. *Current Opinion in Colloid & Interface Science* **2000,** *5,* 101–110.

[92] Brodskaya, E. *Colloid Journal* **2012,** *74,* 154–171.

[93] Polat, B. E.; Lin, S.; Mendenhall, J. D.; VanVeller, B.; Langer, R.; Blankschtein, D. *The Journal of Physical Chemistry B* **2011,** *115,* 1394–1402.

[94] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983,** *4,* 187–217.

[95] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983,** *79,* 926–935.

[96] Bogusz, S.; Venable, R. M.; Pastor, R. W. *The Journal of Physical Chemistry B* **2000,** *104,* 5462–5470.

[97] Bogusz, S.; Venable, R. M.; Pastor, R. W. *The Journal of Physical Chemistry B* **2001,** *105,* 8312–8321.

[98] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995,** *117,* 5179–5197.

[99] Pathiaseril, A.; Woods, R. J. *Journal of the American Chemical Society* **1999,** *122,* 331–338.

[100] Chong, T. T.; Hashim, R.; Bryce, R. A. *The Journal of Physical Chemistry B* **2006,** *110,* 4978–4984.

[101] Abel, S.; Dupradeau, F.-Y.; Raman, E. P.; MacKerell, A. D.; Marchi, M. *The Journal of Physical Chemistry B* **2011,** *115,* 487–499.

[102] Schweighofer, K. J.; Essmann, U.; Berkowitz, M. *The Journal of Physical Chemistry B* **1997,** *101,* 3793–3799.

[103] Bruce, C. D.; Berkowitz, M. L.; Perera, L.; Forbes, M. D. E. *J. Phys. Chem. B* **2002,** *106,* 3788–3793.

[104] Bruce, C. D.; Senapati, S.; Berkowitz, M. L.; Perera, L.; Forbes, M. D. E. *J. Phys. Chem. B* **2002,** *106,* 10902–10907.

[105] Nelson, P. H.; Rutledge, G. C.; Hatton, T. A. *The Journal of Chemical Physics* **1997,** *107,* 10777–10781.

[106] Jusufi, A.; Hynninen, A.-P.; Panagiotopoulos, A. Z. *J. Phys. Chem. B* **2008,** *112,* 13783–13792.

[107] Lazaridis, T.; Mallik, B.; Chen, Y. *J. Phys. Chem. B* **2005,** *109,* 15098–15106.

[108] Stephenson, B. C.; Goldsipe, A.; Beers, K. J.; Blankschtein, D. *The Journal of Physical Chemistry B* **2007,** *111,* 1025–1044.

[109] Stephenson, B. C.; Goldsipe, A.; Beers, K. J.; Blankschtein, D. *The Journal of Physical Chemistry B* **2007,** *111,* 1045–1062.

[110] Stephenson, B. C.; Beers, K. J.; Blankschtein, D. *J. Phys. Chem. B* **2007,** *111,* 1063–1075.

[111] Nagarajan, R.; Ruckenstein, E. *Langmuir* **1991,** *7,* 2934–2969.

[112] Shiloach, A.; Blankschtein, D. *Langmuir* **1998,** *14,* 1618–1636.

[113] Tester, J. W.; Modell, M. *Thermodynamics and Its Applications;* Prentice Hall PTR: Upper Saddle River, New Jersey, 3rd ed.; 1997.

[114] Smith, J. M.; Van Ness, H. C.; Abbott, M. M. *Introduction to Chemical Engineering Thermodynamics;* McGraw-Hill: New York, 6th ed.; 2001.

[115] Shiloach, A.; Blankschtein, D. *Langmuir* **1998,** *14,* 4105–4114.

[116] Goldsipe, A.; Blankschtein, D. *Langmuir* **2006,** *22,* 3547–3559.

[117] Iyer, J.; Blankschtein, D. *J. Phys. Chem. B* **2012,** *116,* 6443–6454.

[118] Smit, B.; Esselink, K.; Hilbers, P. A. J.; Van Os, N. M.; Rupert, L. A. M.; Szleifer, I. *Langmuir* **1993,** *9,* 9–11.

[119] Srinivasan, V. *Theoretical modeling of micellization and solubilization in ionic surfactant systems,* Thesis, Massachusetts Institute of Technology, 2003.

[120] Nagarajan, R.; Chaiko, M. A.; Ruckenstein, E. *J. Phys. Chem.* **1984,** *88,* 2916–2922.

[121] Laschewsky, A. *Current Opinion in Colloid & Interface Science* **2003,** *8,* 274–281.

