Adsorption and Partitioning of Solutes in Chromatographic and Surfactant Systems

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Rebecca K. Lindsey

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

J. Ilja Siepmann

May, 2016

Acknowledgements

First and foremost I would like to thank my adviser, J. Ilja Siepmann for his help in my becoming part of the University of Minnesota graduate student body – I doubt that I would have been here without his support. I also want to thank him for his willingness to answer my often numerous questions, while simultaneously providing an environment conducive to growth as an independent researcher.

The Siepmann group members, past and present, have also been invaluable during my time here. The unique strengths and perspectives of each member has often led to solutions that I alone would never have come up with. Furthermore, the students I've mentored, Sophia Charan, Laura Moore, and Michael Neaton, have kept me inspired, through their infectious curiosity and endless drive. Additionally, I would like extend my thanks to Sumanth Jamadagni, Peter Koenig, and David Eike of The Procter and Gamble Company, and to Mark Schure of Kroungold Analytical for providing support and guidance during my time working with them; I have enjoyed every moment of it.

I want to thank my closest friends who have been like my family. I want to thank those with me at the beginning of my journey, and those there at the end; they were and continue to be major factor in my success and sanity. Finally, I want to thank my parents, sister, and brother for their limitless and unquestioning support during my time here – knowing that I could talk to them at a moment's notice has meant more than they will ever know.

Finally, funding from the National Science Foundation, the Proctor and Gamble Company, and Dow are gratefully acknowledged, as are the Minnesota Supercomputing Institute, Environmental Molecular Sciences Laboratory Molecular Science Computing Facility, and Extreme Science and Engineering Discovery Environment for computing resources.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Interfacial solute/solvent interactions play a crucial role in numerous chemical processes, spanning analytical separations [1], carbon capture [2], enhanced oil recovery [3], templating [4], battery systems [5], personal care products and beyond [6, 7]. The work discussed herein leverages molecular simulation and other computational methods to offer fundamental perspectives on such phenomena and for the development of tools to aid in their study and use, with specific applications towards solute retention in chromatographic systems and uptake in nanostructured soft materials.

1.1 Simulating Chromatographic Systems

1.1.1 Background in Chromatography

Column chromatographic methods, which were developed in the early 1900's, represent the most prevalent analytical separation technique in existence [8, 9, 10]. Though several different column chromatographic techniques exist, the present work will focus on liquid chromatography (LC), in particular. The two schematics for LC, provided in Figure 1.1, demonstrate the general premise of the approach. Essentially, a liquid "mobile phase" is pumped through a series of pipes which connect to a chromatographic column. This column, in turn, is connected to a detector of some type (*i.e.* UV, mass spectrometer, *etc.*) whose purpose is to distinguish the chemicals which exit the column. The mixture of solutes ("solute matrix") that one wishes to separate is injected into the pipeline upstream

of the chromatographic column, and is carried through the column by coersion of the mobile phase. The solute matrix arrives to the column in a single pulse, however, separation of solutes begins upon interaction with the contents of the column. Though flow is present in the system, interactions between the solutes and column contents is typically understood to be an equilibrium processes [8, 10], with solutes transferring between the mobile phase and the stationary contents several times during their passage through the column. It is the difference between the relative favorabilities of solute/stationary phase and solute/mobile phase interactions that eventually result in a separation; those solutes which interact more favorably with the mobile phase than the column will exit first, and those which interact more favorably with the stationary phase will exit last, thus a separation is achieved based on the difference in residence time of solutes within the column, resulting in a measured chromatogram which may look like the example given in Figure 1.1.

The assumption that retention is an equilibrium process yields a particularly interesting result: though retention times are the direct observable of an experimental chromatographic separation, those times can be directly related to transfer free energies for solutes between the stationary and mobile phases, referred to as the retention free energy, ΔG_{ret}:

$$\Delta G_{\text{ret}} = -RT \ln K_{\text{eq}}, \tag{1.1}$$

where R, T, and K_{eq} are the gas constant, temperature, and partition coefficient from the mobile to stationary phases, respectively. The partition coefficient, which is typically understood as a ratio of solute concentrations between two phases, can also be described in terms of a capacity factor, k', and a phase ratio, ϕ. These values are defined as:

$$k' = \frac{t_{\text{r,S}} - t_{\text{d}}}{t_{\text{d}}} = \frac{[\text{S}]_{\text{SP}} V_{\text{SP}}}{[\text{S}]_{\text{MP}} V_{\text{MP}}} = K_{\text{eq}} \phi^{-1}, \tag{1.2}$$

where t is a time, V is a volume, $[\text{S}]_x$ is the concentration of solute S in phase x, and the subscripts r and d refer to the retention time of solute S and the time it takes for an un-retained species to pass through the column ("dead time"), respectively. In practice, ϕ is an exceedingly difficult quantity to define rigorously, however, this issue can be overcome by considering the K_{eq} *ratios* between two species, rather than an absolute K_{eq}. A deeper discussion on ϕ and methods for circumventing the related difficulties is discussed in Chapter 4, Section 4.2.2.

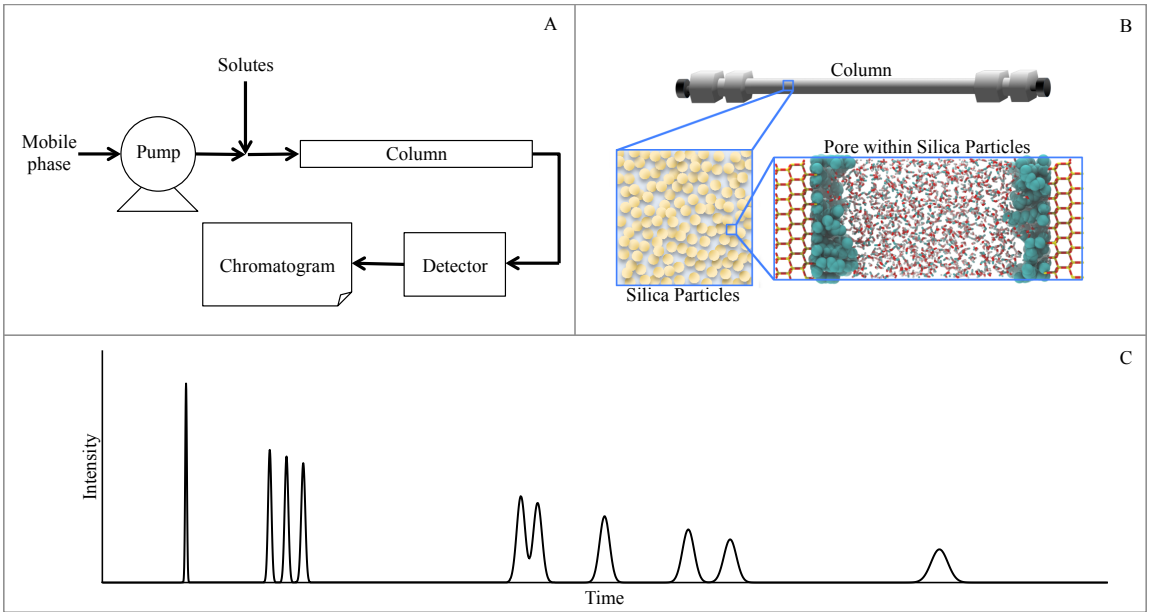

Figure 1.1: A.: Schematic for a liquid chromatography setup. Arrows indicate pipes which connect components. B.: View inside of a liquid chromatography column. Yellow circles in the bottom left box represent silica particles. The system contained in the lower right box corresponds to a chromatographic slit pore. Yellow, red, cyan, and white atoms correspond to silicon, oxygen carbon, and hydrogen respectively. Grafted chains have their carbon atoms drawn with a large radius for emphasis. C.: Example of a chromatogram. Different peaks correspond to different chemicals, and time increases when moving from left to right.

As was mentioned above, LC is a ubiquitous analytical tool. Despite its widespread use, however, there is still much to be learned about how analyte molecules are retained, this being the topic of many debates within the chromatography community [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]. To a large extent, this is due to the plethora of existing LC methods, and the continual development of new approaches, each having a unique retention mechanism. The typical approach when attempting to elucidate retention mechanisms involves the use of van't Hoff analysis and linear solvation free energy relationships [27, 28, 29, 30, 31] (LSER's). Analysis through the former approach yields thermodynamic insights into the retention process, and involves chromatograhic measurements over a range of temperatures and plotting of $\ln k'$ vs T^{-1}.

The following equation is then fit to the data:

$$\ln k' = \frac{-\Delta H}{RT} + \frac{\Delta S}{R} + \ln \phi, \tag{1.3}$$

where $-\Delta H$ and ΔS are the enthalpy and entropy for the solute, respectively, and a good linear fit corresponds to ϕ, $-\Delta H$, and ΔS which are nearly constant over small temperature ranges. Analysis through LSERs is performed by use of the following equation [27, 28, 29, 30, 31]:

$$\mathrm{SP} = c + vV + sS + aA + bB + eE, \tag{1.4}$$

where V, S, A, B, and E are the solute volume, solute dipolarity, H-bond donating ability, H-bond accepting ability, and polarizabilty, respectively, c is a fitting constant, and SP is typically chosen to be $\log k'$. Lowercase letters v, s, a, b, and e correspond to the difference in the ability of the the stationary and mobile phases to interact with the solute by the properties specified for the corresponding capital letter parameter. If one wishes to characterize a column, a solute set whose LSER parameters have been pre-defined can be run through a column, and the resulting $\log k'$ values can be used to fit LSER parameters for the column.

Switching focus to the variety of available LC methods, the evolution of chromatography began with the so-called "normal" phase (NP) configuration [32, 33]. In NPLC, a polar inorganic oxide or cellulose serves as the stationary phase (SP), while a non-polar organic solvent is used as the mobile phase. As one might imagine, solutes of higher polarity are retained for longer in such columns, with less polar solutes exiting the column ("eluting") first. These columns are successful in retaining and separating polar analytes, however, strong analyte-SP interactions often result in peak tailing, which limits the utility of the NPLC approach.

Reversed-phase (RP) columns were then introduced as an alternative to NP columns, having hydrophobic chains grafted to the NP surface [34, 35]. The mobile phase used in RPLC is typically an aqueous organic mixture, and, true to its name, results in a solute retention order which the reverse of what is seen in NPLC. These columns are ideal for biological and pharmaceutical applications, however, highly polar compounds tend to be poorly retained due to their affinity for the aqueous mobile phase.

Since the development of RPLC, numerous other liquid chromatographic methods have been developed, including hydrophilic interaction liquid chromatography (HILIC) [36],

which is particularly well suited for separation of polar compounds, chiral chromatography [37], in which ligands are designed to separate solutes on the basis of their chirality, affinity chromatography [38], which, for example, can be used to separate biological materials having specific binding sites, *etc.*, ion-exchange chromatography [39], which separates materials on the basis of charge properties, and size exclusion chromatography (SEC) [40], which, in principle, separates solutes on the basis of hydrodynamic volume. Furthermore, there have been advances in the materials used in chromatographic systems (for example monolithic materials [41], sub micron particles [42, 43, 44], and zirconia-based particles [45, 46], to name a few) as well as how columns are leveraged (*i.e.* multidimensional methods [47]).

Historically, these complex systems have primarily been studied through the use of experimental methods, however, with recent advances in computing power and methodology, simulation based research has begun to emerge in the field of chromatography. The main advantages stemming from use of simulation include the inherent ability of some molecular simulation methods to produce thermodynamic data simultaneously with microscopic structural data, and the ability to study systems at conditions that are not accessible in experiments. Focusing for a moment on the latter, experimental researchers can be limited by factors such as phase and solute decomposition or unreasonably slow/fast retention times, however, these issues do not exist in simulation, which can be useful when attempting to develop fundamental retention trends. Furthermore, simulation methods can be described as an exceedingly powerful microscope, as coordinates of every atom in the system are directly observable. In the next section, we will briefly examine findings from recent simulation-based RPLC studies.

1.1.2 Review of Previous Simulation Studies

Due to limitations in computing power at the time, early simulations of chromatographic systems focused on solvent-free stationary phases [49, 50, 51, 52]. Solvent was later introduced into the chromatographic simulations [53, 54, 55] and was found to significantly

The contents of this section draw significantly from a review article for which primary contributions were made by the present author [48]

alter chain structure [56], and exhibit a tendency to adsorb to the chain interface. The solvent was also shown to be capable of forming filaments that bridge between the solvent hydrogen bonded to the substrate surface, and solvent in the bulk mobile phase [57, 58, 59]. While inclusion of mobile phase leads to markedly different structure compared to chains in vacuum, alteration of the mobile phase composition has been shown to have a much more subtle effect. Chains were shown exhibit a broad distribution of orientations at all compositions [60], rather than exhibiting a strong preference for a single orientation, and become slightly less ordered in solvent containing higher concentrations of organic modifier [60]. Furthermore, the organic component tends to enrich at the chain/solvent interface [57, 61]. These studies also addressed a long-standing debate within the community, which was whether the significant changes in retention observed when moving from highly organic to highly aqueous mobile phases was due to phase-collapse [62, 63], showing that chains did not exhibit drastic changes, and suggesting that instead, significant losses in solvent wetting at the chain/solvent interface is the likely culprit.

Studies of the influence of chain length on pore structure [64] have shown that longer chain phases exhibit a slightly higher degree of disorder, in terms of both fraction of gauche defects and order parameter (a measure of chain orientation with respect to the substrate normal) in the central region of the chain. To some extent, the apparent increase in disorder is due to the ability of longer chains to fold back on themselves. Furthermore, dewetting at the interface was found to be slightly reduced by longer chains. The influence of chain grafting density has also been studied, and the following has been found: increased grafting density leads to decreases in solvent accessible surface area [65, 66], and begins to exclude solvent from the interior of the stationary phase chains [54, 55]. Furthermore, higher grafting densities were shown to lead to a slightly stronger enrichment of organic at the interface and increased chain homogeneity across the silica surface [65]. Uniformity of grafting sites has been shown to have a significant effect on chain structure at lower grafting density [67], where clustering of chains can lead to chain structure comparable to that at higher grafting densities [68].

The results discussed to this point have all corresponded to a planar slit-type pore, which can be considered to relate to large diameter pore having low curvature, however, there have been simulation studies which examine higher curvature, cylindrical pores [69, 70, 71]. From

an experimental standpoint, pores exhibiting higher curvature require higher operating pressures in order to achieve solvent penetration [63]. Simulation results have found that increased curvature lead to chains that are far more extended, and exhibit significantly lower density near the surface of the silica [71]. When compared to a planar slit pore, residual silanol sites were found to be present in greater quantities for equivalent chain grafting density, as was solvent in the vicinity of the substrate. The increase in penetrating solvent was not found to be due to the increase in H-bonding sites (in curved surfaces, silanol sites tend to be geminal, and thus H-bond to one another), but instead due to the increase in the amount of space between chains, close to the substrate interface [71].

The influence of temperature, pressure, and pH have also been examined through simulation. Pressure is found to have a very small effect below 1000 atm, but has been shown to induce a slight ordering and compression of the grafted chains, an increase in enrichment of organic component at chain/solvent interface, and slightly more H-bonds to surface silanols [67, 71]. Temperature is generally found to increase disorder in the chains [53, 72, 73]. Altering the operation pH of an RPLC column can lead to ionization of silanol groups, and the presence of free cations near the substrate surface. These cations are shown to sit at ionized silanol sites, and greatly enhance water penetration, while excluding nearly all organic solvent, fairly independent of solvent composition, though water does remain depleted at chain/solvent interface [71]. Furthermore, ions are found to induce stronger ordering at base of SP chains, leading to a stronger alignment with surface normal, fewer gauche defects, and a larger end-to-end distance.

Embedded polar (EP) groups are often introduced into RPLC grafted chains to improve separation of polar compounds. Simulation studies of systems containing ether- and amide-type embedded polar groups [71, 74] demonstrated that they can alter pore structure by increasing solvent penetration and thereby, solvent H-bonding to substrate silanols. These moieties were also shown to increase overall solvent wetting at the chain/solvent interface, while reducing enhancement of the organic component. Furthermore, EP groups were shown to encourage extension and ordering of chains, and to lead to overall smaller tilt angles with respect to the interface normal.

Solutes (in particular, methane) were first introduced into these chromatographic simulations in the late 1990's [75, 76]. Beyond simply providing predictions of retention free

energies, these simulations also produced solute location-dependent free energy profiles, yielding valuable insights into the preferred location of solutes within the retentive pore. These simulations studies represent some of the first steps in relating system microstructure (rather than using thermodynamic extrapolations) to provide insights into the retention mechanism in RPLC. Since that time, numerous studies have been conducted, predicting incremental retention free energies for significantly more complex solute series, in excellent agreement with experimental results [64, 66, 71, 74, 77, 78]. These findings have helped to determine the driving forces in RPLC, and to understand how those driving forces change as a function of, for example, mobile phase composition. In this document, chromatographic simulation work will focus on developing similar insights for pressurized hot water RPLC systems, and towards development of a computational tool to aid in column selection for multi-column separations.

1.2 Simulating Surfactant Systems

1.2.1 Surfactants as Solute Carriers

Surfactants represent an incredibly versatile class of chemical compounds, which are found in everything from injections for enhanced oil recovery [3] to our own lungs [79]. The utility of these amphiphilar molecules lies in their ability to decrease surface tension between otherwise highly insoluble species, and to create microemulsions having diverse but tunable micellar structures. Indeed, a surfactant solution can form a variety of thermodynamically stable microemulsions under the right conditions; for example, by heating a 50/50 weight percent mixture of water and pentaethylene glycol mono-*n*-dodecyl ether (C12E5), one can observe hexagonally packed rod micelles, spherical micelles, and a (layered) lamellar phase, to name a few [80]. Furthermore, surfactants can form monolayers exhibiting their own rich phase diagrams, which have utility in sensor applications [81] and as simplified models for biological bilayers [82].