[122] Lutz, J.-F.; Laschewsky, A. *Macromolecular Chemistry and Physics* **2005,** *206,* 813–817.

[123] Nagarajan, R. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **1993,** *71,* 39–64.

[124] Blankschtein, D.; Thurston, G. M.; Benedek, G. B. *The Journal of Chemical Physics* **1986,** *85,* 7268–7288.

[125] Kjellander, R. *Journal of the Chemical Society, Faraday Transactions 2* **1982,** *78,* 2025.

[126] Carale, T. R.; Pham, Q. T.; Blankschtein, D. *Langmuir* **1994,** *10,* 109–121.

[127] Napper, D. H. *Polymeric Stabilization of Colloidal Dispersions;* Academic Press: Orlando, 1983.

[128] Zoeller, N.; Lue, L.; Blankschtein, D. *Langmuir* **1997,** *13,* 5258–5275.

[129] McMillan, W. G.; Mayer, J. E. *The Journal of Chemical Physics* **1945,** *13,* 276–305.

[130] Corkill, J. M.; Goodman, J. F.; Walker, T.; Wyer, J. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **1969,** *312,* 243–255.

[131] Mukerjee, P. *J. Phys. Chem.* **1972,** *76,* 565–570.

[132] Reif, I.; Mulqueen, M.; Blankschtein, D. *Langmuir* **2001,** *17,* 5801–5812.

[133] Kamrath, R. F.; Franses, E. I. *J. Phys. Chem.* **1984,** *88,* 1642–1648.

[134] Williams, R. J.; Phillips, J. N.; Mysels, K. J. *Transactions of the Faraday Society* **1955,** *51,* 728–737.

[135] Corrin, M. *Journal of Colloid Science* **1948,** *3,* 333–338.

[136] Hall, D. G. *Transactions of the Faraday Society* **1970,** *66,* 1351–1358.

[137] Bockris, J. O.; Reddy, A. K. N. *Modern Electrochemistry;* volume 1 Plenum Press: New York, 1970.

[138] Onsager, L. *J. Am. Chem. Soc.* **1936,** *58,* 1486–1493.

[139] Hildebrand, J. H.; Prausnitz, J. M.; Scott, R. L. *Regular and Related Solutions;* Van Nostrand Reinhold: New York, 1970.

[140] Eriksson, F.; Eriksson, J. C.; Stenius, P. *Colloids and Surfaces* **1981,** *3,* 339–356.

[141] Maibaum, L.; Dinner, A. R.; Chandler, D. *J. Phys. Chem. B* **2004,** *108,* 6778–6781.

[142] Defay, R.; Prigogine, I.; Bellemans, A. *Surface Tension and Adsorption;* Longmans, Green & Co: London, 1966.

[143] Morgan, J. L. R.; Griggs, M. A. *J. Am. Chem. Soc.* **1917,** *39,* 2261–2275.

[144] Tolman, R. C. *The Journal of Chemical Physics* **1949,** *17,* 333–337.

[145] Koenig, F. O. *The Journal of Chemical Physics* **1950,** *18,* 449–459.

[146] Buff, F. P. *The Journal of Chemical Physics* **1951,** *19,* 1591–1594.

[147] Naor, A.; Puvvada, S.; Blankschtein, D. *The Journal of Physical Chemistry* **1992,** *96,* 7830–7832.

[148] Goldsipe, A.; Blankschtein, D. *Langmuir* **2007,** *23,* 5942–5952.

[149] Szleifer, I.; Carignano, M. A. *Macromolecular Rapid Communications* **2000,** *21,* 423–448.

[150] Helfand, E.; Reiss, H.; Frisch, H. L.; Lebowitz, J. L. *The Journal of Chemical Physics* **1960,** *33,* 1379–1385.

[151] Lebowitz, J. L.; Helfand, E.; Praestgaard, E. *The Journal of Chemical Physics* **1965,** *43,* 774–779.

[152] Fraser, D. P.; Zuckermann, M. J.; Mouritsen, O. G. *Physical Review A* **1991,** *43,* 6642–6656.

[153] Bockris, J. O.; Reddy, A. K. N. *Modern Electrochemistry;* volume 2 Plenum Press: New York, 1970.

[154] Shiloach, A.; Blankschtein, D. *Langmuir* **1997,** *13,* 3968–3981.

[155] Ohshima, H.; Healy, T. W.; White, L. R. *Journal of Colloid and Interface Science* **1982,** *90,* 17–26.

[156] Moré, J. J.; Garbow, B. S.; Hillstrom, K. E. "User guide for MINPACK-1", Technical Report ANL-80-74, Argonne Nat. Lab., Argonne, IL, 1980.

[157] Kaufman, E.; Leeming, D.; Taylor, G. *Numerical Algorithms* **1995,** *9,* 25–37.

[158] Magid, L. J. *J. Phys. Chem. B* **1998,** *102,* 4064–4074.

[159] Boden, N.; Corne, S.; Holmes, M.; Jackson, P.; Parker, D.; Jolley, K. *Journal de Physique* **1986,** *47,* 2135–2144.

[160] Fromherz, P.; Röcker, C.; Rüppel, D. *Faraday Discussions of the Chemical Society* **1986,** *81,* 39.

[161] Srinivasan, V.; Blankschtein, D. *Langmuir* **2005,** *21,* 1647–1660.

[162] Al-Anber, Z. A.; Bonet i Avalos, J.; Floriano, M. A.; Mackie, A. D. *The Journal of Chemical Physics* **2003,** *118,* 3816–3826.

[163] Missel, P. J.; Mazer, N. A.; Benedek, G. B.; Young, C. Y.; Carey, M. C. *J. Phys. Chem.* **1980,** *84,* 1044–1057.

[164] Bernheim-Groswasser, A.; Zana, R.; Talmon, Y. *J. Phys. Chem. B* **2000,** *104,* 4005–4009.

[165] Kozlov, M. M.; Lichtenberg, D.; Andelman, D. *J. Phys. Chem. B* **1997,** *101,* 6600–6606.

[166] Szleifer, I.; Ben-Shaul, A.; Gelbart, W. M. *The Journal of Chemical Physics* **1985,** *83,* 3612–3620.

[167] Kreyszig, E. *Advanced Engineering Mathematics;* John Wiley & Sons: New York, 10th ed.; 2011.

[168] May, S.; Bohbot, Y.; Ben-Shaul, A. *J. Phys. Chem. B* **1997,** *101,* 8648–8657.

[169] Imae, T.; Kamiya, R.; Ikeda, S. *Journal of Colloid and Interface Science* **1985,** *108,* 215–225.

[170] Magid, L. J.; Li, Z.; Butler, P. D. *Langmuir* **2000,** *16,* 10028–10036.

[171] van Os, N. M.; Haak, J. R.; Rupert, L. A. M. *Physico-Chemical Properties of Selected Anionic, Cationic and Nonionic Surfactants;* Elsevier: Amsterdam, 1993.

[172] Briganti, G.; Puvvada, S.; Blankschtein, D. *J. Phys. Chem.* **1991,** *95,* 8989–8995.

[173] Imanishi, K.; Einaga, Y. *J. Phys. Chem. B* **2005,** *109,* 7574–7581.

[174] Balmbra, R. R.; Clunie, J. S.; Corkill, J. M.; Goodman, J. F. *Transactions of the Faraday Society* **1962,** *58,* 1661.

[175] Ravey, J.-C. *Journal of Colloid and Interface Science* **1983,** *94,* 289–291.

[176] He, X.; Shinoda, W.; DeVane, R.; Anderson, K. L.; Klein, M. L. *Chemical Physics Letters* **2010,** *487,* 71–76.

[177] Palazzesi, F.; Calvaresi, M.; Zerbetto, F. *Soft Matter* **2011,** *7,* 9148.

[178] Yoshii, N.; Okazaki, S. *Chemical Physics Letters* **2006,** *425,* 58–61.

[179] Yan, H.; Yuan, S.-L.; Xu, G.-Y.; Liu, C.-B. *Langmuir* **2010,** *26,* 10448–10459.

[180] Rakitin, A. R.; Pack, G. R. *J. Phys. Chem. B* **2004,** *108,* 2712–2716.

[181] Tobias, D. J.; Klein, M. L. *J. Phys. Chem.* **1996,** *100,* 6637–6648.

[182] Sterpone, F.; Briganti, G.; Pierleoni, C. *Langmuir* **2001,** *17,* 5103–5110.

[183] Marcus, Y. *Biophysical Chemistry* **1994,** *51,* 111–127.

[184] Lee, B.; Richards, F. *Journal of Molecular Biology* **1971,** *55,* 379–IN4.

[185] Reynolds, J. A.; Gilbert, D. B.; Tanford, C. *Proceedings of the National Academy of Sciences* **1974,** *71,* 2925–2927.

[186] Menger, F. M.; Doll, D. W. *J. Am. Chem. Soc.* **1984,** *106,* 1109–1113.

[187] Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988,** *110,* 1657–1666.

[188] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996,** *118,* 11225–11236.