Beyond versatility in micelle structure, surfactants also exhibit significant diversity in chemical composition. For example, the C12E5 surfactant mentioned above belongs to the "polyoxyethylene" class of non-ionic surfactants, and by simply changing the number of alkyl ("C") or oxyethylene ("E") groups, one can manipulate accessible micellar structures,

by way of changing preferred interfacial curvature [80, 83]. Ionic surfactants exhibit similar tunability, where decreasing salt concentration has an effect comparable to increasing the number of "E" groups in a polyoxyethylene surfactant.

Research on surfactant solutions has been conducted in several areas. In fields geared towards industrial applications, for example, one is often concerned with the uptake of pigment or fragrant compounds for personal care formulations [6, 84], where product texture and viscosity are also of importance [85]. As such, there have been tremendous advances in phase diagram mapping [80, 83, 86, 87], development of mathematical and thermodynamic models relating surfactant and microemulsion structure [80, 88] and characterization of corresponding mechanical properties [85]. These phase diagrams, which typically focus on surfactant/water or surfactant/oil/water systems at varied temperature, have become measureable largely due to advances in spectroscopic [86, 87, 89] and NMR methods [90, 91], but are still quite difficult to determine.

In pharmaceutical research, on the other hand, a common theme is solubilization of drug molecules in spherical-type micelles [92, 93, 94, 95, 96]. Towards this front, there has been extensive work done with the aim of understanding the thermodynamics of solute partitioning into micellar structures. This type of study typically involves measurement of micelle/water partition coefficients for solutes, which are then used to develop linear solvation free energy relationships (LSERs), as discussed in section 1.1. In this application LSER parameters correspond to either the solute, or the micellar phase, and while this approach is very useful for identifying trends in solubility [94, 96, 97], which can help guide formulation scientists, LSERs are based on the assumption of infinite diliuton, which is not always valid for practical applications [98]. Nevertheless, many suggestions have been put forth regarding the mechanism of solute uptake in surfactant materials, as will be discussed in Chapters 6 and 7. Both of the aforementioned surfactant research areas would benefit from a fundamental and systematic understanding of how solute uptake and micellar structure are interrelated. Given the immensity of physiochemical space, however, this would be a very expensive experimental undertaking, whether in terms of chemical resources, human cost, or generated waste. Computer simulations, on the other hand, provide a feasible means to begin to tackle this problem.

1.2.2 Towards Large Scale Studies

Over the past 100 years, computers have transitioned from cumbersome collections of vacuum tubes only accessible to a select few, to extremely compact and affordable devices. Furthermore, trends in computing resources have faithfully followed Moore's law [99], being orders of magnitude more powerful than they once were. With the advances and increased affordability of these machines, molecular simulation tools have concomitantly become more sophisticated, and are being used to solve some of the world's most challenging problems in science, technology, enginering, and mathmatics [100, 101, 102]. In the realm of surfactant research, advances in simulation tools have manifested through introduction of enhanced sampling techniques [103, 104, 105, 106], new algorithms [107, 108], and coarse-grained force fields specifically designed for the study of surfactant systems and their micellar phase transitions [109, 110].

Coupled with the recent interest in large scale screening of materials (*i.e.* The Nanoporous Materials Genome Center [111] or The Materials Project [112]), one might envision a similar initiative for soft materials, such as those formed by surfactants; the aim of such a study would entail construction of a database of phase diagrams, solubility data, mechanical properties, *etc.*. Though such an initiative for soft materials is still far off (the fluid nature of soft materials requires more advanced computational resources for large scale studies), steps towards this eventual goal can be made by attempting to refine the related simulation details (*viz.* the protocols). For example, with the multitude of force field resolutions available, it would be very beneficial to know where the trade off between accuracy and efficiency lies. Along similar lines, one needs to be aware of how simulation approach impacts the computed result.

Given the relevance of surfactant materials, several simulation studies have already been conducted, which are quite useful for guiding future work. Given large length and time scales over which structured surfactant materials exist (*i.e.* micrometer and microsecond, respectively [113, 114]), much of the early computational studies focused on use of coarse grained "toy" models, designed not for quantitave agreement with any specific system, but rather for elucidation of trends in structure and phase transitions. These early simulations often also employed the use of lattice models, making possible the study systems which would otherwise be considered prohibitively large, at the time. For example,

in 1992, Larson used lattice Monte Carlo simulations to study the phase transitions in model surfactant/oil/water systems, and was able to observe spherical micelles, hexagonally packed rods, lamellae, and bicontinious phases [115]. Several years later, Floriano *et al.* used a coarse grained force fields in conjunction with lattice Monte Carlo in an open (grand canonical) ensemble, where surfactants were allowed to enter and exit the simulation box [116]. Using this approach, Floriano *et al.* were able to predict the critical micelle concentrations (CMC) for their models and compute micelle size distributions at a given surfactant chemical potential. Toy coarse-grained models have also been used to study the influence of alcohol-type molecules on the structural integrity of model bilayer systems [98], the influence of solutes on micelle formation [117], the influence of head group geometry on micelle shape [113], and in attempts to understand the abnormally high D_2O reflectivity experimentally observed in neutron scattering experiments for air/surfactant/water systems [118].

More sophisticated coarse-grained force fields have been developed since then [119, 120], which are intended to reproduce physical properties of specific chemicals. For example, the Shinoda-DeVane-Klein force field has been parameterized to model polyoxyethylene and alkylbenzene sulfonate surfactants as well as linear alkanes, alcohols, ethers, benzene, phenols, and fullerene, and have been used extensively to study surfactant system phase transitions [109, 110, 121, 122, 123]. Still others have conducted simulations using generic all-atom force fields, in attempts to gain structural insights at a higher resolution. As an example, Faeder *et al.* examined the structure and dynamics of water confined in reverse micelles of varing radius, and found that the structure varied significantly as a function of distance from the interior surface [124]. Stephenson and coworkers used all atom simulations in a particularly interestingly way, using structural insights from ibuprofen/surfactant/water simulations to inform a predictive thermodynamic model of micelle composition and ibuprofen solubility as a function of surfactant concentration [93]. In the present work, protocols will be developed for both coarse-grained and molecular simulations of solute loading in surfactant systems, and insights into the relationship between solute uptake, system structure, and stability will be discussed for both unary and binary solute matrices.

Chapter 2

Theory and Methods

Discussions in Chapter 1 hinted at the fact that, often times, scientific insights are limited by the tools at hand. Indeed, scientists are only as powerful as the techniques they wield, thus the next chapter will focus on developing background in the primary tools utilized in this body of work. First, statistical mechanics will be reviewed, allowing for discussion of the Monte Carlo simulation approach utilized in the present work. From there, advanced algorithms which enhance simulation precision and efficiency will be introduced, followed by an overview of the force fields leveraged.

2.1 Statistical Mechanics

Suppose one desires to describe the equilibrium properties of a large system having a constant number of particles, volume, and temperature. One way this might be accomplished is by treating this large system as a collection of several smaller systems, having consistent properties. For example, consider a collection of small systems, each having the same fixed volume, number of particles and energy. Suppose then that this collection of systems is immersed in a large temperature bath, allowed to reach thermal equilibrium, removed from the bath, and then insulated. All small systems now have the same temperature, although each system may have a unique energy. At this point, the collection of small systems,

henceforth referred to as an "ensemble," has the following features:

$$\sum_i^{n_\mathrm{s}} V_i = V_\mathrm{T}, \quad \sum_i^{n_\mathrm{s}} N_i = N_\mathrm{T}, \quad \sum_i^{n_\mathrm{s}} E_i = E_\mathrm{T}, \tag{2.1}$$

where the summations are over n_s, which is the number of systems in the ensemble, and V, N, and E, correspond to volume, number of particles, and energy, respectively. The subscript "T" refers to the total values for the ensemble.

To progress further, the first postulate of statistical mechanics must be applied [125], which states that, *for an ensemble having fixed N, V, and E, each system may be in any of the ensemble's accessible energy states.* Worded otherwise, each small system can exhibit a unique value of energy (so long as total ensemble energy is conserved), and energy states can be described in terms of occupancy, a. For example, if three systems have energy $\varepsilon = 4$, then the occupancy of the energy state $\varepsilon = 4$ is said to be 3. Thus, the energy sum in equation 2.1 can be rewritten as:

$$E_\mathrm{T} = \sum_i^{n_\varepsilon} a_i \varepsilon_i, \tag{2.2}$$

where the summation is now over energy states, a_i is the occupancy of energy state i, $\boldsymbol{a}$ is the set of all a_i for a given set of systems ($\boldsymbol{a} = \{a_1, a_2, a_3 ...\}$), and the sum of all a_i in $\boldsymbol{a}$ is equal to the number of systems in the ensemble, n_s. There are many $\boldsymbol{a}$ which satisfy equation 2.2 while summing to n_s, thus it can be said that, given a set of occupation numbers for an ensemble, $\boldsymbol{a}$, the fraction of systems in energy state a_i is $a_i / \sum_i^{n_\varepsilon} a_i = a_i / n_\mathrm{s}$.

The second and third postulates of statistical mechanics [125], which state that *an ensemble average is equivalent to a time average,* and *the weighted average of a mechanical property over all accessible states of an ensemble is equivalent to the a thermodynamic property,* can now be used to extract experimental measurables from the ensemble. In order to do so, the number of possible $\boldsymbol{a}$ for the ensemble must be determined. For a set of distinguishable objects (such as a_i's), statistics states that the number of possible arrangements is given by:

$$W(\boldsymbol{a}) = \frac{(\sum_i^{n_\varepsilon} a_i)!}{\prod_i^{n_\varepsilon} (a_i!)} = \frac{n_\mathrm{s}}{\prod_i^{n_\varepsilon} (a_i!)}, \tag{2.3}$$

thus, considering that the energy state of a system can be described by a_i belonging to

many $\boldsymbol{a}$, the overall probability that a system is in a state a_i is given by:

$$P_i = \frac{\overline{a}_i}{n_\mathrm{s}} = \frac{1}{n_\mathrm{s}} \frac{\sum_j^{n_{\boldsymbol{a}}} W(\boldsymbol{a}_j)\, a_{ij}}{\sum_j^{n_{\boldsymbol{a}}} W(\boldsymbol{a}_j)}, \tag{2.4}$$

where $n_{\boldsymbol{a}}$ is the number $\boldsymbol{a}$ that satisfies equation 2.2. Any ensemble average property, $\overline{M}$, can then be computed by:

$$\overline{M} = \sum_i^{\varepsilon} M_i P_i, \tag{2.5}$$

where M_i is the value of the property of interest in state i. A caveat of equations 2.4 and 2.5 is that the sum over $\boldsymbol{a}$ is very large, making these equations exceedingly difficult to compute. To overcome this, a convenient feature of $W(\boldsymbol{a})$ is utilized, namely that it is sharply peaked about a single maximum value when n_s is large, and has the greatest value when all a_i are equivalent. The implication of this is that there exists a single distribution, $\boldsymbol{a^*}$ with $a_1 = a_2 = a_3 ... = a_i^*$, that corresponds to the most probable distribution for the ensemble. Furthermore, when n_s is large, $\boldsymbol{a}$ not equal to $\boldsymbol{a^*}$ become negligible, allowing equation 2.4 to be rewritten as:

$$P_i = \frac{1}{n_\mathrm{s}} \frac{W(\boldsymbol{a^*})\, a_i^*}{W(\boldsymbol{a^*})} = \frac{a_i^*}{n_\mathrm{s}} = \frac{\overline{a}_i}{n_\mathrm{s}}, \tag{2.6}$$

and what remains is to find a_i^*. To do so, the Legrange method of undetermined multipliers is applied in conjunction with Stirling's approximation and the maximum term method to yield [125]:

$$P_i = \frac{a_i^*}{n_\mathrm{s}} = \frac{\exp(-\beta\varepsilon_i)}{\sum_i^{n_\varepsilon} \exp(-\beta\varepsilon_i)} \tag{2.7}$$

where β can be shown to be equivalent to $1/k_\mathrm{B}T$, and the summation in the denominator is commonly referred to as the "canonical" partition function, Q_{NVT}. The current formulation of Q_{NVT} has been derived for quantum mechanics, and thus is given as a discrete summation, but as the classical limit is approached, it may be written as a continuous function:

$$Q_{NVT} = \frac{1}{N! h^{3N}} \int \dots \int \exp\left[-\frac{H(\boldsymbol{p},\boldsymbol{q})}{k_\mathrm{B}T}\right] d\boldsymbol{p}\; d\boldsymbol{q}. \tag{2.8}$$

where h is Planck's constant, H is Hamiltonian, which is the sum of potential and kinetic energy, and $\boldsymbol{p}$ and $\boldsymbol{q}$ are the momenta and coordinates for the system, respectively, where the integrals are over all possible coordinates and momenta. By separating kinetic and

potential energy in the Hamiltonian and integrating over momenta, Q_{NVT} can be further simplified to:

$$Q_{NVT} = \frac{1}{N!\Lambda^{3N}} \int \dots \int \exp\left[-\frac{U(\boldsymbol{q})}{k_{\mathrm{B}}T}\right] d\boldsymbol{q} = \frac{1}{N!\Lambda^{3N}} Z \tag{2.9}$$

where Λ is the de Broglie wavelength, U is the potential energy, and Z is the configurational integral. Within the context of the present work, Monte Carlo simulations are utilized in an attempt to compute Z.

2.2 Metropolis Monte Carlo

Before delving into the details of Monte Carlo (MC), the concept of "phase space" must be introduced. Phase space is a collection of momenta and coordinates that specify every possible configuration of a given system [126]. In a Cartesian formulation, at any given time, a system's location within this space can be specified with a set of $6N$ coordinates, corresponding to $p_x,\ p_y,\ p_z,\ q_x,\ q_y,$ and q_z for each of the N particles in a given system; as was shown above, the configurational integral is related to the set of $\boldsymbol{q} = \{q_x,\ q_y,\ q_z\}$ of a system's phase space coordinates.

The second postulate of statistical mechanics states that, for adequate sampling of phase space, the ensemble average is equivalent to the time average, which is what allows methods such as molecular dynamics (MD) and MC to be used to predict macroscopic properties; however, due to the deterministic nature of MD, situations may arise wherein adequate phase space sampling is not achieved over reasonable simulation lengths. For example, if large energetic barriers exist between to sequential phase space coordinates, a deterministic simulation may become trapped, hampering sampling. Similarly, if the phase space of a system is very large, it may take a considerable amount of time for a molecular dynamics simulation to sample all relevant locations therein. Monte Carlo methods can overcome these barriers because sampling (*i.e.* system evolution) is dependent on the *probability* to be in a given state rather dynamics, thus, a MC simulation can hop between likely states, despite the potential that they are very far apart or may be separated by large energy barriers.

In its simplest form, MC is a sample-mean integration method [127], which is conducted through repeated random system perturbations, followed by acceptance or rejection of the

attempt, according to some acceptance rule. The most straightforward application, then, is to make use of equations 2.5, 2.7, and 2.9, yielding the following for a canonical ensemble, when the integrals over all phase space coordinates are replaced with a sum over system configurations (Γ), which gives:

$$\overline{M}_{NVT} \approx \frac{\sum_i^{n_\Gamma} M_i \exp\left(\frac{-U_i}{k_\mathrm{B}T}\right)}{\exp\left(\frac{-U_i}{k_\mathrm{B}T}\right)} = \overline{M}_{\mathrm{sim}}, \tag{2.10}$$

where n_Γ is the total number of configurations in the set $\mathbf{\Gamma}$ such that $\mathbf{\Gamma} = \Gamma_1, \Gamma_2, \Gamma_3 ... \Gamma_{n_\Gamma}$ In reality, however, this approach is highly inefficient, as only a small fraction of configurations contribute significantly to the sums. With this in mind, one may choose to use an improved approach such as importance sampling. In this framework, configurations are now chosen from a distribution, $\rho(\Gamma)$, which is *known* to contribute significantly to the sums.

The challenge, now, is to define $\rho(\Gamma)$ such that *i.* the partition function need not be known, and *ii.* states are sampled according to $\rho(\Gamma_{\mathrm{ens}})$. This can be achieved through use of the Metropolis MC scheme [128]. Here, a so-called "Markov chain" of states is utilized [126, 127], in which each state depends only on the previous state. This chain can be described through the following:

$$\boldsymbol{\rho_2} = \boldsymbol{\pi}_\mathrm{T} \boldsymbol{\rho_1}, \tag{2.11}$$

such that:

$$\boldsymbol{\rho_n} = \boldsymbol{\pi}_\mathrm{T}^{n-1} \boldsymbol{\rho_1}, \tag{2.12}$$

where in the above, $\boldsymbol{\rho}$ is a probability distribution containing the probabilities for a system to be in a given state $\rho(\Gamma_i)$, $\boldsymbol{\pi}_\mathrm{T}$ is the stochastic probability matrix, whose elements, π_{ij}, describe the probability to transition from $\rho(\Gamma_i)$ to $\rho(\Gamma_j)$, and where π_{ij} has the property that $\sum_j \pi_{ij} = 1$, and the superscript on $\boldsymbol{\pi}_\mathrm{T}$ implies the number of applications. As successive applications of $\boldsymbol{\pi}_\mathrm{T}$ approach infinity, a distribution is achieved for which further applications return the same distribution. The distribution, Γ^*, which satisfies this condition $\rho(\Gamma_n) = \boldsymbol{\pi}_\mathrm{T} \rho(\Gamma_n)$ is known as the limiting distribution, and once this condition is met, the system is said to be at steady state. At this point, the system has achieved "detailed balance," signifying that the Markov chain is ergodic, meaning any state of the system can be accessed [126, 127].

At this point, one needs to find the $\boldsymbol{\pi}_{\mathrm{T}}$ which yields the limiting distribution. To aid in this process, an additional, unnecessarily strong constraint is applied:

$$\rho(\Gamma_i)\pi_{ij} = \rho(\Gamma_j)\pi_{ji}, \tag{2.13}$$

which is known as "microscopic reversibility [126, 127]." Utilizing this, and recalling the significance of π_{ij} (namely that it is the probability to transition from a state i to j), equation 2.13 can be rewritten as:

$$\rho(\Gamma_i)\alpha_{ij}P_{ij} = \rho(\Gamma_j)\alpha_{ji}P_{ji}, \tag{2.14}$$

where α_{ij} is called the "underlying matrix [126, 127]," and gives the probability to attempt a move from i to j, and P_{ij} is the probability to accept that move.