[189] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008,** *4,* 435–447.

[190] van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual version 4.5.4;* Uppsala University: Sweden, 2010.

[191] Canongia Lopes, J. N.; Deschamps, J.; Pádua, A. A. H. *The Journal of Physical Chemistry B* **2004,** *108,* 2038–2047.

[192] Canongia Lopes, J. N.; Pádua, A. A. H. *The Journal of Physical Chemistry B* **2004,** *108,* 16893–16898.

[193] Canongia Lopes, J. N.; Pádua, A. A. H. *J. Phys. Chem. B* **2006,** *110,* 19586–19592.

[194] Canongia Lopes, J. N.; Pádua, A. A. H.; Shimizu, K. *The Journal of Physical Chemistry B* **2008,** *112,* 5039–5046.

[195] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *Journal of Computational Chemistry* **1997,** *18,* 1463–1472.

[196] Darden, T.; York, D.; Pedersen, L. *The Journal of Chemical Physics* **1993,** *98,* 10089–10092.

[197] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *The Journal of Chemical Physics* **1995,** *103,* 8577–8593.

[198] Deserno, M.; Holm, C. *The Journal of Chemical Physics* **1998,** *109,* 7678–7693.

[199] Lennard-Jones, J. E. *Proceedings of the Physical Society* **1931,** *43,* 461–482.

[200] Bussi, G.; Donadio, D.; Parrinello, M. *The Journal of Chemical Physics* **2007,** *126,* 014101–014101–7.

[201] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *The Journal of Chemical Physics* **1984,** *81,* 3684–3690.

[202] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids;* Clarendon Press: Oxford, 1987.

[203] Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications;* volume 1 of *Computational Science: From Theory to Applications* Academic Press: San Diego, 2nd ed.; 2002.

[204] Lin, S. "A computer simulation and molecular-thermodynamic framework to model the micellization of ionic branched surfactants in aqueous solution", Master's thesis, Massachusetts Institute of Technology, 2008.

[205] Rezus, Y. L. A.; Bakker, H. J. *Physical Review Letters* **2007,** *99,* 148301.

[206] Frank, H. S.; Evans, M. W. *The Journal of Chemical Physics* **1945,** *13,* 507–532.

[207] Fidler, J.; Rodger, P. M. *J. Phys. Chem. B* **1999,** *103,* 7695–7703.

[208] Southall, N. T.; Dill, K. A.; Haymet, A. D. J. *J. Phys. Chem. B* **2001,** *106,* 521–533.

[209] Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M. *Journal of Computational Chemistry* **2004,** *16,* 273–284.

[210] Shrake, A.; Rupley, J. *Journal of Molecular Biology* **1973,** *79,* 351–371.

[211] Wishnia, A. *J. Phys. Chem.* **1963,** *67,* 2079–2082.

[212] Shinitzky, M.; Dianoux, A. C.; Gitler, C.; Weber, G. *Biochemistry* **1971,** *10,* 2106–2113.

[213] Flory, P. J. *Statistical Mechanics of Chain Molecules;* Oxford University Press: New York, 1989.

[214] Szleifer, I.; Ben-Shaul, A.; Gelbart, W. M. *The Journal of Chemical Physics* **1986,** *85,* 5345–5358.

[215] Mattice, W. L.; Santiago, G. *Macromolecules* **1980,** *13,* 1560–1567.

[216] Abe, A.; Jernigan, R. L.; Flory, P. J. *J. Am. Chem. Soc.* **1966,** *88,* 631–639.

[217] Mattice, W. L. *Macromolecules* **1975,** *8,* 644–651.

[218] Bates, T. W.; Stockmayer, W. H. *Macromolecules* **1968,** *1,* 12–17.

[219] Ben-Shaul, A.; Gelbart, W. M. *Annual Review of Physical Chemistry* **1985,** *36,* 179–211.

[220] Jaynes, E. T. *Physical Review* **1957,** *106,* 620–630.

[221] Jaynes, E. T. *Physical Review* **1957,** *108,* 171–190.

[222] Shannon, C. E. *SIGMOBILE Mob. Comput. Commun. Rev.* **2001,** *5,* 3–55.

[223] Tribus, M. *Journal of Applied Mechanics* **1961,** *28,* 1–8.

[224] Tribus, M. *Thermostatics and Thermodynamics;* D. Van Nostrand Co.: Princeton, NJ, 1961.

[225] Schrödinger, E. *Statistical Thermodynamics;* Cambridge University Press: Cambridge, 1952.

[226] Szleifer, I.; Ben-Shaul, A.; Gelbart, W. M. *The Journal of Chemical Physics* **1987,** *86,* 7094–7109.

[227] Mattice, W. L.; Suter, U. W. *Conformational Theory of Large Molecules: The Rotational Isomeric State Model in Macromolecular Systems;* John Wiley & Sons: New York, 1994.

[228] Baker, J.; Chan, F. *Journal of Computational Chemistry* **1996,** *17,* 888–904.

[229] Rehahn, M.; Mattice, W. L.; Suter, U. W. *Rotational Isomeric State Models in Macromolecular Systems;* volume 131/132 of *Advances in Polymer Science* Springer-Verlag: Berlin Heidelberg, 1997.

[230] Patterson, G. D.; Flory, P. J. *J. Chem. Soc., Faraday Trans. 2* **1972,** *68,* 1098–1110.

[231] Hoeve, C. A. J. *The Journal of Chemical Physics* **1961,** *35,* 1266–1267.

[232] Ben-Shaul, A.; Szleifer, I.; Gelbart, W. M. *The Journal of Chemical Physics* **1985,** *83,* 3597–3611.

[233] Breneman, C. M.; Wiberg, K. B. *Journal of Computational Chemistry* **1990,** *11,* 361–373.

[234] Frisch, M. J. *et al.* "Gaussian 03, Revision C.02", Gaussian, Inc., Wallingford, CT, 2004.

[235] Anderson, P. M.; Wilson, M. R. *Molecular Physics* **2005,** *103,* 89–97.

[236] Berr, S.; Jones, R. R. M.; Johnson, J. S. *The Journal of Physical Chemistry* **1992,** *96,* 5611–5614.

[237] Dupuy, C.; Auvray, X.; Petipas, C.; Rico-Lattes, I.; Lattes, A. *Langmuir* **1997,** *13,* 3965–3967.

[238] Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *Journal of Computational Chemistry* **1997,** *18,* 1955–1970.

[239] Abraham, M. H. *Journal of the Chemical Society, Faraday Transactions 1* **1984,** *80,* 153.

[240] Moroi, Y.; Sugii, R.; Akine, C.; Matuura, R. *Journal of Colloid and Interface Science* **1985,** *108,* 180–188.

[241] Moroi, Y.; Nishikido, N.; Uehara, H.; Matuura, R. *Journal of Colloid and Interface Science* **1975,** *50,* 254–264.

[242] Nilsson, P. G.; Lindman, B. *J. Phys. Chem.* **1983,** *87,* 4756–4761.

[243] Klevens, H. B. *The Journal of physical and colloid chemistry* **1948,** *52,* 130–148.

[244] Aniansson, E. A. G.; Wall, S. N.; Almgren, M.; Hoffmann, H.; Kielmann, I.; Ulbricht, W.; Zana, R.; Lang, J.; Tondre, C. *J. Phys. Chem.* **1976,** *80,* 905–922.

[245] Gershman, J. W. *J. Phys. Chem.* **1957,** *61,* 581–584.

[246] Barry, B. W.; Wilson, R. *Colloid & Polymer Science* **1978,** *256,* 251–260.

[247] Shinoda, K.; Hirai, T. *J. Phys. Chem.* **1977,** *81,* 1842–1845.

[248] Legrand, D.; Gaines Jr., G. *Journal of Colloid and Interface Science* **1973,** *42,* 181–184.

[249] Gómez-Díaz, D.; Navaza, J. M.; Sanjurjo, B. n. *J. Chem. Eng. Data* **2007,** *52,* 889–891.

[250] Evans, D. F.; Allen, M.; Ninham, B. W.; Fouda, A. *Journal of Solution Chemistry* **1984,** *13,* 87–101.

[251] Škerjanc, J.; Kogej, K.; Cerar, J. *Langmuir* **1999,** *15,* 5023–5028.

[252] Qu, G.; Cheng, J.; Wei, J.; Yu, T.; Ding, W.; Luan, H. *Journal of Surfactants and Detergents* **2011,** *14,* 31–35.

[253] Carignano, M. A.; Szleifer, I. *The Journal of Chemical Physics* **1995,** *102,* 8662–8669.

[254] Frindi, M.; Michels, B.; Zana, R. *J. Phys. Chem.* **1992,** *96,* 6095–6102.

[255] Fournial, A.-G.; Zhu, Y.; Molinier, V.; Vermeersch, G.; Aubry, J.-M.; Azaroual, N. *Langmuir* **2007,** *23,* 11443–11450.

[256] Mahajan, R. K.; Nandni, D. *Ind. Eng. Chem. Res.* **2012,** *51,* 3338–3349.