The Metropolis method requires a symmetric underlying matrix [128], such that $\alpha_{ij} = \alpha_{ji}$, allowing Equation 2.14 to be rewritten as:

$$\rho(\Gamma_i)P_{ij} = \rho(\Gamma_j)P_{ji}. \tag{2.15}$$

Using the canonical ensemble as an example, it was shown that:

$$\rho_i = \exp\left(\frac{-U_i}{k_{\mathrm{B}}T}\right) Q_{NVT}^{-1}, \tag{2.16}$$

thus:

$$\frac{\rho(\Gamma_i)}{\rho(\Gamma_j)} = \frac{\exp(-U_i/(k_{\mathrm{B}}T))}{\exp(-U_j/(k_{\mathrm{B}}T))} = \exp(-\Delta U/(k_{\mathrm{B}}T)), \tag{2.17}$$

where U_i is the energy of configuration Γ_i, demonstrating that Q_{NVT}, need not be calculated. From this point, Metropolis suggested the following acceptance rules [128]: if the probability to be in state j is greater than that to be in state i, accept the transition to state j; otherwise, accept with the probability emerging from equation 2.17.

2.3 The Gibbs Ensemble

In the above section, the Metropolis Monte Carlo method for simulations in a canonical (NVT) ensemble was derived. In practice, however, this ensemble is of limited use, as many phenomena of interest occur at constant pressure (rather than volume), thus one might consider use of an isobaric-isothermal ensemble (NpT). Furthermore, simulations in the

NVT or NpT ensemble, which are limited to a a closed system, can become problematic if one desires to study problems relating equilibria of bulk phases. For example, if one desires to study the equilibrium of a vapor and liquid, a very large simulation box having an explicit interface would be required. An efficient alternative to the aforementioned setup involves introduction of an open "Gibbs" ensemble [107, 108]. In such an approach, multiple simulation boxes are utilized, having no explicit interface, but being in thermal, chemical, and mechanical equilibrium with each other.

The partition function for a multicomponent Gibbs–NpT ensemble can be derived following an approach similar to what was done for the canonical ensemble, where now the system is allowed to fluctuate in volume, and has an additional constraint relating to the sum of molecules of each component. The result is the following partition function [107, 108, 126]:

$$\Delta_{N_1,N_2,\ldots N_c} = \prod_i^{n_\mathrm{s}} \prod_j^{n_\mathrm{comp}} \frac{1}{\Lambda^{3N_{ij}} N_{ij}!} \int V_i^{N_{ij}} \exp\left(\frac{-pV_i}{k_\mathrm{B}T}\right) dV_i \times \int \exp\left(\frac{-U(\boldsymbol{s}^{N_{ij}}; L)}{k_\mathrm{B}T}\right) d\boldsymbol{s}^{N_{ij}} \tag{2.18}$$

where the first and second products are over number of interacting systems and number of components and $\boldsymbol{s}$ corresponds to the system coordinates when scaled by the length of the simulations box, L. The following acceptance rules emerge for the Gibbs ensemble [107, 108, 126]:

1. Particle displacement:
$$P_\mathrm{acc} = \min\left[1, \exp\left(\frac{-\Delta U}{k_\mathrm{B}T}\right)\right] \tag{2.19}$$

2. Volume change for a single box from state o$\rightarrow$ n
$$P_\mathrm{acc} = \min\left[1, \exp\left(\frac{-\Delta U - p\Delta V + k_\mathrm{B}TN\ln(V_\mathrm{n}/V_\mathrm{o})}{k_\mathrm{B}T}\right)\right] \tag{2.20}$$

3. Particle exchange from box A$\rightarrow$ B
$$P_\mathrm{acc} = \min\left[1, \frac{N_\mathrm{A}V_\mathrm{B}}{(N_\mathrm{B}+1)V_\mathrm{A}} \exp\left(\frac{-\Delta U_\mathrm{A} - \Delta U_\mathrm{B}}{k_\mathrm{B}T}\right)\right] \tag{2.21}$$

where P_{acc} is the probability to accept the perturbation and the subscripts "o" and "n" refer to the new and old configurations, respectively. The first acceptance rule recovers equation 2.17, and ensures thermal equilibration of the system, while the second rule leads to mechanical equilibration. The first rule is common to all ensembles, and the second with the NpT ensemble, whereas the third rule is unique to a Gibbs ensemble and allows for chemical equilibrium between two systems which are not in direct contact.

2.4 Enhanced Sampling Techniques

Monte Carlo simulation methods offer extreme flexibility in choice of sampling technique, and one's choice of sampling approach has a significant influence on the resulting simulation efficiency. In this section, an overview of advanced sampling methods utilized in the present work will be discussed.

2.4.1 Transfer Free Energy and Biasing Potentials

Computation of the free energy of transfer of a solute between two phases can yield valuable insights into the thermodynamics of solute uptake. Because free energy is a state function, the process of transferring species i between coexisting phases A and B can be broken down into two components: the free energy of solvation for species i in phase A, ΔG_i^{A}, and the free energy of solvation for species i in phase B, ΔG_i^{B}. These solvation free energies can be defined as [129]:

$$\Delta G_i^{\text{A}} = \mu_i^{*\text{A}} - \mu_i^{*\text{g}}, \tag{2.22}$$

where $\mu_i^{*\text{A}}$ and $\mu_i^{*\text{g}}$ are the pseudo chemical potential of i in phase A and an ideal gas phase, respectively. In this context, the pseudo chemical potential is defined as [129]:

$$\mu_i^{*x} = \mu_i^x - k_{\text{B}}T \ln(\rho_i^x \Lambda_i), \tag{2.23}$$

where the pseudo potential of i in phase x, μ_i^{*x} describes the change in free energy when a solute is inserted into phase x at a fixed position, the rightmost term corresponds to a "liberation free energy," for when the solute is released from its fixed position, and μ_i^x is the true chemical potential for the solute in phase x. The variables ρ_i^x and Λ_i are the number density of i in phase x and the de Broglie wavelength for species i.

The transfer free energy for i from phase A to B can then be computed through:

$$\begin{aligned} \Delta G_{\mathrm{A}\to\mathrm{B}} &= \Delta G_i^{\mathrm{B}} - \Delta G_i^{\mathrm{A}} \\ &= (\mu_i^{*\mathrm{B}} - \mu_i^{*\mathrm{g}}) - (\mu_i^{*\mathrm{A}} - \mu_i^{*\mathrm{g}}) \\ &= \mu_i^{*\mathrm{B}} - \mu_i^{*\mathrm{A}}. \end{aligned} \tag{2.24}$$

Using equation 2.23, equation 2.24 can be rewritten:

$$\Delta G_{\mathrm{A}\to\mathrm{B}} = \mu_i^{\mathrm{B}} - k_{\mathrm{B}}T\ln(\rho_i^{\mathrm{B}}\Lambda_i) - \mu_i^{\mathrm{A}} - k_{\mathrm{B}}T\ln(\rho_i^{\mathrm{A}}\Lambda_i). \tag{2.25}$$

Recalling the condition of chemical equilibrium, namely that $\mu_i^{\mathrm{A}} = \mu_i^{\mathrm{B}}$, equation 2.25 reduces to simply:

$$\Delta G_{\mathrm{A}\to\mathrm{B}} = -k_{\mathrm{B}}T\ln\left(\frac{\rho_i^{\mathrm{B}}}{\rho_i^{\mathrm{A}}}\right). \tag{2.26}$$

In many systems of interest the solute can have very low solubility in one or more of the phases. From the MC standpoint, low solubility in a phase means that the probability to observe the solute in that phase is low, and that exceedingly long simulations would be required to collect adequate statistics (observations) to provide reliable values of ρ_i [130]. It turns out, however, that a simple trick can be used to circumvent this problem. By introduction of a uniform biasing potential on i in troublesome phases, the solute can be forced to spend an equal amount of simulation time in each phase, greatly enhancing sampling efficiency [66]. Furthermore, the influence of the biasing potential on the resultant free energy can be immediately accounted for through:

$$\Delta G_{2\to1} = \Delta G_{2\to1}^{\mathrm{biased}} - (E_1 - E_2), \tag{2.27}$$

where $\Delta G_{2\to1}^{*}$ is the free energy resulting from the biased simulation, and E_j is the biasing potential applied to phase j.

2.4.2 Molecule Identity Switch

Introduction of biasing potentials encourages acceptance of particle transfer moves by compensating for the large transfer free energies that can be present for a solute between coexisting phases. An additional way to boost the acceptance rate is to circumvent the

simultaneous problems of solvent cavitation and solvent/solute association. Consider a system where the transfer of a homologue series of linear alcohols, ranging carbon chain length from one to ten, between a bulk water and vapor phase. Because of its small size, methanol will have the least trouble finding volume to insert itself into, however larger molecules like decanol will struggle to find ample volume in the new phase. Furthermore, the transfer of hydrogen bonding solutes from, for example, vapor to hydrogen bonding solvent phases can be troublesome if the solute is not able to form sufficient hydrogen bonds to offset the loss of entropy experienced when moving from a vapor to condensed phase. A way one might circumvent this issue is through introduction of molecule identity switch (or "swatch") moves [66, 131, 132]. Here, in contrast to particle transfer moves, which "swap" an entire molecule between boxes, a pre-specified pair of solutes which occupy different simulation boxes have their identity exchanged. For example, if methanol and ethanol are chosen as a swatch pair, the ethanol molecule in its phase has a segment removed and has its interaction parameters converted to those of methanol, and *vice versa* for the methanol molecule. This approach can be expanded to the following:

$$\mathrm{CH_3OH} \overset{\mathrm{CH_2}}{\rightleftharpoons} \mathrm{CH_3CH_2OH} \overset{\mathrm{CH_x}}{\rightleftharpoons} ... \overset{\mathrm{CH_2}}{\rightleftharpoons} \mathrm{CH_3(CH_2)_9OH}, \tag{2.28}$$

where CH_2 is a methylene segment. Relating this to the previous discussion of transfer free energies, this approach allows for the transfer free energy of decanol to be computed by only swapping methanol, which yields a reference free energy, and using swatch moves on intermediate molecules to obtain incremental free energies for the addition of interaction sites. This approach which will be shown to be quite useful in Chapter 4.

2.4.3 Configurational-bias Monte Carlo

In practice, most interesting problems involve molecules with complex structures, which necessitates introduction of advanced MC move types for sampling of molecular conformations. Perhaps the most intuitive way to go about this is to simply perform atom translations, however, this approach can be highly inefficient. Intramolecular interactions tend to have high energetic penalties associated with small perturbations from the equilibrium structure. For this reason, atom translations tend to be limited to very small displacements [126]. Molecular conformation also comes into play when attempting to insert a

flexible molecule into a condensed phase, as is often done in the Gibbs ensemble. As one might imagine, the probability of accepting a rigid insertion of a large flexible molecule into such a phase has a very low acceptance probability. Thus, what is needed for this type of MC move is a "smart" method for inserting a molecule while allowing its conformation to change.

Introduced in the early 1990's as an extension of the Rosenbluth and Rosenbluth random walk algorithm for lattice polymers [133], configurational-bias Monte Carlo (CBMC) provides an efficient method to sample molecular conformations in continuous space [104, 105, 106]. In this approach, structures are sampled by selecting a molecule, removing all (or some portion) of it, and attempting to re-grow the eliminated segments in a new conformation through a series of several trials. In order to demonstrate how this algorithm works, we will consider the complete regrowth of a linear molecule having rigid bonds, and n_{s} segments, following the approach of Vlugt *et al.* [105]; a move which might be attempted when transferring a molecule between simulation boxes in a Gibbs ensemble. The first task at hand is to place the first segment of the chain, s_0. This is accomplished by randomly generating sites within the simulation box. One of the trial sites, is then selected for further development with the following probability:

$$P_{s_0,t}^{\mathrm{place}}(\boldsymbol{t}) = \frac{\exp\left(-\beta U_{s_0,t}^{\mathrm{NB}}\right)}{w_{s_0}^{\mathrm{new}}}, \tag{2.29}$$

where $P_{s_0,t}^{\mathrm{place}}(\boldsymbol{t})$ is the probability to choose site t from the set of $\boldsymbol{t}$ trial locations for placement of the first segment. In the above, $U_{s_0,t}^{\mathrm{NB}}$ is the non-bonded energy of s_0, and $w_{s_0}^{\mathrm{new}}$ is the Rosenbluth weight of the set of trials, $\boldsymbol{t}$, given by:

$$w_{s_0}^{\mathrm{new}} = \sum_{t=0}^{n_t} \exp\left(-\beta U_{s_0,t}^{\mathrm{NB}}\right), \tag{2.30}$$

where n_t is the number of generated trial sites. With the first segment placed, the remaining units are grown sequentially, through the following procedure. Let i be the index of the segment being grown. Rather than generating trial sites randomly as was done for the s_0, s_i sites are generated based on the Boltzmann weight of the resulting conformation:

$$P_{s_i,t}^{\mathrm{generate}}(\boldsymbol{t})d\boldsymbol{t} = \frac{\exp\left(-\beta U_{s_i,t}^{\mathrm{B}}\right) d\boldsymbol{t}}{\int \exp\left(-\beta U_{s_i,t}^{\mathrm{B}}\right) d\boldsymbol{t}}, \tag{2.31}$$

where $U_{s_0,t}^{\mathrm{NB}}$ is the intramolecular energy of the chain being regrown. A trial site is then selected for s_i using the criteria given in 2.29, and replacing s_0 with s_i. This process is repeated until all segments of the chain are regrown, resulting in the following total probability ($P_{\mathrm{chain}}^{\mathrm{grow,new}}$)and Rosenbluth factor ($W_{\mathrm{chain}}^{\mathrm{grow,new}}$) for chain regrowth:

$$P_{\mathrm{chain}}^{\mathrm{grow,new}} = \prod_{i=1}^{n_{\mathrm{s}}} P_{i,t_{\mathrm{choose}}}^{\mathrm{place}} P_{i,t_{\mathrm{choose}}}^{\mathrm{generate}}, \tag{2.32}$$

and

$$W_{\mathrm{chain}}^{\mathrm{grow,new}} = \prod_{i=1}^{n_{\mathrm{s}}} w_i^{\mathrm{new}}. \tag{2.33}$$

In equations 2.32 and 2.32, the product is over grown segments, and the probabilities on the right-hand-side of equation 2.32, are for the chosen trial of the i^{th} segment.

Before the move can be accepted or rejected, the bias introduced while regrowing the chain must be removed. To do so, a total Rosenbluth weight, $W_{\mathrm{chain}}^{\mathrm{grow,new}}$ is computed for the old configuration by repeating the entire regrowth process for $n_t - 1$ trials, and using the chain old configuration for the n_t^{th} trial. The probability for the old and new configurations are then given by:

$$P_{\mathrm{new}} = W_{\mathrm{chain}}^{\mathrm{grow,new}} P_{\mathrm{chain}}^{\mathrm{grow,new}} \quad \text{and} \quad P_{\mathrm{old}} = W_{\mathrm{chain}}^{\mathrm{grow,old}} P_{\mathrm{chain}}^{\mathrm{grow,old}} \tag{2.34}$$

resulting in the following acceptance rule:

$$P_{\mathrm{acc}} = \min \left[1, \frac{W_{\mathrm{chain}}^{\mathrm{grow,new}}}{W_{\mathrm{chain}}^{\mathrm{grow,old}}} \right], \tag{2.35}$$

where P_{acc} is the probability to accept the entire regrowth.

Since its initial creation, the CBMC algorithm has been continually improved and increased in complexity. The first major advancement of the algorithm was incorporation of a "coupled-decoupled" growth [134], allowing for application to branched molecules. In order to understand how this extension functions, consider a pseudo molecule, which takes the form of the carbon atoms in 2-methylbutane. If one was to regrow the two terminal units bonded to the branch point ("A" and "B"), three angles and two dihedrals would need to be specified, where the three angles include the angle between groups ("A" and "B"). In a coupled-decoupled growth, each angle would be determined independently, and in a

fashion decoupled from determination of the torsion, and selection of the torsion then be coupled with the non-bonded contributions to energy. In departure from previous methods, use of this approach for branched molecules leads to angles which are all generated on the proper distribution, prevents the need to simultaneously regrow entire molecules such as the pseudo branched molecule described above (which is highly inefficient), and increases the number of trials that can be attempted for angle generations at a given computational expense.

Aiding further in simulation efficiency, "dual-cutoff" CBMC was introduced [134]. This supplementation further increased the efficiency of CBMC calculations by using a shorter-ranged non-bonded interaction energy when computing P^{generate}, and then correcting for the truncated energy the acceptance rule. The final variety of CBMC to be discussed here, "self-adapting-fixed-endpoint," or "SAFE" CBMC [135] presents a solution to the difficult task of regrowing interior segments of molecules, which is particularly useful when simulating structures that are cyclic, of large molecular weight, and/or tethered to a surface. This extension introducing a guiding bias, which drives the regrowth to a point where the molecular structure is not broken; without this guiding bias, the probability of closing the structure would be prohibitively low.

2.5 Molecular Mechanics Force Fields

In order to compute the energies or forces required to drive a molecular simulation, a set of equations and corresponding parameters known as a force field (FF), is required, which describes the inter- and intra-molecular interactions between molecules. The nature of a force field (*i.e.* the functional form and parameters) is a key component to determining the accuracy of a simulation. As an example, consider a molecular angle bending vibration. Many force fields treat these vibrations as harmonic, despite the fact that they are better captured through an anharmonic representation. The assumption of angle harmonicity holds reasonably well for most molecules at typical simulation conditions, where vibrations remain close to the potential minimum, however problems can arise at extreme conditions (*i.e.* geothermal), where an angle becomes more likely to sample high-energy conformations. In principle, one could exclusively use models which treat all angles anharmonically, however an important factor to keep in mind is the balance between force field complexity (which

is often correlated with performance) and computational expense/usability.