[257] Hu, C.; Li, R.; Yang, H.; Wang, J. *Journal of Colloid and Interface Science* **2011,** *356,* 605–613.

[258] Nilsson, F.; Söderman, O.; Johansson, I. *Journal of Colloid and Interface Science* **1998,** *203,* 131–139.

[259] Matsumura, S.; Imai, K.; Yoshikawa, S.; Kawada, K.; Uchibor, T. *Journal of the American Oil Chemists' Society* **1990,** *67,* 996–1001.

[260] He, L.; Garamus, V. M.; Funari, S. S.; Malfois, M.; Willumeit, R.; Niemeyer, B. *J. Phys. Chem. B* **2002,** *106,* 7596–7604.

[261] Liljekvist, P.; Kronberg, B. *Journal of Colloid and Interface Science* **2000,** *222,* 159–164.

[262] Walter, A.; Suchy, S. E.; Vinson, P. K. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1990,** *1029,* 67–74.

[263] Frindi, M.; Michels, B.; Zana, R. *J. Phys. Chem.* **1992,** *96,* 8137–8141.

[264] Chai, J.; Yang, X.; Gao, Y.; Chen, J.; Wang, Z. *Colloid Journal* **2008,** *70,* 221–226.

[265] Rappoport, Z. *CRC Handbook of Tables for Organic Compound Identification;* CRC Press: Boca Raton, 3rd ed.; 1967.

[266] Bai, G.; Wang, J.; Yan, H.; Li, Z.; Thomas, R. K. *J. Phys. Chem. B* **2001,** *105,* 3105–3108.

[267] Newcomb, L. F.; Gellman, S. H. *J. Am. Chem. Soc.* **1994,** *116,* 4993–4994.

[268] Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *Journal of the Chemical Society, Perkin Transactions 2* **2001,** 651–669.

[269] He, X.; Guvench, O.; MacKerell, A. D.; Klein, M. L. *J. Phys. Chem. B* **2010,** *114,* 9787–9794.

[270] van Os, N.; Daane, G.; Bolsman, T. *Journal of Colloid and Interface Science* **1988,** *123,* 267–274.

[271] Dill, K. A.; Flory, P. J. *Proceedings of the National Academy of Sciences* **1981,** *78,* 676–680.

[272] Besserman, M. *A study of the aqueous solutions of the positional isomers of sodium dodecanesulfonate as colloidal electrolytes,* Thesis, University of Washington, 1953.

[273] Gunnarsson, G.; Joensson, B.; Wennerstroem, H. *J. Phys. Chem.* **1980,** *84,* 3114–3121.

[274] Schick, M. J.; Fowkes, F. M. *J. Phys. Chem.* **1957,** *61,* 1062–1068.

[275] Van Os, N.; Daane, G.; Bolsman, T. *Journal of Colloid and Interface Science* **1987,** *115,* 402–409.

[276] Bauernschmitt, D.; Hoffmann, H.; Platz, G. *Berichte der Bunsengesellschaft für physikalische Chemie* **1981,** *85,* 203–210.

[277] Jaynes, E. T. *Proceedings of the IEEE* **1982,** *70,* 939– 952.

[278] Polak, J.; Lu, B. C.-Y. *Canadian Journal of Chemistry* **1973,** *51,* 4018–4023.

[279] Sutton, C.; Calder, J. A. *J. Chem. Eng. Data* **1975,** *20,* 320–322.

[280] Mooij, G. C. A. M.; Frenkel, D. *Journal of Physics: Condensed Matter* **1994,** *6,* 3879–3888.

[281] Bennett, C. H. *Journal of Computational Physics* **1976,** *22,* 245–268.

[282] Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *The Journal of Chemical Physics* **1982,** *76,* 637–649.

[283] Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *The Journal of Chemical Physics* **1998,** *108,* 1964–1977.

[284] Siu, S. W. I.; Pluhackova, K.; Böckmann, R. A. *J. Chem. Theory Comput.* **2012,** .

[285] Huggins, M. L. *The Journal of Chemical Physics* **1941,** *9,* 440–440.

[286] de Gennes, P. G. *The Journal of Chemical Physics* **1980,** *72,* 4756–4763.

[287] Pincus, P. *The Journal of Chemical Physics* **1981,** *75,* 1996–2000.

[288] Forrest, B.; Toral, R. *Journal of Statistical Physics* **1994,** *77,* 473–489.

[289] McBain, J. W.; Lissant, K. J. *The Journal of Physical Chemistry* **1951,** *55,* 655–662.