Generally, force fields can be described by extent of resolution, where higher resolution yields finer microscopic details, but lower resolution allows for larger simulations. For example, in the "all-atom" description (AA), each atom has its own interaction site. The next level of resolution is the "united-atom" class (UA), where hydrogen atoms not belonging to polar functional groups are combined with their parent heavy atom to create a single interaction site; for example, methane would be treated as a single bead. From there, force fields colloquially referred to as "coarse-grained" (CG), generally refer to any resolution having multiple heavy atoms combined into a single interaction site, and that can potentially be used to study systems on the mesoscopic level (*i.e.* falls between micro- and macroscopic).

The present work heavily utilizes the Transferable Potentials for Phase Equilibria (TraPPE) force field [134, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145]. In contrast to the bulk of existing force fields, which are generally parameterized only at ambient or biological conditions (Charmm [146], Amber [147], and Gromos [148], to name a few), the TraPPE force field is specifically designed to reproduce phase equilibria over a wide range of temperatures and pressures, and as such, is particularly well suited for studies of uptake and retention. Another particularly noteworthy aspect of the TraPPE force field is, as its name suggests, transferability. Rather than requiring unique parameters for every single atom, bond, angle, and dihedral, depending on the parent molecule, TraPPE defines atoms as building blocks that can be used to construct a variety of different molecule types.

The TraPPE-UA force field, which assumes energy is a pair-wise additive quantity, is comprised of a non-bonded and a bonded component, where the former contains terms for van der Waals and Coulomb interactions, while the latter contains terms for bonds, angles, and torsions. The non-bonded component of the force field between two interaction sites takes the following form [137]:

$$E_{\text{non-bonded}} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \tag{2.36}$$

where $E_{\text{non-bonded}}$, ϵ_{ij}, σ_{ij}, and r_{ij} are the non-bonded energy, potential well depth, average bead diameter, and interaction site separation distance, respectively, while q_i and ϵ_0 are the partial charge of species i, and the permittivity of vacuum, respectively. The first term of equation 2.36 is called the Lennard-Jones potential [149], and the second term, the

Coulomb potential. The Coulomb potential is particularly long ranged, decaying as r_{ij}^{-1}, and is typically handled in one of two ways. In the first approach, known as the "group-based-cutoff" method, molecules are divided into smaller groups having net-zero charge. In this way, interactions between groups can be considered as multipole-multipole interactions, rather than point charge-point charge interactions, resulting in a decay which goes as r^{-3} at the slowest, corresponding to dipole/dipole interactions. The Ewald summation [127] approach provides a more accurate method, at the cost of higher computational expense. In this approach, charge interactions are broken into long and short range interactions, where the former are handled in reciprocal space (where the summation over interactions rapidly converges) and the latter, in real space.

The bonded component of the TraPPE-UA force field generally has the form [137]:

$$E_{\text{bonded}} = \sum_{i}^{\text{angles}} \frac{k_\theta}{2}(\theta_{\text{eq}} - \theta_i)^2 + \sum_{i}^{\text{bonds}} \frac{k_r}{2}(r_{\text{eq}} - r_i)^2 + \sum_{i}^{\text{dihedrals}} \sum_{n=0}^{\text{terms}} a_n \left(1 + \cos(n\phi_i)\right), \quad (2.37)$$

where E_{bonded}, k_θ, θ_{eq}, and θ are the bonded energy, angle bending force constant, equilibrium bond angle, and bond angle, respectively. The values k_r, r_{eq}, r, a_i, and ϕ correspond to the bond stretching force constant, equilibrium bond distance, bond distance, fourier constant, and dihedral angle, respectively. In the TraPPE-UA force field, bonded potential parameters are either taken from experiment or extracted from relaxed quantum mechanical energy surface scans of the relevant degrees of freedom. Note that traditionally, the TraPPE-UA force field assumes rigid bonds (*i.e.* the second term of Equation 2.37 is set to zero). Non-bonded parameters for the TraPPE-UA force field are typically fit to vapor-liquid coexistence data, though recently additional fitting properties have been included such as the dielectric constant, triple point, and binary phase equilibria [150].

The Shinoda-DeVane-Klein (SDK) coarse grained force field [109, 110, 121, 122, 123] is also used for a significant portion of surfactant based work in this dissertation. This force field typically uses a two- or three-to-one heavy atom to interaction site mapping (*i.e.* propane would be represented as a single bead), and three water molecules are treated as a single interaction site. The SDK force field was chosen for the coarse-grained surfactant simulations because it has been specifically designed for use in such systems, and has been demonstrated to perform well in such applications [114]. Furthermore, it has been developed specifically to reproduce surface tension data, which is one of the major aspects contributing

to surfactant/solvent/solute phase behavior and solubility.

The SDK force field is also pair-wise additive [109], with the non-bonded component of force field taking the form of a Mie potential [151], as is shown below for a pair of interaction sites i and j:

$$E_{\text{non-bonded}} = \left(\frac{n}{n-m}\right)\left(\frac{n}{m}\right)^{m/(n-m)} \epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{n} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{m}\right], \tag{2.38}$$

where n and m control the range of van der Waals interactions, and all other parameters are as given in equation 2.36. The n,m parameters are taken to be 12,4 for any interaction with water, and 9,6 otherwise. Coulomb interactions are not included, which, when combined with the reduced number of interaction sites required to describe a molecule, significantly enhances computational efficiency, while still providing microscopic-level details. Bonded interactions are only specified through harmonic two- and three-bead interactions, like those given in equation 2.37, and are fit to structural data generated from all-atom simulations.

Chapter 3

An Iterative Configurational-Bias Monte Carlo Annealing Approach for Semi-Automated Fitting of Intramolecular Bonded Potentials

Acceptance of computer simulations as a laudable means of chemical research has grown significantly in recent years, with the 2013 Nobel Prize being awarded to Martin Karplus, Michael Levitt, and Arieh Warshel for their contributions towards simulation models for chemical systems [152], and the 2016 American Chemical Society Meeting having the central theme of "Computers in Chemistry" [153]. This surge of interest has lead to a significantly increased demand for molecular mechanics force fields which are diverse, accurate, and user friendly. For those working in force field development, this demand has driven continual development of parameter sets to improve existing descriptions of chemical compounds [154, 155, 156, 157, 158]. In general, however, the process of force field fitting is difficult and tedious and has created a market for new force field development tools [159, 160, 161, 162, 163, 164, 165].

Historically, the bonded components of molecular mechanics force fields, such as bonds, angles, and dihedrals, were either taken from experimental data or obtained by fitting force

field functions directly to quantum mechanical (QM) scans of the degree of freedom (DOF) of interest [136, 146, 166, 167, 168, 169, 170], however, both of these cases suffer from the assumption that conformational DOF are decoupled, which can lead to strained structures in complex molecules. Furthermore, approaches such as those mentioned above result in parameters which depend on the chemical composition of the entire molecule rather than just neighboring atoms, meaning the resulting force field is not inherently transferable (*i.e.* parameters can not *necessarily* be used as building blocks to describe new compounds).

Recently, approaches have been developed which attempt to address the issue of fitting bonded potentials for coupled degrees of freedom. For example, Guvench *et al.* developed a simulated annealing tool which simultaneously fits all of the torsional potentials within a given molecule [159]. This method is particularly well suited to capture concerted dihedral rotation in flexible cyclic compounds, however, it only fits across a single molecule at a time and does not account for angle/dihedral, bond/dihedral, or bond/angle coupling. Grimme created QMDFF software, which simultaneously fits all degrees of freedom within a single molecule through a series of QM-based calculations [161]. This program addresses the issue of coupled DOF within a molecule, but is non-transferable and uses an unconventional force field form. Betz *et al.* created Paramfit software, which uses hybrid minimization/genetic algorithm to address several of the aforementioned issues, providing simultaneous fits and doing so using a relatively small set of QM calculations, however, the program is limited to the AMBER force field [160].

Here, we attempt to build upon the aforementioned approaches to develop a method which *i.* can capture concerted motion due to coupling between DOF, *ii.* is mindful of non-bonded contributions to conformational energy (*i.e.* interactions such as Pauli-exclusion, van der Waals attraction, Coulomb contributions, etc.), *iii.* is aware of existing parameters and thus can be used to fill gaps in existing force field parameters, and *iv.* is flexible with respect to force field form and interfacing software.

The remainder of the chapter is divided as follows: A basic overview of the fitting approach is provided in Section 3.1, while characteristics of the approach are discussed in Section 3.2. Sample applications are provided in Section 3.3, followed by conclusions in Section 3.4.

3.1 Method Overview

The present approach, referred to herein as the Iterative Configurational-bias Monte Carlo Simulated Annealing (ICBSA) method, makes use of the Monte Carlo algorithm [128], described in Chapter 2, Section 2.2, the configurational-bias algorithm [104, 105, 106, 134], described in Chapter 2, Section 2.4, and the simulated annealing approach [171, 172], which drives approximate global optimization of a target function. The ICSBA works as follows: First, a set of force field equations and a molecule set are specified. Because this method is only intended for fitting bonded DOF, parameters for non-bonded interactions are treated as fixed values. A set of initial guess parameters are then provided for the bonded potentials, which can be approximations stemming from existing force fields, experimental data, or values obtained from exploratory QM calculations. The molecule set and corresponding parameters are then used to generate a collection of perturbed configurations for each molecule; in the present work, this is accomplished by running a short configurational-bias Monte Carlo (CBMC) calculation on each isolated molecule. The CBMC approach is particularly well suited for this type of application due to its ability to quickly generate uncorrelated structures in a fashion that is uninhibited by, for example, the presence of high rotational barriers that may be found for some molecules, however in practice, other structure perturbation methods such as atom translations or molecular dynamics could also be utilized. In the next step, the QM energy (E^{QM}) and non-bonded molecular mechanics (MM) energy ($E^{\mathrm{MM,NB}}$) are computed for each structure. If a united-atom (UA) force field is being used, where hydrogen atoms not belonging to functional groups are excluded, the missing hydrogen atoms are added to the structure and have their positions optimized during the QM energy calculation. Following, the new set of bonded parameters are fit through simulated annealing, which aims to minimize the difference between E^{QM} and the sum of bonded and non-bonded molecular mechanics energies, $E^{\mathrm{MM,B}} + E^{\mathrm{MM,NB}}$.

The annealing process progresses according to the following acceptance criteria:

$$P_{\mathrm{acc}} = \min\left[1, \exp\left(-\frac{F_{\mathrm{new}} - F_{\mathrm{old}}}{k_{\mathrm{B}} T_{\mathrm{A}}}\right),\right] \tag{3.1}$$

where k_{B} is the Boltzmann constant, T_{A}, is the annealing temperature, and F is a fitness function. The annealing temperature is decreased throughout the ICBSA calculation, that drives acceptance of new parameter sets, which increase F with a progressively smaller

probability. The fitness function is given by:

$$F = \sqrt{\frac{\sum_{i=0}^{n_{\text{molecs}}} \sum_{j=0}^{n_{\text{conf}}} w_{ij} \left[\frac{E_{ij}^{\text{QM}} - \left(E_{ij}^{\text{MM,NB}} + E_{ij}^{\text{MM,B,FIT}} \right)}{n_{\text{DOF},i}} \right]^2}{\sum_{i=0}^{n_{\text{molecs}}} \sum_{j=0}^{n_{\text{conf}}} w_{ij}}}, \tag{3.2}$$

where n_{molecs}, n_{conf}, and n_{DOF} are the number of molecules, number of configurations per molecule, and the number of degrees of freedom in each molecule, respectively. The QM energy, MM non-bonded energy, and MM bonded energy from the current fitting parameters are given by E_{ij}^{QM}, $E_{ij}^{\text{MM,NB}}$, and $E_{ij}^{\text{MM,B,FIT}}$, respectively. The weights, w_{ij}, are defined as:

$$w_{ij} = \exp\left[\frac{-E_{ij}^{\text{QM}}}{k_{\text{B}} T_{\text{w}} \left(n_{\text{DOF},i} \right)} \right], \tag{3.3}$$

where T_{w} is the weighting temperature. Throughout the annealing process, parameters for the bonded force field are randomly perturbed, and the resulting new sets are accepted or rejected according to 3.1. Maximum displacements in each parameter type are decreased at the end of each cycle, which, in combination with the decreasing temperature, damps fluctuations in parameter values. Due to the random nature of the Monte Carlo method, final resulting parameters may not reflect the parameter set that had the smallest fitness, so throughout the annealing, "best" parameters are tracked.

3.2 Software Capabilities and Characteristics

The present software is designed to be flexible in terms of annealing controls, force field form, and interfacing software. Currently, two cooling schedules for T_{A} are supported:

$$T_{\text{A}} = T_0 \exp\left(-\tau C / n_{\text{C}} \right), \tag{3.4}$$

for exponential temperature programming or

$$T_{\text{A}} = T_0 \left(1.0 - C / n_{\text{C}} \right), \tag{3.5}$$

for linear programming, where T_0 is the initial annealing temperature, τ is the decay constant, for which larger values yield more rapid cooling, C is the current cycle, and n_{C} is

the total number of cycles. Additionally, weighting within the fitness function is optional; if weighting is not requested, all w_{ij} are set to 1.

Fully supported non-bonded force field equations include harmonic bonds and angles (Equation 3.6), and either Charmm- or Ryckaert-Bellemans-type dihedrals (Equations 3.7 and 3.8), respectively [146, 173]:

$$E_{\text{harmonic}} = \frac{k}{2}\left(x_{\text{eq}} - x\right)^2, \tag{3.6}$$

$$E_{\text{Charmm}} = c_0 + \sum_{i=1}^{n} c_i \left[1 + \cos\left(i\phi + \delta\right)\right], \tag{3.7}$$

$$E_{\text{Ryckaert-Bellemans}} = c_0 + \sum_{i=1}^{n} c_i \left[\cos^i\left(\phi - \delta\right)\right], \tag{3.8}$$

where in Equation 3.6, x_{eq} and k correspond to force constants and equilibrium values, respectively. In Equations 3.7 and 3.8, c_i, ϕ, and δ are the Fourier constants, dihedral angle, and phase shift, respectively. Ancillary support is also included for Morse bonds and anharmonic type bonds and angles.

Calculation of QM energies is currently supported for Gaussian09 [174] and NWChem [175] packages, and in the present work, CBMC structure generation and calculation of non-bonded energies is accomplished through the use of MCCCS software [176], however, any tool that can produce .xyz format structure files and can compute molecular mechanics energies can be readily utilized. Decoupling of the annealing portion of the code from the MM energy calculation allows for complete flexibility in the nature of the non-bonded aspect of the molecular mechanics force field, in contrast to other similar semi-automated fitting software [160, 161]. Furthermore, this software has been programmed using object-oriented C++, meaning that extensions to include other bonded potentials forms can be readily accomplished.

The accuracy of parameters generated by the ICBSA method is dependent on the compatibility of the target potential energy surface (PES) and the functional form of the force field, while the efficiency is dependent on the quality of generated structures. The latter point is the primary reason the approach is iterative; if generated structures are poor, the resulting QM energies to which the force field is being fit can correspond to regions of the

PES which are not well captured by the form of the MM force fields, and, if weighting is being used, contribute minimally to the MC acceptance rule. Alternatively, if the generated structures are not distinct enough, features such as dihedral maxima may not be well captured. Somewhat related, choice of weighting function can also influence results. For a harmonic bond or angle, one typically desires to fit preferentially to the minima, which would favor a low weighting temperature. In contrast, one typically desires to capture both minima and maxima for a dihedral potential fit, meaning a relatively higher weighting temperature would be desirable, particularly when large rotational barriers are present. One such way this can be circumvented is by performing each annealing run in several pieces, where one alternates between annealing dihedral potentials with a relatively high weighting temperature, and annealing bond/angle potentials with a lower weighting temperature. This approach will be discussed in more detail in the following section. Additional considerations for annealing efficiency include the allowed size of parameter space and the choice of T_0. To the former point, it is possible that physically unreasonable parameter sets can yield relatively good F values. Thus, by bounding possible parameter choices, one limits the annealed results to values which are most likely to capture realistic system physics. To the latter point, the higher T_0, the more likely that accepted parameters will move far away from the initial guess; thus, if one assumes that the initial parameter guess corresponds to a force field which is not far from the target PES, a high T_0 can be counterproductive.

3.3 Sample Applications

As a first sample application, an ICBSA calculation is run for a simple parameter-recovery test case. Here, an eight-bead chain having two bead types, either end-type (ET) or interior-type (IT) is utilized, resulting in two bond types (ET–IT and IT–IT), two angle types (ET–IT–IT and IT–IT–IT), and two dihedral types (ET–IT–IT–IT and IT–IT–IT–IT). For this application, rather than fitting to a QM-derived PES, a psuedo-PES is created using harmonic bonds and angles, and Charmm type dihedrals. Because the fitting force field equations exactly capture the physics of the target system, this test allows for demonstration of method performance when the MM force field form adequately captures QM physics.

One hundred configurations are generated for the chain molecule through random atom perturbations and dihedral rotations. Only a single iteration is performed, consisting of

Table 3.1: Allowed annealing parameter ranges.

	$k_{\min}$	$k_{\max}$	$x_{\min}$	$x_{\max}$
Bonds[a]	250×10^3	350×10^3	1.52	1.57
Angles[b]	50×10^3	120×10^3	108	117
	n	$c_{\min}$	$c_{\max}$	
Dihedrals[c]	4	-3000	3000	

[a]Values for k are given in K, while x are given in Å.
[b]Values for k are given in K, while x are given in degrees.
[c]Values for c are given in K.

20×10^3 annealing cycles of 5000 MC steps (*i.e.* annealing temperature is decreased after each 5000 MC steps). An exponential cooling schedule is used with an initial temperature of 10 and a τ of 4; weighting is not used. Initial parameter guesses are chosen to be intentionally poor, to demonstrate recovery of generating parameters, and were constrained to the ranges listed in Table 3.1. Figure 3.1 provides the an overview of the ICBSA calculation. Despite the initially poor correlation between the configuration energies produced from the initial guess parameters, and the target energy, the annealing procedure is able to locate parameters which recover the target configuration energies very well. Figure 3.1 also shows a typical feature of simulated annealing approaches; best fitness decreases rapidly in the early stages of the annealing run, but exhibit a decreasingly steep slope as the calculation progresses. This behavior is related to the relative ease of finding parameters which decrease F early in the calculation, when parameters are poor, relative to the later stages, in combination with the decrease in maximum parameter displacement as the calculation progresses.

In the next sample application, parameters are annealed for united-atom linear alkanes. The fitting molecule set is comprised of ethane, propane, butane, and octane, and molecules are described by a combination of CH_3 and CH_2 bead types, yielding 3 unique bond types, and 3 unique angle types. For this fitting procedure, dihedrals are only considered unique on the basis of the central bond, thus the present linear alkane set contains only a single dihedral type. One hundred configurations are generated for each of the fitting molecules by selecting every 100^{th} structure for each compound from a 10,000 step CBMC simulation

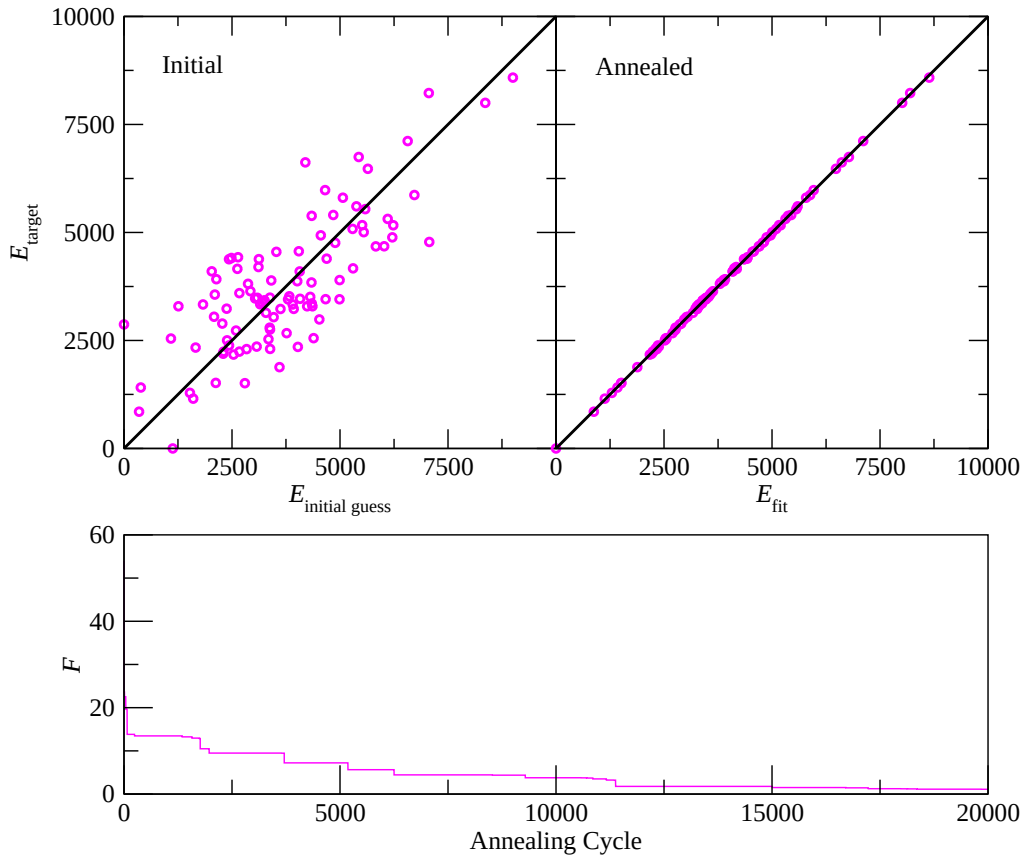

Figure 3.1: Top: comparison of the target energy, E_{target} for each configuration of the eight-bead chain with the energy computed for each configuration based on parameter initial guesses, $E_{\text{initial guess}}$ (left) and the annealed parameters E_{fit} (right). The diagonal line is drawn as a guide to the eye and represents perfect correlation (*i.e.* has an intercept and slope of 0 and 1, respectively). Each data point corresponds to one of the 100 generated structures for the chain molecule. Bottom: Trajectory of the best fitness over the course of the annealing calculation.

on the isolated molecule, at 300 K. The QM energies are taken from single-point energy calculations of each structure computed using the Gaussian09 software suite [174], where hydrogen positions are allowed to optimize, if added to the structure.

The annealing procedure involves 3 full iterations (*i.e.* configuration generation and annealing), where the first iteration uses parameters from the TraPPE-UA force field as an initial guess; for bonds, which are assumed rigid by the TraPPE-UA force field and thus do not contribute to the energy, k is assumed to be 300×10^3 K. Note that parameters for these annealing calculations are also constrained by the values listed in Table 3.1. In any given iteration, 6 annealing calculations are performed, each consisting of 40×10^3 cycles of 5000 steps, and using a τ of 4. The first two calculations in any given iteration use $T_0 = 100$, while the next and last two use $T_0 = 10$ and 1, respectively; odd numbered calculations use $T_\mathrm{w} = 100$ and hold dihedral parameters fixed and even calculations use $T_\mathrm{w} = 1000$ and hold bond/angle parameters fixed. The "best" parameter set resulting from each of the six calculations are used as input for the next calculation. Alternate schemes where T_w were held at an intermediate value and all parameters were allowed to fluctuate were also tested, however alternating between low and high weighting temperatures in the present fashion was found to result in better simultaneous capture of harmonic minima and Fourier series maxima/minima.

Figure 3.2 provides a comparison of the correlation between the QM-derived energy and the total MM energy for the initial guess of the first iteration to the final parameters yielded by the third iteration. As with the previous sample application, annealing leads to much better reproduction of the target (in this case, QM) energies. In contrast to the previous sample calculation, however, the present annealed parameter set predicts energies for each configuration which are less correlated to the QM values, at higher energy. This results from use of weighting, and is intentional; unlike the previous case where the physics of the target and fitting PES's were identical, the molecular mechanics used here do not replicate the full quantum physics of alkane molecules. For example, bonds and angles all exhibit a small amount of anharmonicity, which is not captured by the harmonic fitting functions used here.

To ensure that the improved correlation shown in Figure 3.2 is not due to gratuitous cancellation of errors, Figure 3.3 provides a direct comparison between QM scans of each

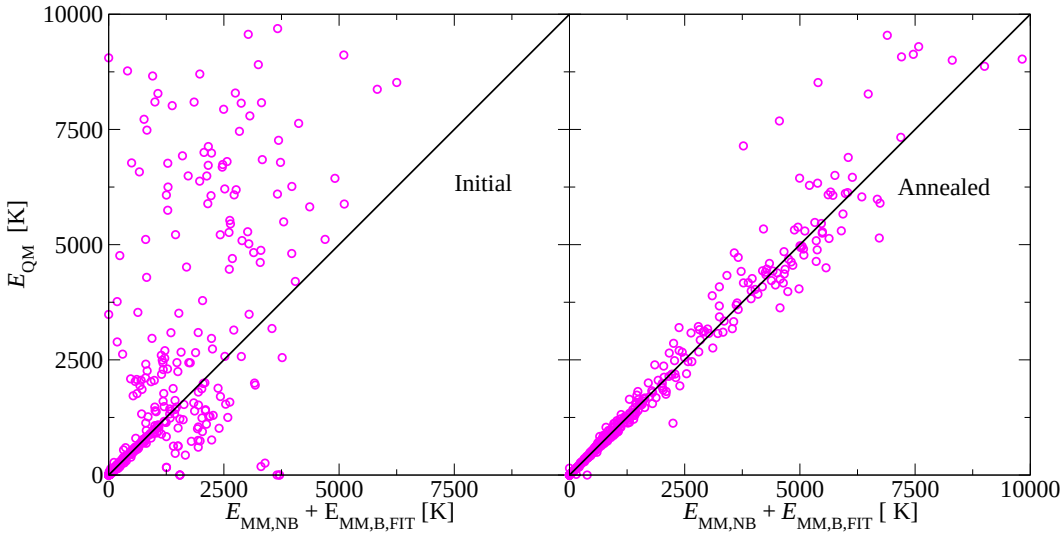

Figure 3.2: Comparison of the QM energy for each configuration of each of the linear alkanes in the fitting set with the energy computed for each configuration based on parameter initial guesses (left), and the energy based on annealed parameters (right). The diagonal line is drawn as a guide to the eye and represents a perfect correlation (*i.e.* has an intercept and slope of 0 and 1, respectively). Each data point corresponds to one of the 100 structures generated for a given molecule in the annealing set.

bond, angle and dihedral type in the present set of fitting molecules, along with the MM energies predicted for the corresponding structure, by the annealed parameters. A few important results emerge from this figure, the first being that the annealed parameters capture the QM PES shape very well. Furthermore, this figure demonstrates that the assumption that dihedrals containing the same central bond are identical works reasonably well for linear alkanes. More importantly, however, the data demonstrate that the ICBSA can be used to fit *shared* parameters across multiple molecules, simultaneously.

As a final example of ICBSA method performance, parameters are annealed for a collection of molecules containing –CH(CH_3)– type branch points, namely isobutane, 2-methylbutane, and 2,3-dimethylbutane. Using the same bond/angle/dihedral description convention as for the linear alkanes, this set of branched molecules has 4 bond types, 4 angle types, and 2 dihedral types, however the CH_2–CH_3 bond type is held fixed at the annealed value from the linear alkanes. Other than the use of 6 iterations (rather than 3),

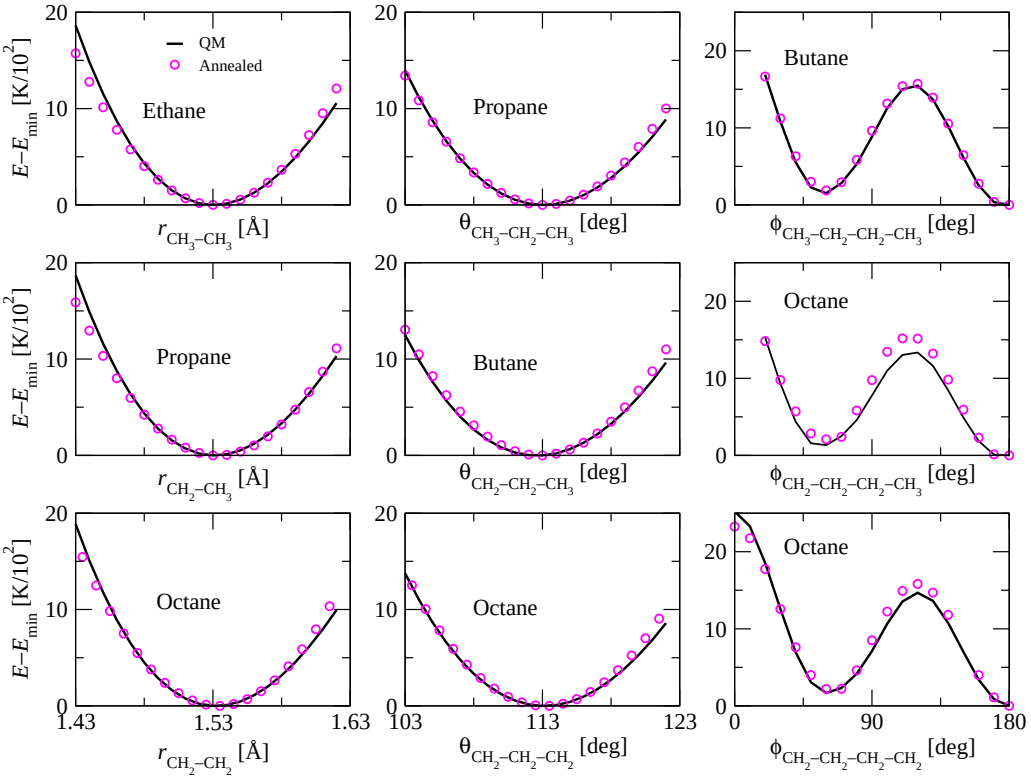

Figure 3.3: Relaxed scans for bonds, angles, and dihedrals in the linear alkane fitting set. The black line provides the QM energy when a given bond, angle, or dihedral is fixed at the value specified by the x-axis, and the rest of the molecule is optimized. The magenta circles are the MM energy predicted for the same exact structure, using annealed parameters.

annealing, calculations are run using the same protocols as the linear alkanes.

Comparisons between the QM and MM energies for each configuration are provided in Figure 3.3. Once again, correlation between QM and MM energies are greatly enhanced by the annealing process. Overall, the correlation is not as high as for the linear alkanes, however, due in part to the increase in anharmonicity of the bonds and angles. Building upon what has been learned from the previously discussed annealing examples, the present results indicate that the ICBSA method can also be applied to more complex structures, and can yield reasonable results even when the form of the force field and the true PES are somewhat mismatched.

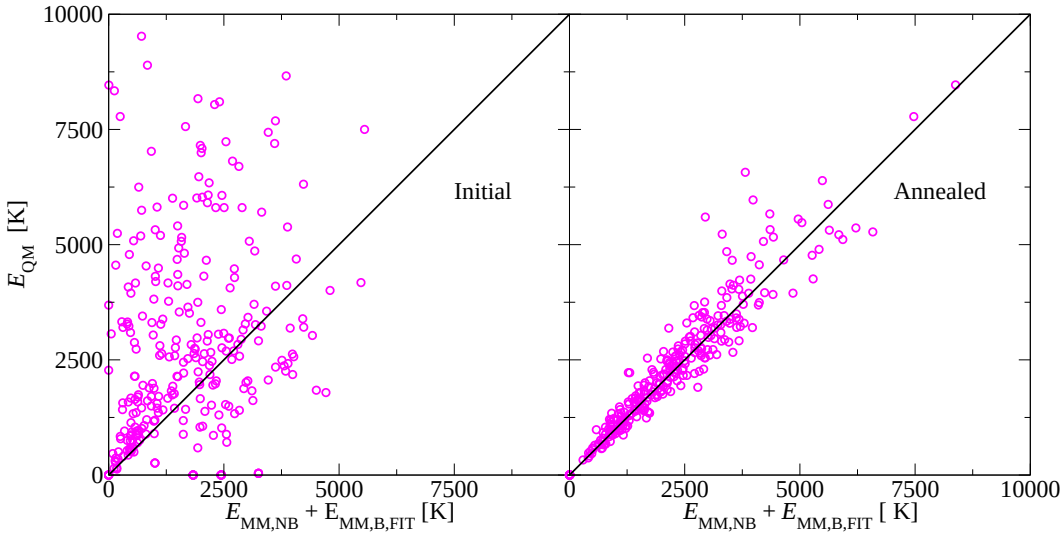

Figure 3.4: Comparison of the QM energy for each configuration of each of the branched alkanes in the fitting set with the energy computed for each configuration based on parameter initial guesses (left), and the energy based on annealed parameters (right).The diagonal line is drawn as a guide to the eye and represents a perfect correlation (*i.e.* has an intercept and slope of 0 and 1, respectively). Each data point corresponds to one of the 100 structures generated for a given molecule in the annealing set.

3.4 Conclusions

A tool has been developed which introduces the ICBSA method for semi-automated fitting of the non-bonded component of molecular mechanics force fields. The present work has demonstrated that the ICBSA method works well when the the physics of the target system and the form of the force field are compatible. In departure from existing methods for automated and semi-automated fitting of the non-bonded portion of force fields, the ICBSA method is able to identify candidate parameters for groups of several molecules at a time, for which parameters may overlap. The tool is designed to be versatile with respect to the form of the non-bonded force field by decoupling the annealing process from calculation of the non-bonded energy. Furthermore, the tool is programmed in an object-oriented fashion, making addition of new bonded force field equations straightforward, and can interface with multiple QM software packages.

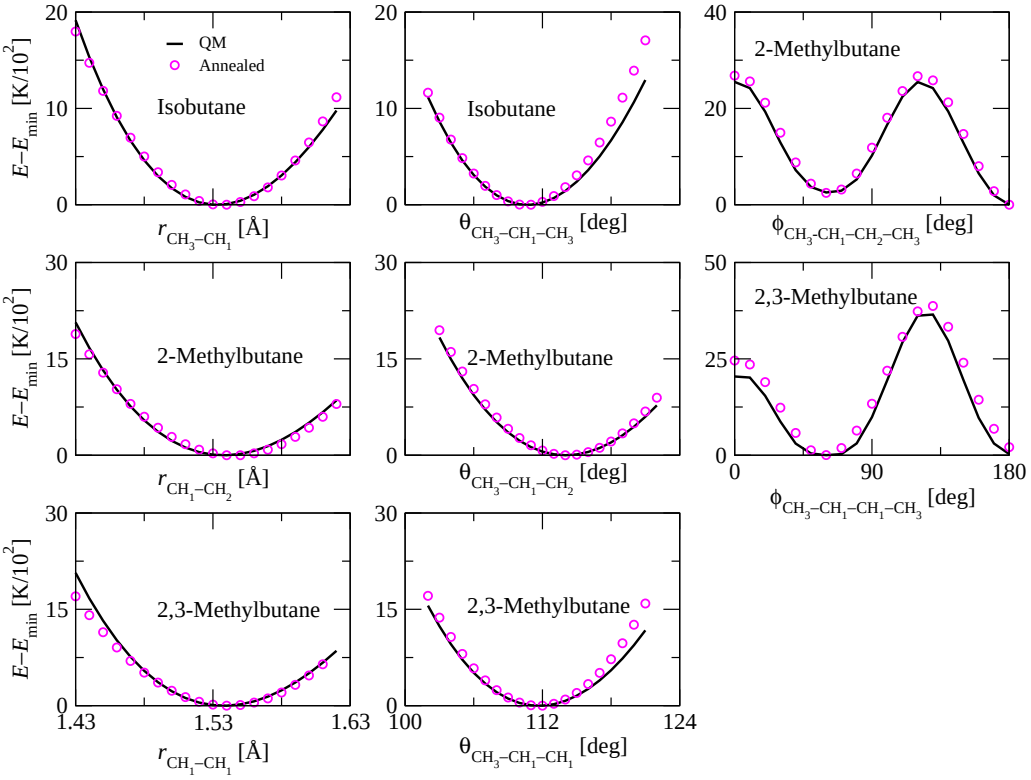

Figure 3.5: Relaxed scans for bonds, angles, and dihedrals in the branched alkane fitting set. The black line provides the QM energy when a given bond, angle, or dihedral is fixed at the value specified by the x-axis, and the rest of the molecule is optimized. The magenta circles are the MM energy predicted for the same exact structure, using annealed parameters.