[290] Tokuoka, Y.; Uchiyama, H.; Abe, M.; Christian, S. D. *Langmuir* **1995,** *11,* 725–729.

[291] Stephenson, B. C.; Rangel-Yagui, C. O.; Junior, A. P.; Tavares, L. C.; Beers, K.; Blankschtein, D. *Langmuir* **2006,** *22,* 1514–1525.

[292] Donahue, D. J.; Bartell, F. E. *The Journal of Physical Chemistry* **1952,** *56,* 480–484.

[293] Mulligan, C.; Yong, R.; Gibbs, B. *Engineering Geology* **2001,** *60,* 371–380.

[294] Karney, C. F. F. *Journal of Molecular Graphics and Modelling* **2007,** *25,* 595–604.

[295] Goldstein, H.; Poole, C.; Safko, J. *Classical Mechanics;* Addison Wesley: New York, 3rd ed.; 2001.

[296] Fattal, D. R.; Ben-Shaul, A. *Biophysical Journal* **1994,** *67,* 985–995.

[297] Vince, J. *Quaternions for Computer Graphics;* Springer: London, 2011.

[298] Parsons, J.; Holmes, J. B.; Rojas, J. M.; Tsai, J.; Strauss, C. E. M. *Journal of Computational Chemistry* **2005,** *26,* 1063–1068.

[299] Metropolis, N.; Ulam, S. *Journal of the American Statistical Association* **1949,** *44,* 335–341.

[300] Newman, M. E. J.; Barkema, G. T. *Monte Carlo Methods in Statistical Physics;* Clarendon Press: Oxford, 1999.

[301] Gruen, D. W. R. *J. Phys. Chem.* **1985,** *89,* 146–153.

[302] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *The Journal of Chemical Physics* **1953,** *21,* 1087–1092.

[303] Rubinstein, R. Y. *Simulation and the Monte Carlo Method;* John Wiley & Sons: New York, 1981.

[304] Landau, D. P.; Binder, K. *A Guide to Monte-Carlo Simulations in Statistical Physics;* Cambridge University Press: New York, 3rd ed.; 2009.

[305] Sokal, A. D. Monte Carlo Methods for the Self-Avoiding Walk. In *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*; Binder, K., Ed.; Oxford University Press: New York, 1995.

[306] Diudea, M. V.; Gutman, I.; Lorentz, J. *Molecular Topology;* Nova Science Publishers: Hauppauge, New York, 2001.

[307] Diestel, R. *Graph Theory;* Springer-Verlag: New York, 2000.

[308] Verlet, L. *Physical Review* **1967,** *159,* 98–103.

[309] Haile, J. M. *Molecular Dynamics Simulation: Elementary Methods;* John Wiley & Sons: New York, 1992.

[310] Chandrasekhar, S. *Reviews of Modern Physics* **1943,** *15,* 1–89.

[311] Marsaglia, G.; Bray, T. A. *SIAM Review* **1964,** *6,* 260–264.

[312] Lemieux, C. *Monte Carlo and Quasi-Monte Carlo Sampling;* Springer Series in Statistics Springer: New York, 2009.

[313] Matsumoto, M.; Nishimura, T. *ACM Trans. Model. Comput. Simul.* **1998,** *8,* 3–30.

[314] Jäckel, P. *Monte Carlo Methods in Finance;* Wiley Finance Series John Wiley & Sons: Chichester, West Sussex, 2002.

[315] Martelli, A. *Python in a Nutshell: A Desktop Quick Reference;* O'Reilly: Sebastopol, 2nd ed.; 2006.

[316] Ryckaert, J.-P.; Bellemans, A. *Faraday Discuss. Chem. Soc.* **1978,** *66,* 95–106.

[317] Aris, R. *Vectors, Tensors, and the Basic Equations of Fluid Mechanics;* Dover Publications: New York, 1989.

[318] Shimizu, K.; Almantariotis, D.; Gomes, M. F. C.; Pádua, A. A. H.; Canongia Lopes, J. N. *The Journal of Physical Chemistry B* **2010,** *114,* 3592–3600.

[319] McDonald, N. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998,** *102,* 8049–8059.

[320] Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999,** *121,* 4827–4836.

[321] Price, M. L. P.; Ostrovsky, D.; Jorgensen, W. L. *Journal of Computational Chemistry* **2001,** *22,* 1340–1352.

[322] Watkins, E. K.; Jorgensen, W. L. *The Journal of Physical Chemistry A* **2001,** *105,* 4118–4125.