Chapter 4

Elucidating the Driving Forces in Pressurized Hot Water Reversed-Phase Liquid Chromatography

4.1 Introduction

With the recent push towards greener practices in chemistry [177], separation scientists are looking to replace common organic solvents with safer, environmentally friendly alternatives. Within the field of chromatography, this has manifested as an increased interest in replacing the aqueous-organic solvents used in reversed-phase liquid chromatography (RPLC) with neat, pressurized hot water (PHW), giving rise to the PHW-RPLC mode [178]. In a conventional RPLC regime, hydrocarbon chains (C8–C30), which are grafted to nanoporous silica particles act as a stationary phase and an aqueous organic solvent serves as a mobile phase, resulting in retention times which are typically longer for hydrophobic compounds than those that are hydrophilic. Heating the neat water mobile phase has a few benefits including improved eluting strength (*i.e.*, solutes exit the column faster) and decreased viscosity, however, the PWH-RPLC method is not without difficulties. In a typical RPLC

separation, the hydrophilic character of the mobile phase can be tuned by modifying solvent composition, temperature, and to some extent, pressure. This ability is of particular importance, as solutes can become "stuck" in the stationary phase or to be retained too weakly, if the solvent and stationary phase character differ by too much. While retention can be tuned by adjusting temperature and pressure in PHW-RPLC, the mobile phase only contains a single, highly lipophobic component. Moreover, the effects of varying T and p are not immediately obvious. For example, increasing T can decrease the penalty associated with inserting a solute into a condensed phase (*i.e.* the cavitation entropy) however increasing p has the potential to counter this effect.

Due to the ubiquitous nature of RPLC, numerous studies have been conducted both experimentally and through the use of computer simulations, with the aim of elucidating the mechanistic and thermodynamic factors contributing to retention [1, 48, 179, 180]. The literature on PHW as a mobile phase for RPLC, however, is not nearly as mature or comprehensive as for more common solvents, and generally falls short in one of two ways: either the knowledge gained is not general enough to be applied to other systems, or the insights gained do not provide detailed microscopic explanations for the observed macroscopic and thermodynamic trends. To a large extent, this is due to experimental limitations associated with PHW-RPLC. For example, development of stationary phase materials with the robustness necessary to handle high T and p conditions is still a recent advancement [41, 42, 43, 44, 45, 46, 181, 182, 183, 184, 185], meaning that earlier experimental studies would be limited by phase decomposition. Additional experimental difficulties include solute decomposition, and prohibitively long retention times when run with highly non-polar solutes. While simulations can circumvent these problems, there have been a lack of studies for the PHW-RPLC system, to date.

Nevertheless, many valuable insights into this system have been made. Much of the foundational work in the area of PHW-RPLC was focused on the observation that increasing the temperature of a neat water mobile phase leads to retention which was consistent with experiments run at lower temperature (298–328 K) and with organic modifiers [186, 187]. Coupled with the observation that increasing the temperature of a neat water mobile phase decreases cohesive energy and dielectric constant, researchers arrived to the conclusion that the decreases in the polarity of the solvent and disruption of the H-bonding network

could be achieved by both heating a neat water mobile phase and through introduction of organic modifier [188, 189, 190, 191]. Later studies showed, however, that this was an incomplete picture. For example, it was demonstrated that the T-dependent selectivity changes in PHW-RPLC do not coincide with those observed upon addition of modifier, suggesting that the mechanism is quite different between the two systems [192]. More recent studies [178, 193, 194], which have focused on the retention mechanism in PHW-RPLC, have suggested that changes in H-bonding is the dominant factor affecting retention behavior, and for this reason, the PHW-RPLC mechanism appears more similar to that of RPLC utilizing H-bonding organic modifier (which is more comparable to a neat water solvent than to a non-H-bonding organically modified solvent). It has also been suggested that T-induced increases in stationary phase wetting can decrease retention [193].

The present work distinguishes itself from previous work in the area of PHW-RPLC by leveraging molecular simulations to provide a more fundamental view of the effects governing retention in PHW-RPLC. Because partitioning of solutes into stationary phase chains can contribute significantly to the retention mechanism in RPLC systems [195], it can be useful to employ a simplified bulk partitioning setup, such as hexadecane (C16)/water as approximation of the chromatographic system. In addition to simplifying a considerably complex system, this framework allows researchers to address otherwise obfuscated details, such as whether retention is driven by solvophobic or lipophilic interactions, [11, 12, 196, 197] in other words, whether retention is driven by unfavorable interactions between the solute and the mobile phase, or by favorable interactions of the solute with the stationary phase. Furthermore, a C16/water framework can offer insight into solvent-based extraction and uptake processes. Thus, in addition to simulation data for a PHW-RPLC slit pore, the present work will also include simulations of bulk PHW-C16 systems.

4.2 Simulation Details

4.2.1 System Description and Force Fields

In the present work, simulations are used to study the transfer of n-alkanes (CH_4–C_4H_{10}) and 1-alcohols (CH_3OH–C_4H_9OH) between a neat water phase, an ideal gas reservoir, and one of two retentive phases: either bulk C16 or a chromatographic slit pore, at temperatures

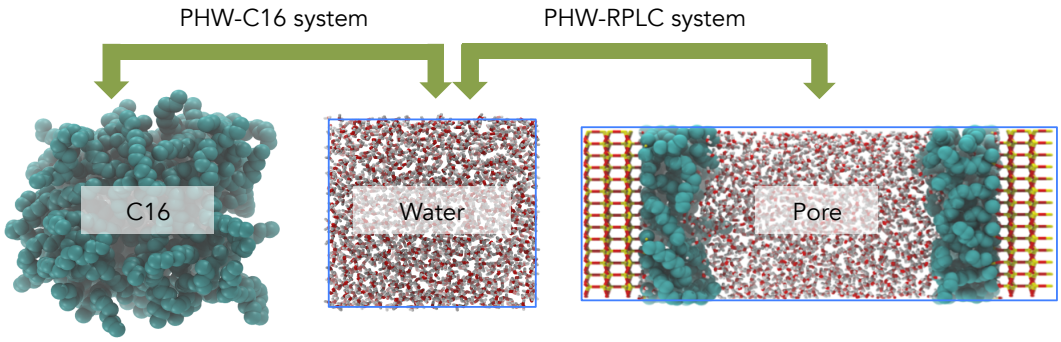

Figure 4.1: Simulation setup for PHW-C16 and PHW-RPLC simulations. The vapor box has been excluded for clarity. Yellow, red, white, and cyan beads correspond to silicon, oxygen, hydrogen and carbon beads, respectively.

ranging from 323 to 473 K and pressures from 30 to 300 atm. Figure 4.1 provides a schematic for the system setup. The pore box is 90 Å in length and contains two roughly 35×35 Å planar substrates, corresponding to the (1,1,1) surface of β-cristobalite, separated by 70 Å along the z-direction. Dimethyl octadecylsilane (ODS) chains are grafted to these surfaces at a density of 2.9 μmol/m^2, leaving residual silanols at a surface density of 4.8 μmol/m^2. One solute of each type is introduced into the system, as is enough water to maintain a mobile phase box of roughly 40^3 Å^3 and fill the regions of slit pore box, which are not occupied by silica or stationary phase chains. Further details of this setup are described previously [198].

The MCCCS-MN software [176] is used to drive the isobaric isothermal Gibbs ensemble [107, 108], coupled-decoupled configurational-bias Monte Carlo simulations [104, 106, 134]. Solutes and water are allowed to transfer between each phase, but hexadecane molecules are confined to their respective box due to their low solubility in water. The chromatographic pore stationary phase silica and grafted chains and similarly confined to the pore box. Particle transfer *via* swap and identity exchange moves [66, 131, 132] between the three phases ensures chemical equilibrium, while volume fluctuations (in all but the slit pore) and particle translations/rotations/conformation changes allow for mechanical and thermal equilibrium, respectively. Volume moves are attempted with a probability $2/N$, where N is the number of molecules in the system excluding silica atoms, resulting in ≈ 1

accepted volume move per cycle, where 1 cycle is equivalent to N Monte Carlo moves. Volume, particle swap and identity exchange moves have a cumulative probability of 0.3, where swaps and identity exchanges occur with equal probability, and the remaining 0.7 of moves are distributed among configurational-bias, translation, and rotation moves, according to the degrees of freedom in the system. Stationary phase chains, which are tethered to the silica surface, undergo conformational changes with the additional help of the self-adapting fixed-endpoint (SAFE) CBMC algorithm, which allows for regrowth of interior segments of a chain [135]. Direct particle transfers between two condensed phases are not allowed, due to the low acceptance rates associated with moving a solute between two dense phases - instead, the ideal gas (IG) reservior serves as a transfer medium. Additionally, a uniform biasing potential is applied to each solute in each condensed phase box to ensure sufficient sampling of all three phases, while allowing no more than two solute molecules in the retentive phase, on average [66], and to ensure that predicted solute retention is not influenced by the presence of other solutes.

The TraPPE-UA [106, 137, 141], and TIP4P water [199] force fields are used in conjunction to describe solute, mobile phase, and stationary phase chain interactions, while a zeolite force field [200, 201] is used to describe silica atoms. All simulations use a cutoff distance of 10 Å and the Ewald summation technique is used to compute long-ranged electrostatics, with a screening parameter of $\kappa = 3.2/r_{\text{cut}}$ and $K_{\text{max}} = \text{int}(\kappa \text{L}_{\text{box}} + 1)$ for the upper bound of the reciprocal space summation [127]. Eight/sixteen independent simulations were run for each condition in the PHW-C16/PHW-RPLC systems, consisting of at least 100,000 cycles of equilibration and 500,000 cycles of production.

Thermal expansivities and isothermal compressibilities are also computed for neat C16 and PHW phases. These values are obtained by running CBMC simulations in the isobaric-isothermal ensemble for an 80 molecule neat C16 system, and a 1500 molecule neat water system. Thermal expansivities are computed by running simulations at $p = 30$ atm and $T =$ 323, 373, 423, and 473 K, while isothermal compressibilities were computed from simulations run at $T = 423$ K and $p = 30$, 100, and 300 atm. Derivatives were evaluated through finite differences. Equilibration and production were run for 50,000 and 100,000 cycles, respectively. Cumulative move type probabilities were distributed as (volume, CBMC, translation, rotation) = (0.1, 0.4, 0.7, 1.0) and (0.0003, 0.0, 0.5, 1.0) for the C16 and PHW

phases, respectively, where 20% of the CBMC moves attempted for the C16 phase were executed using the SAFE-CBMC algorithm, and eight independent simulations were run at each condition. All errors provided in this work correspond to the 95% confidence interval, taken across independent simulations.

4.2.2 Data Interpretation and Analysis

One of the primary objectives of this work is to compute the retention free energies, ΔG_{ret} of solutes between retentive stationary and mobile phases, "RP" and "MP" respectively. Obtaining these values for the bulk partitioning system is straight forward, and can be accomplished *via*:

$$\Delta G_{\text{ret}} = -RT \ln \left(\frac{n_{i,\text{RP}} V_{\text{MP}}}{n_{i,\text{MP}} V_{\text{RP}}} \right) = -RT \ln(k' \phi), \tag{4.1}$$

where R, T, and $n_{i,y}$ are the gas constant, temperature, and number of molecules of species i in phase y. The terms k' and ϕ are the capacity factor and phase ratio, and are equivalent to $n_{i,\text{RP}}/n_{i,\text{MP}}$ and $V_{\text{MP}}/V_{\text{RP}}$, respectively.

The phase ratio is an easily defined quantity for the PHW-C16 system due to the distinct and homogeneous nature of each phase, however, the PHW-RPLC pore contains regions which correspond to both the RP and MP, as is shown in Figure 4.2. Thus, one must take structure into consideration when attempting determine ϕ. Experimentally, ϕ is an exceedingly difficult quantity to determine, so instead, studies focus on measurement of *incremental* free energies ($\Delta\Delta G$), which correspond to the change in free energy that results from addition of a given moiety. In other words, the incremental retention free energy between species i and j can be written as:

$$\Delta\Delta G_{\text{ret},j\to i} = -RT \ln(k_i' \phi) - RT \ln(k_j' \phi), \tag{4.2}$$

where the phase ratios cancel out.

In the present work, $\Delta\Delta G$ values are computed through the following: first a $\rho(z)$ profile, which has been symmetrized across the pore, is constructed for the solute of interest, i. The $\rho(z)$ profile will approach the number density of the bulk MP at the center of the pore, indicating the presence of bulk-like solvent. If integrated in its current state, the resulting $n_{i,\text{RP}}$ will also contain solutes belonging to the mobile phase, such that, if the pore box

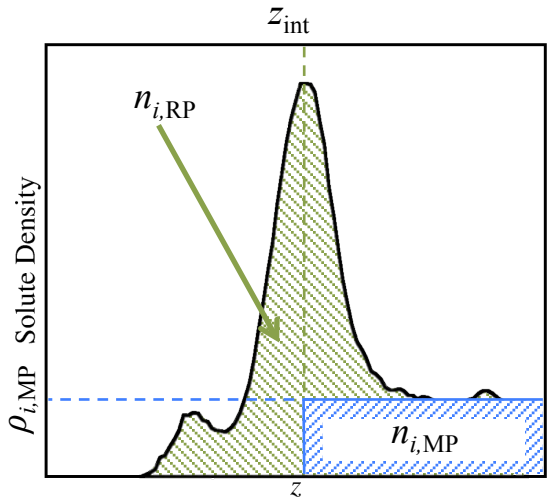

Figure 4.2: Schematic of how $n_{i,\text{RP}}$ and $n_{i,\text{MP}}$ are defined for use in computation of incremental free energies. The black line corresponds to the symmetrized density profile for a solute, i across the slit pore. The blue shaded region defines the mobile phase portion of the slit pore, while the green shaded region corresponds to the stationary phase portion. This figure has been adapted from a presentation prepared by Jake L. Rafferty.

length was extended towards infinity, so would the resulting $n_{i,\text{RP}}$, despite the additional solute density corresponding to the mobile phase. To account for this effect, we define the mobile phase region of the slit pore box as the portion which extends from the retentive phase/mobile phase interface to the center of the box.

The number of solutes in the retentive phase is then defined as:

$$n_{i,\text{RP}} = \int_0^{0.5z_{\text{max}}} \rho(z)dz - \int_{z_{\text{int}}}^{0.5z_{\text{max}}} dz, \tag{4.3}$$

where the first integral goes from the silica surface to the middle of the pore. The second integral goes from the retentive phase/mobile phase interface to the center of the box and gives $n_{i,\text{MP}}$. Note that in the present work, the location of the interface is taken to be the midpoint of the 10/90% region of the symmetrized solvent density profile within the slit pore and incremental free energies, such as $\Delta\Delta G_{\text{CH}_x}$ and $\Delta\Delta G_{\text{OH}}$, correspond to the change in retention free energy a solute experiences upon the addition of a CH_x or OH group, respectively.

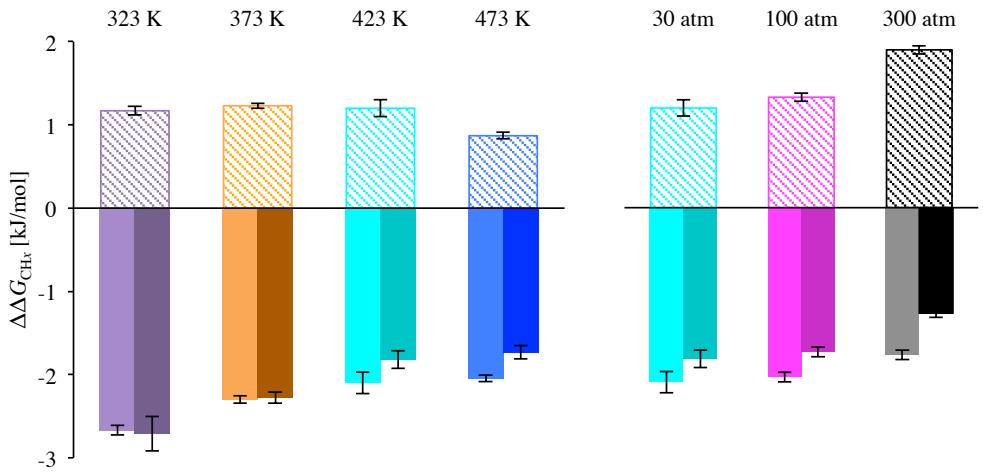

Figure 4.3: Incremental transfer free energy for a methylene segment. For any given temperature, the textured wide column, solid light column, and solid dark column corresponds to VAP→WAT, VAP→C16, and VAP→ODS transitions, respectively.

4.3 Results and Discussion

4.3.1 Alkane Partitioning and Retention

Separation scientists often use the concept of solvophobicity and lipophilicity to understand the driving forces behind solute retention [11, 12, 13, 14, 34, 197]. In this framework, if one assumes a polar MP, sufficiently apolar solutes will be driven out of the mobile phase, causing the solute to be retained by the RP. Along similar lines, if a solute has a sufficiently high affinity for the RP, it will be pulled into the stationary phase, driving a higher retention. In this way, solvophobicity and lipophilicity can be rationalized as the tendency for a solute to be either "pushed" from the mobile phase or "pulled" into the stationary phase, respectively. These two effects can be evaluated by studying the incremental free energy of a methylene unit in the system, $\Delta\Delta G_{\mathrm{CH}x,\mathrm{ret}}$. Assuming the process of transferring a solute from the MP to RP can be broken into two components, VAP→MP and VAP→RP, a positive $\Delta\Delta G_{\mathrm{CH}x,\mathrm{VAP}\rightarrow\mathrm{MP}}$ is indicative of a solvophobic driving force, whereas a negative $\Delta\Delta G_{\mathrm{CH}x,\mathrm{VAP}\rightarrow\mathrm{SP}}$ indicates a lipophilic driving force.

Figure 4.3 provides the incremental free energies for transfer of a CH_x unit. Where

comparisons are available, data are in good agreement with experiment and previous simulations. For example, experimental measurements of $\Delta\Delta G_{\mathrm{CH}x,\mathrm{ret}}$ and $\Delta\Delta G_{\mathrm{CH}x,\mathrm{VAP}\rightarrow\mathrm{C16}}$ for a WAT-C16 system at conventional RPLC conditions have yielded values of -3.84 and -2.70 kJ/mol, respectively [197], while previous simulations have computed respectively values of -3.76_8 and -2.71_1 kJ/mol [196], in excellent agreement with the values of -3.84_3 and -2.67_6 found in the present work. Furthermore, measurements for a PHW-RPLC system utilizing an ODS stationary phase found $\Delta\Delta G_{\mathrm{CH}x,\mathrm{ret}}$ to be -2.7 ± 1.3 kJ/mol for $T = 453 - 468$ K [178], compared to the present value of -2.54_7, for $T = 473$ K, while simulation studies for the WAT-RPLC system at conventional conditions have yielded $\Delta\Delta G_{\mathrm{CH}x,\mathrm{ret}} = -3.6$ kJ/mol [78] compared to -3.9_3 in the present work. Note that errors for previously published work have been included where available.

Focusing on data as a function of temperature, it is apparent that all transfers from the ideal gas to mobile phases result in positive free energies, meaning the CH_x units are pushed out of the mobile phase by solvophobic driving forces. At the same time, transfers from the ideal gas to stationary phases remain negative, indicating that CH_x units are simultaneously pulled into the stationary phase through lipophilic forces. As temperature is increased, the strength of solvophobic and lipophilic forces are decreased, resulting in incremental retention free energies that are less favorable. Increases in $\Delta\Delta G_{\mathrm{CH}_x}$, such as those seen here, are a desirable trait, and indicate prevention of the overly-long retention times one would expect for alkanes when a neat water mobile phase is used. Quantitatively, results are found to be comparable between the PHW-C16 and PHW-RPC systems at all but the highest temperature. The solvophobic view of retention in chromatographic systems [34] was developed based on the observation that solute retention involving RPLC retentive is highly sensitive to mobile phase phase composition. The present work indicates changes in lipophilicity are relatively larger than solvophobicity as temperature is increased, in agreement with experiment for similar systems [178], and with a more recent description of retention, which also accounts for the role of changes in the stationary phase character [11].

As was mentioned in the introduction, alteration of operation temperature in PHW-RPLC is used to tune the separation whereas mobile phase concentration would be used in conventional RPLC. Changes in mobile phase composition in conventional RPLC generally has little effect on $\Delta\Delta G_{\mathrm{CH}x,\mathrm{VAP}\rightarrow\mathrm{RP}}$, however changes in $\Delta\Delta G_{\mathrm{CH}x,\mathrm{VAP}\rightarrow\mathrm{MP}}$ when adding

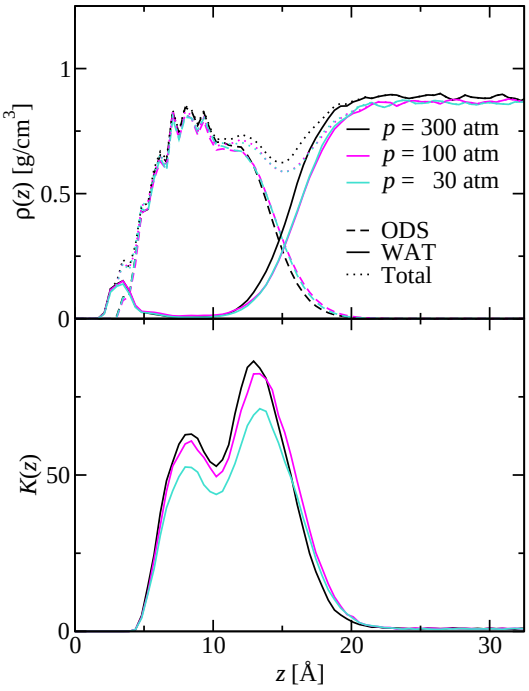

Figure 4.4: PHW-RPLC system density profile (top) and propane $K(z)$ (bottom) at 423 K. A z-coordinate of zero corresponds to the silica surface.

30% molfraction organic modifier (*i.e.*, methanol or acetonitrile) to a neat water mobile phase is enough to change the retention mechanism from a combination of solvophobic and lipophilic driving forces, as is observed for PHW-RPLC, to purely lipophilic [78]. In contrast, changing temperature in PHW-RPLC has a smaller effect on $\Delta\Delta G_{\mathrm{CH}x,\mathrm{VAP}\rightarrow\mathrm{MP}}$, but also significantly influences $\Delta\Delta G_{\mathrm{CH}x,\mathrm{VAP}\rightarrow\mathrm{RP}}$. Thus, despite such different driving forces, increasing the temperature in a PHW-RPLC separation from 323 to 473 K is found to be roughly equivalent [78] to adding 15% molfraction organic modifier to a neat water RPLC system, changing the retention free energy by $1-1.5$ kJ/mol. Experimental studies [193] yielded consistent results when using a polystyrene-coated zirconia phase, reporting $\Delta\Delta G_{\mathrm{CH}x,\mathrm{ret}}$ values which were equivalent within error for a neat water mobile phase at 373 K and a $\approx 20\%$ methanol/water mobile phase at 303 K.

Changing pressure leads to different and more subtle trends; generally, higher pressure

results in larger solvophobicity, but also in decreased lipophilicity. For the bulk system, changes are minimal when moving from $p = 30 - 100$ atm, and become more noticeable at 300 atm. Increases in solvophobicity are greater than decreases in lipophilicity, leading to slightly more favorable $\Delta\Delta G_{\mathrm{CH}x,\mathrm{ret}}$ values. The slit pore system exhibits somewhat different trends. Again, only small changes are observed when moving from $p = 30 - 100$ atm, however lipophilicity becomes noticeably smaller at 300 atm while solvophobicity becomes larger. Overall, changes in lipophilicity and solvophobicity balance one another leading to $\Delta\Delta G_{\mathrm{CH}x,\mathrm{ret}}$ for the PHW-RPLC system which are relatively unchanged as pressure is increased. The sudden changes in lipophilicity and solvophobicity observed at 300 atm is consistent with the density profiles at the top of Figure 4.4, which show that system structure at the highest pressure has ODS chains which are slightly more compressed, and exhibits slightly increased interfacial wetting. The change in system structure can be understood in terms of isothermal compresibility:

$$\beta_{\mathrm{T}} = \frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_{\mathrm{T}}. \tag{4.4}$$

Water is found to be incompressible, exhibiting values of 0.096_9 and 0.10_3 atm$^{-1} \times 10^3$ at $p = 200$ and 65 atm, respectively, while C16 changes from 0.25_1 to 0.32_4 atm$^{-1} \times 10^3$. As a result, increased pressure allows water to push on the ODS chains, leading to the observed structural changes. The observed increase in ODS compression works in tandem with increased interfacial wetting to squeeze the solute out of the ODS chains. Previous studies [71, 202, 203] of pressure effects in conventional RPLC systems using aqueous organic and purely organic mobile phases find an opposite, albeit subtle trend, where $\Delta\Delta G_{\mathrm{CH}x,\mathrm{ret}}$ is decreased as pressure is increased. This result is somewhat unsurprising, as the mobile phases used in those studies are more compressible than the neat water phase used in the present work.

Previous work [178, 193, 204] describes an inversion in solute shape selectivity when moving from traditional RPLC to PHW-RPLC conditions (*i.e.* less conformationally constrained molecules exit the column last in PHW-RPLC), which was suggested to indicate a decrease in conformational order within the stationary phase, with the rationale that a less conformationally constrained molecule has the ability to rearrange itself in order to maximize favorable contact with the stationary phase, whereas a more rigid structure cannot. In the present work, system structure and ordering were directly evaluated from the

Table 4.1: Structural properties of ODS chains in PHW-RPLC systems.

T [K]	p [atm]	w_{int} [Å]	$\cos\theta_{\text{ete}}$	r_{ete}[Å]	f_{gauche}	$\langle S \rangle$
473	30	7.1_2	0.57_1	14.9_1	0.35_1	0.03_4
423	30	5.6_1	0.52_1	15.0_1	0.33_1	0.01_3
373	30	4.8_1	0.47_4	15.2_2	0.30_1	-0.04_3
323	30	3.9_4	0.44_1	15.7_1	0.28_1	-0.07_2
423	300	6.0_1	0.50_4	15.0_1	0.33_1	-0.02_2
423	100	6.0_2	0.51_3	14.9_1	0.33_1	-0.02_3

simulation data by examining the fraction of gauche defects (f_{gauche}), the chain end-to-end distance (r_{ete}), and the overall and local chain order ($\langle S \rangle$ and S_{1-3}, respectively) as defined by equation 7.1:

$$S_{1-3} = \frac{1}{2} \left\langle 3\cos^2\theta - 1 \right\rangle, \tag{4.5}$$

where θ is the angle between the silica surface normal and a given 1–3 vector along the chain. An S of 1 and -0.5 corresponds to chains which are perpendicular and parallel to the surface, respectively, while a value of zero indicates no preferential alignment. The overall order parameter, $\langle S \rangle$, is taken as the average of all S_{1-3} in the chain. Table 4.1 and Figure 4.5 demonstrate that, consistent with Allmon *et al.*'s suggestion [193], the RP chains become more disordered, as we see increased f_{gauche} and S_{1-3} which approach zero as temperature is increased, particularly at the terminus. The simulations also indicate that chains are slightly less extended (r_{ete}), but overall, more perpendicular with respect to the silica surface ($\langle S \rangle$ and $\cos\theta_{\text{ete}}$). Changing pressure has little effect on the aforementioned structural properties, with the exception of an increase in interfacial width from 5.6_1 to 6.0_2 and chains which are slightly more parallel to the silica surface, when moving from 30 to 100 atm.

As mentioned in the introduction, partitioning of solutes between a bulk C16 and mobile phase is often used as a simplified framework for describing the retention mechanism in chromatographic systems which posess structure and heterogeneity more akin to the chromatographic pore studied here. The utility of this approximation depends on whether the solute partitions into the stationary phase chains of the chromatographic system, or adsorbs

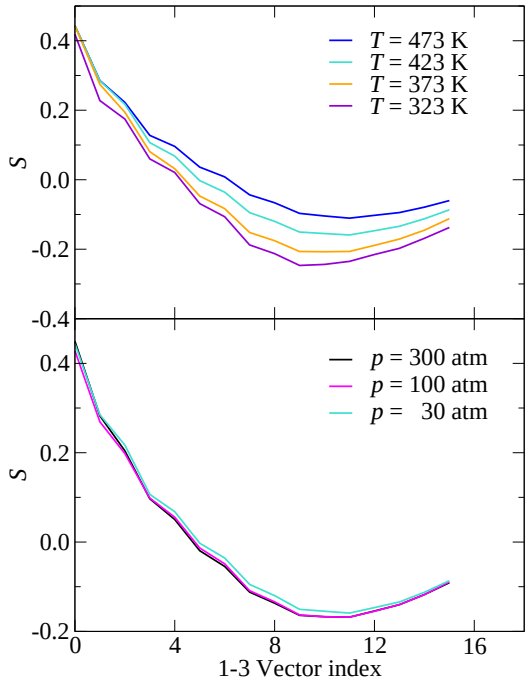

Figure 4.5: Order parameter along the chain backbone in PHW-RPLC systems at 30 atm with varied temperature (top) and 423 K with varied pressure (bottom). The index 0 is used for the 1-3 vector closest to the silicon atom of the stationary phase chain.

to the stationary phase/mobile phase interface. Density profiles for the slit pore systems with varied temperature are shown in Figure 4.6, as are the z-dependent distribution coefficients, $K(z) = \rho(z)_{\text{pore}}/\rho_{\text{WAT}}$, for a propane molecule. It is immediately apparent that the distribution of alkane solutes within these systems is bimodal, consistent with previous simulations studies with conventional RPLC solvents and conditions [78]. One peak sites well within the tethered ODS chains, where solutes experience a solvation environment similar to that of the bulk C16 partitioning system, while the other peak is located at the interface. The unintuitive tendency for alkanes to adsorb at the interface given the nearby ODS chains is partially driven by the solvent depletion at the interface (*i.e.* the dip in total density at $z \approx 15$ Å), which decreases the cavitation entropy penalty at that location. Interfacial width, which is defined as the thickness of the 10–90% portion of the water

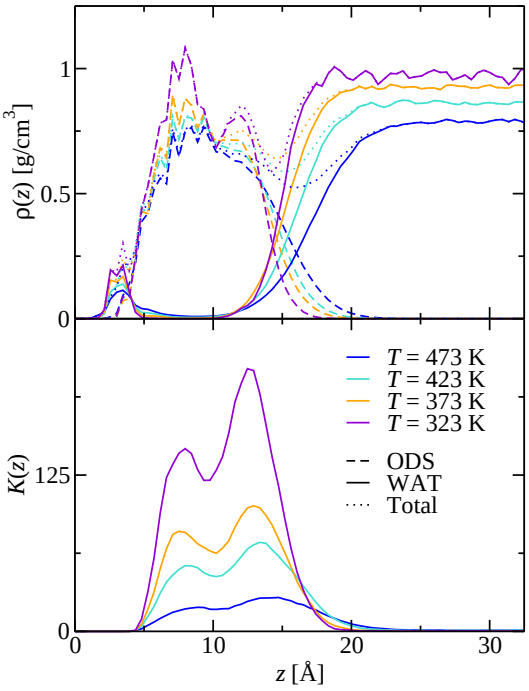

Figure 4.6: PHW-RPLC system density profile (top) and propane $K(z)$ (bottom) at 30 atm. A z-coordinate of zero corresponds to the silica surface.

density profile is given in Table 4.1, and is shown to increase with temperature. Height of the $K(z)$ profiles is shown to decrease as temperature increases, indicating an overall less favorable ΔG_{ret}.

Figure 4.7 provides the retention free energies, ΔG_{ret}, for alkanes in the PHW-C16 systems at different temperatures and pressures, where more negative values of ΔG_{ret} correspond to more strongly retained solutes. For any given n_{CH_x}, retention becomes less favorable when moving from $T = 373$ to 473 K, however ΔG_{ret}'s for $T = 323$ K fall roughly between those for the next two highest temperatures. Figure 4.8 demonstrates this trend more clearly by showing that the retention free energy of propane as a function of temperature goes through a minimum, leading to nearly overlapping free energies at the three lowest temperatures.

The bottom panel of Figure 4.8 shows that the observed retention behavior coincides

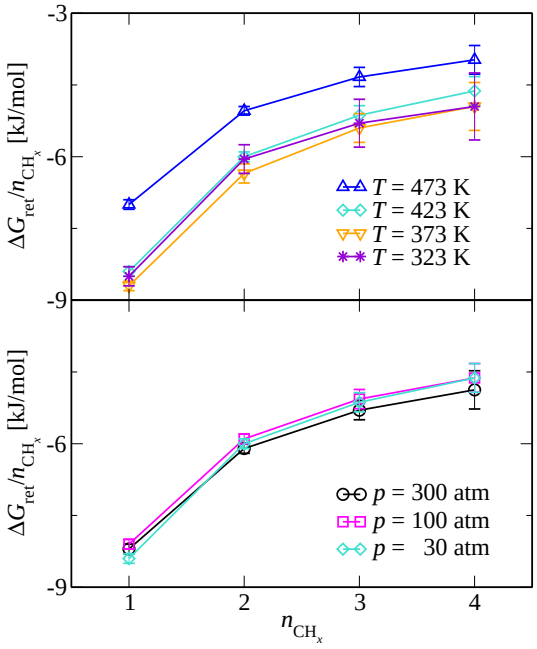

Figure 4.7: Retention free energies for alkanes in a PHW-C16 system as a function of temperature, at 30 atm (top) and pressure, at 423 K (bottom).

with changes in the thermal expansivity of each phase, α_{V}, defined as:

$$\alpha_{\mathrm{V}} = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_{\mathrm{p}}, \tag{4.6}$$

where V and p are the volume and pressure, respectively. This quantity relates to the change in volume of a phase upon changing temperature, where larger values correspond to a phase which undergoes greater expansion. Figure 4.8 shows that initially α_{V} is greater for the bulk C16 phase, however, as temperature is increased, α_{V} grows more quickly for the water phase, driven by the critical temperature of water being less than that of C16. A possible explanation for the observed ΔG_{ret} trend relates to the formation of hydrogen bonded "cages" around non-polar solute molecules [205, 206, 207]. This cage formation has an associated negative enthalpy which offsets the positive enthalpy of mixing the non-polar species with water. As temperature is increased, however, this effect plays a lesser role, as entropy becomes more important. Figure 4.8 also provides ΔG_{ret} for alkanes as pressure is increased, and shows no appreciable differences when moving from $p = 30$ to 100 atm,

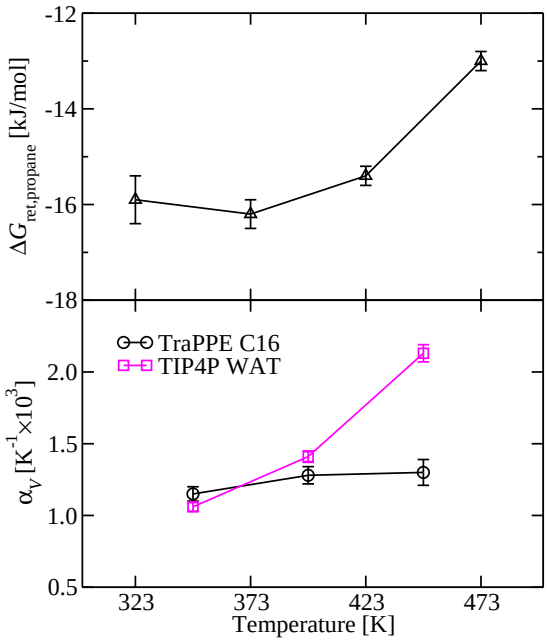

Figure 4.8: Top: propane retention free energy as a function of temperature for a PHW-RPLC system at 30 atm. Bottom: thermal expansivities of TraPPE C16 and TIP4P water.

but slightly more favorable ΔG_{ret}'s when moving to 300 atm, consistent with solubility measurements for saturated fatty acids in water at similar conditions [206].