[323] Foresman, J. B.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods;* Gaussian, Inc.: Pittsburgh, 2nd ed.; 1996.

[324] Mø ller, C.; Plesset, M. S. *Physical Review* **1934,** *46,* 618–622.

[325] Dunning, T. H. *The Journal of Chemical Physics* **1989,** *90,* 1007.

[326] Dunning, T. H.; Peterson, K. A.; Wilson, A. K. *The Journal of Chemical Physics* **2001,** *114,* 9244–9253.

[327] Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997,** *119,* 5908–5920.

[328] Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. *J. Phys. Chem. A* **1999,** *103,* 1913–1928.

[329] Feller, D. *Journal of Computational Chemistry* **1996,** *17,* 1571–1586.

[330] Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. *Journal of Chemical Information and Modeling* **2007,** *47,* 1045–1052.

[331] Halgren, T. A. *Journal of Molecular Structure: THEOCHEM* **1988,** *163,* 431–446.

[332] Nilsson, K.; Lecerof, D.; Sigfridsson, E.; Ryde, U. *Acta Crystallographica Section D Biological Crystallography* **2003,** *59,* 274–289.

[333] Lüder, K.; Kjellander, R. *J. Phys. Chem. B* **2006,** *110,* 15514–15524.

[334] Westbrook, J. D.; Levy, R. M.; Krogh-Jespersen, K. *Journal of Computational Chemistry* **1992,** *13,* 979–989.

[335] Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *Journal of Computational Chemistry* **1986,** *7,* 230–252.

[336] Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors;* volume 11 of *Methods and Principles in Medicinal Chemistry* Wiley-VCH: New York, 2000.

[337] Tropsha, A.; Golbraikh, A. Predictive Quantitative Structure-Activity Relationships Modeling. In *Handbook of Chemoinformatics Algorithms*; Faulon, J.-L.; Bender, A., Eds.; CRC Press: Boca Raton, 2010.

[338] Gavezzotti, A. *J. Am. Chem. Soc.* **1983,** *105,* 5220–5225.

[339] Bondi, A. *J. Phys. Chem.* **1964,** *68,* 441–451.

[340] Tanaka, M.; Girard, G.; Davis, R.; Peuto, A.; Bignell, N. *Metrologia* **2001,** *38,* 301–309.

[341] Kier, L. B.; Murray, W. J.; Randić, M.; Hall, L. H. *Journal of Pharmaceutical Sciences* **2006,** *65,* 1226–1230.

[342] Spencer, C. F.; Danner, R. P. *Journal of Chemical & Engineering Data* **1972,** *17,* 236–241.

[343] Yalkowsky, S. H.; Valvani, S. C. *Journal of Pharmaceutical Sciences* **1980,** *69,* 912–922.

[344] Isnard, P.; Lambert, S. *Chemosphere* **1989,** *18,* 1837–1853.

[345] Jain, N.; Yalkowsky, S. H. *Journal of Pharmaceutical Sciences* **2001,** *90,* 234–252.

[346] Abraham, M. H.; Le, J. *Journal of Pharmaceutical Sciences* **1999,** *88,* 868–880.

[347] Hall, L. H.; Kier, L. B.; Murray, W. J. *Journal of Pharmaceutical Sciences* **2006,** *64,* 1974–1977.

[348] Nirmalakhandan, N. N.; Speece, R. E. *Environ. Sci. Technol.* **1988,** *22,* 328–338.

[349] Sutton, C.; Calder, J. A. *Environ. Sci. Technol.* **1974,** *8,* 654–657.

[350] Sherblom, P. M.; Gschwend, P. M.; Eganhouse, R. P. *J. Chem. Eng. Data* **1992,** *37,* 394–399.

[351] Mackay, D.; Shiu, W. Y. *Journal of Physical and Chemical Reference Data* **1981,** *10,* 1175–1199.

[352] McAuliffe, C. *J. Phys. Chem.* **1966,** *70,* 1267–1275.

[353] Price, L. C. *AAPG Bulletin* **1976,** *60,* 213–244.

[354] McAuliffe, C. *Science* **1969,** *163,* 478–479.

[355] Hiemenz, P. C.; Rajagopalan, R. *Principles of Colloid and Surface Chemistry;* Marcel Dekker: New York, 1997.

[356] Fowkes, F. M. *Ind. Eng. Chem.* **1964,** *56,* 40–52.

[357] Girifalco, L. A.; Good, R. J. *J. Phys. Chem.* **1957,** *61,* 904–909.

[358] Demond, A. H.; Lindner, A. S. *Environ. Sci. Technol.* **1993,** *27,* 2318–2331.