4.3.2 Alcohol Partitioning and Retention

Retention trends are inverted for alcohols in the PHW-C16 system; as is shown in Figure 4.9, increasing temperature leads to more favorable retention free energies. This increase in retention can be understood in terms of H-bonding. In the present work, the criteria for a H-bond is an O$\cdots$O distance $\leq$ 3.3 Å, an O$\cdots$H distance $\leq$ 2.5 Å, and an O$\cdots$H$-$O angle cosine ≤ -0.1. Table 4.2 indicates that, as T increases, the number of water molecules in the C16 phase increases significantly, concomitantly increasing the number of H-bonds to propanol within the phase. Increases in T also decrease H-bonding in the neat water phase, further contributing to increased retention. The highest number of H-bonds to propanol in C16, however, are observed at the lowest temperature, when there are the least water molecules in the C16 phase. Increasing pressure is shown to decrease the amount of water in

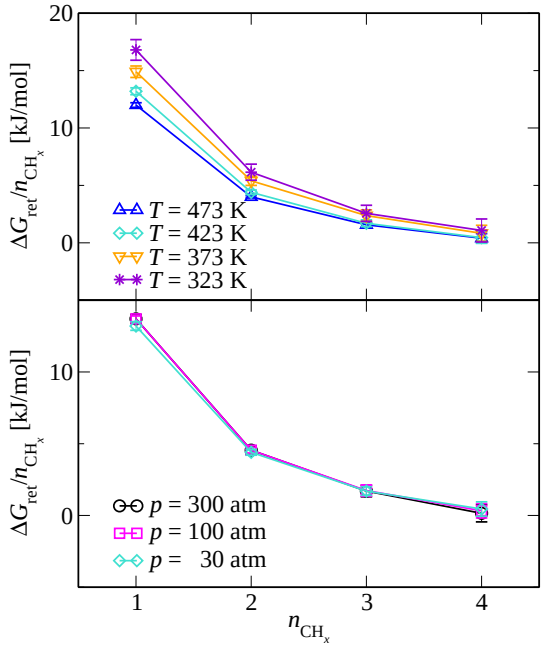

Figure 4.9: Retention free energies for alcohols in a PHW-C16 system as a function of temperature, at 30 atm (top) and pressure, at 423 K (bottom).

C16 by only a small amount, however the number of H-bonds between water and propanol are not significantly influenced. H-bonding in the water phase is increased with pressure, though also by a very small amount. Figure 4.9 demonstrates that, as a result, retention free energies for the alcohols in the C16 system are not significantly influenced by increased pressure.

Figure 4.10 provides the $K(z)$ profiles for alcohols in the PHW-RPLC system, with varied T. In departure from what was observed for alkanes, only a single large peak emerges, at the retentive phase/mobile phase interface, indicating that retention is almost entirely determined by adsorption to the interface. This result is consistent with previous simulation studies conductcted for RPLC systems at conventional temperatures, and using aqueous organic solvents [78]. The small peak near $z = 5$ Å corresponds to hydrogen bond formation with residual surface silanols, and decreases with increased temperature. Analysis of the incremental retention free energies, provided in Figure 4.11, show that, unlike alkanes, retention trends differ significantly between the C16 and slit pore systems. Pressure was

Table 4.2: Hydrogen bonding in bulk water and hexadecane[a], and number of water molecules in hexadecane.

		$n_{\mathrm{H-bond}}$			
		WAT/WAT	WAT/PrOH	WAT/PrOH	n_{WAT}
T [K]	p [atm]	in WAT	in WAT	in C16	
473	30	2.99_1	1.91_1	0.19_1	5.7_1
423	30	3.31_1	2.07_2	0.12_1	1.55_3
373	30	3.54_1	2.27_1	0.13_3	0.31_1
323	30	3.74_1	2.35_1	0.30_8	0.037_2
423	300	3.38_1	2.13_2	0.14_2	1.19_3
423	100	3.35_1	2.12_2	0.13_1	1.44_5

[a] The notation "A/B in C" denotes H-bonds between A and B in phase C, per molecule of B.

not found to have a significant influence on propanol $K(z)$ profiles, thus they are not shown. For all T's, however, $\Delta\Delta G_{\mathrm{OH,VAP\to MP}}$ is negative, but becomes less favorable as temperature is increased due to decreased H-bonding. The changes in $\Delta\Delta G_{\mathrm{OH,VAP\to MP}}$ observed when moving from $T = 323$ to 473 are comparable to the changes observed when replacing a neat water solvent with pure methanol, in a conventional RPLC system [78].

Incremental transfer free energies from vapor to the PHW-RPLC pore stationary phase are also negative for all temperatures, though, as was observed for $\Delta\Delta G_{\mathrm{OH,VAP\to MP}}$, values become less favorable as temperature is increased. Nevertheless, changes for $\Delta\Delta G_{\mathrm{OH,VAP\to RP}}$ are larger than for $\Delta\Delta G_{\mathrm{OH,VAP\to MP}}$, resulting in $\Delta\Delta G_{\mathrm{OH,ret}}$ for the slit pore system which are slightly less favorable with increased temperature. Though the interfacial structural changes shown in Figure 4.10 and listed in Table 4.1 suggest increasing temperature leads to an adsorption site which becomes increasingly entropically favorable (*i.e.* an interfacial region that widens and exhibits greater de-wetting), a likely explanation for the overall decreasing $\Delta\Delta G_{\mathrm{OH}}$ is the commensurately decreases in H-bonding indicated in Table 4.2.

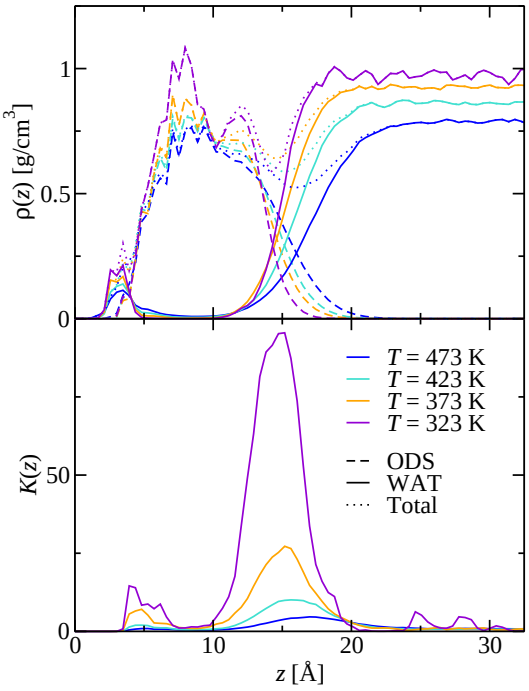

Figure 4.10: PHW-RPLC system density profile (top) and propanol $K(z)$ (bottom) at 30 atm. A z-coordinate of zero corresponds to the silica surface.

4.4 Conclusions

Simulations have been used to investigate the driving forces in PHW-RPLC. Using a fundamental model (*i.e.*, n-alkanes and 1-alcohols in an PHW-RPLC pore) we find trends in incremental retention free energy that are consistent with experiment, and are able to provide a molecular-level view into the system. We find that alkane retention is governed by a combination of partitioning and adsorption whereas the latter governs alcohol retention. Despite differences in retention mechanism (*i.e.* purely partitioning for PHW-C16 system compared to a combination of partitioning and adsorption in the PHW-RPLC pore), retention trends are similar for alkanes with increasing T in the two systems. In contrast, application of pressure results in opposing trends in the two systems, due to the heterogeneity of the PHW-RPLC pore. Retention for alcohols between the two systems was found to

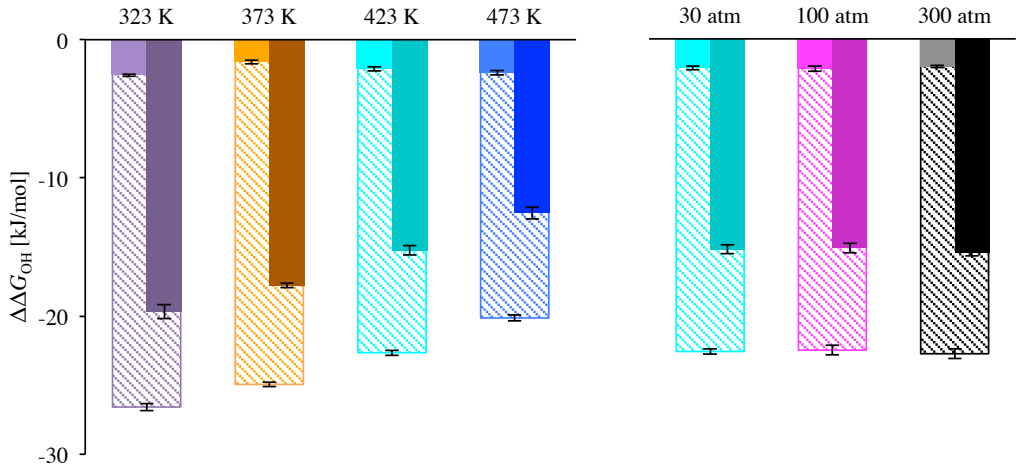

Figure 4.11: Incremental transfer free energy for a hydroxyl segment. For any given temperature, the textured wide column, solid light column, and solid dark column corresponds to VAP→WAT, VAP→C16, and VAP→ODS transitions, respectively.

be substantially different, where in the C16 system, increasing temperature resulted in more favorable free energies due to the enhanced presence of solubilized water in the hexadecane phase, whereas retention decreased in the PHW-RPLC system. Pressure was not found to have a significant effect on $\Delta\Delta G_{\mathrm{OH}}$ for either system.

More generally, we find that, although both solvophobic and lipophilic forces are present in the PHW-RPLC system, decreases in lipophilicity have a larger influence on the overall incremental retention free energy, as temperature is increased. Our results, which are consistent with those of Allmon *et al.* [193], suggest that a less bimodal separation utilizing a neat water mobile phase can be achieved by *i.* increasing pore wetting, which would discourage alkane adsorption, and *ii.* decreasing the H-bonding capability of the mobile phase to further decrease alkane retention, and encourage alcohol retention. These changes could be achieved by introducing ligands with embedded polar groups, which would have the added benefit of increasing alcohol retention by enhancing solvent penetration in the ODS chains, and through addition of an environmentally friendly salt to the mobile phase to suppress H-bonding [193], however temperature robustness of the retentive phase in addition to salt corrosivity would need to be taken into consideration.

Chapter 5

A New Approach to Column Pair Screening for 2DLC

5.1 Introduction

The search for faster and higher resolution separation methods has led to a vast variety of different technologies for column chromatography utilizing a single column. Innovations in column technologies, such as superficially porous particles [184, 43] and sub-2μm particles [42, 44, 208] increase the efficiency and speed of single-dimension liquid phase separations. These advances enable impressive separation of many samples typically encountered in drug development, clinical chemistry, environmental and food monitoring [209, 210, 211, 212]. However, a single column is not always sufficient for complex separation resolution, as is the case in fields such as proteomics and metabolomics where the number of components may exceed one thousand. These samples require a separation technique with a highly enhanced peak capacity.

One method of implementing highly complex separations is to use two-dimensional liquid chromatography (2DLC) [213, 214, 215]. 2DLC is a rapidly developing technology whereby the second dimension column samples the output of the first dimension column using a sample valve and storage loop [216]. This configuration is run in a sequential manner, typically under computer control [214, 217, 218, 219]. 2DLC requires a fast second-dimension column so that the first dimension separation is rapidly and completely sampled by the

second column [214], in the so-called "comprehensive" mode. When properly implemented, this technique can provide dramatic enhancements in chromatographic resolution [220, 221].

The 2DLC method development process is largely empirical but specific schemes have been suggested [222, 223, 224, 225]. In one scheme [222], the "cardinal rules" of 2DLC method development are given and the first step and perhaps most important step is the selection of columns. Usually, this is a semi-empirical process where the solute class(es) are known *a priori* and columns are chosen to maximize speed and resolution where applicable. Of particular concern is the maximization of orthogonality [226], specifically the maximization of an individual orthogonality metric or products of orthogonality metrics [226], so that zones are spread as uniformly as possible across the maximum available separation space. Unless the column selection process is performed carefully, orthogonality maximization may not yield the desired separation performance that specific column pairs may offer.

Orthogonality evaluation and subsequent optimization takes place upon completion of the separation, [214] *i.e.* orthogonality metrics are utilized for an *a posteriori* performance evaluation. However, some schemes have been suggested that allow for method development prior to running the chromatogram. Utilizing the Snyder–Dolan hydrophobic subtraction model (HSM) [227, 228], orthogonality can be evaluated *a priori* using the weighted Euclidian distance metric [227, 228] and selectivity triangles and cubes [229, 230].

Column-pair rankings based solely on the concept of column orthogonality may suffer from the significant drawback that information regarding column order and operating time is not considered, but these factors are pertinent for defining the practical ability of a given column pair. For example, using a column that may fully resolve an analyte matrix but exhibits strong analyte/column interactions for a few compounds leading to unreasonably long elution times should be avoided, especially if this column is to be used in the second dimension.

In this work, we present a computational method that leverages the HSM to predict 2D chromatograms which, in turn, allows for ranking of column pair performance. In departure from other ranking methods, this method considers analyte/column interactions explicitly, allowing rankings based on the distance between peaks with consideration for column-pair order, mixing from the limited number of first-dimension samples, and different operating times. In the following sections a description of this novel column-pair ranking approach,

a comparison with two other orthogonality metrics, and suggestions towards the goal of defining optimal column pairs will be presented.

5.2 Methods for Column-Pair Ranking

The new column-pair ranking method, henceforth referred to as the 2DHSM method, utilizes virtual 2D chromatograms that are generated utilizing the HSM [227, 228, 231, 232, 233, 234, 235]. A retention factor, k'_i, can be computed for each analyte, i, on each column through application of the HSM retention equation as follows [227, 228]:

$$\log\left(\frac{k_i}{k_{\mathrm{EB}}}\right) \equiv \log k'_i = \eta_i H - \sigma_i S^* + \beta_i A + \alpha_i B + \kappa_i C, \tag{5.1}$$

where k_i and k_{EB} are the partition coefficients or net retention times for the analyte of interest and for ethyl benzene, the HSM reference analyte, respectively. The HSM column parameters H, S^*, A, B, and C refer to the hydrophobicity, steric hindrance, H-bond acidity, H-bond basicity, and cation exchange activity of the column, respectively, while the parameters η_i, σ_i, α_i, β_i, and κ_i denote the corresponding complimentary analyte properties.

The k' value is then used to compute the retention time for this analyte on a specific column but differentiating between a column utilized for the first or second dimension:

$${}^{d}t_i = \frac{(k'_i\, t^*_{\mathrm{EB}} + 1)\ {}^{d}t_{\max}}{k'_{\max}\, t^*_{\mathrm{EB}} + 1}, \tag{5.2}$$

where $t^*_{\mathrm{EB}} = (t_{\mathrm{EB}} - t_0)/t_0$ with t_{EB} and t_0 being the retention time of ethyl benzene and the column dead time; $k'_{\max}$ and ${}^{d}t_{\max}$ are the maximum retention factor (*i.e.*, an analyte with a retention factor greater than $k'_{\max}$ does not elute from the column during its maximum operating time) and the dimension-specific maximum operating time. To reduce the computational demands, the current work assumes that both t^*_{EB} and $k'_{\max}$ are independent of whether a column is used in the first or second dimension. However, it would also be possible to repeat the calculation of the 2D chromatograms utilizing a smaller $k'_{\max}$ for the second dimension.

In practice, analytes in a given sample taken from the first-dimension separation are mixed upon injection into the second dimension leading to a loss of resolution. To account

for this effect, a computed retention time in the first dimension is placed into a bin and assigned the time corresponding to the center of this bin according to:

$$n^{\text{bin}} = \left\lceil \frac{{}^1t_i \; n_{\text{sample}}}{{}^1t_{\text{max}}} \right\rceil \qquad\qquad {}^1t_i^{\text{bin}} = \frac{{}^1t_{\text{max}}}{n_{\text{sample}}} \left(n^{\text{bin}} + 0.5 \right), \tag{5.3}$$

where $\lceil \quad \rceil$ is the ceiling function; ${}^1t_i$ and n_{sample} refer the retention time of analyte i in the first dimension computed from Equation 5.2 and the number of samples injected into the second dimension, respectively. It is important to note that assigning a single n^{bin} value makes the assumption that an analyte will never elute into multiple samples taken from the first dimension. Relaxing this constraint would increase the complexity of the calculation of the 2D chromatograms.

Once ${}^1t_i^{\text{bin}}$ and the retention time in the second dimension, ${}^2t_i$, have been computed for a set of analytes and for all column pairs, the column pairs are ranked on the basis of how well separated the peaks are according to the following scoring function:

$$S_{\text{2DHSM}} = \sum_{i \neq j}^{\text{analytes}} \sum \delta \left({}^1t_i^{\text{bin}} - {}^1t_j^{\text{bin}} \right) \exp \left[\frac{-|{}^2t_i - {}^2t_j|}{10 \; ({}^2\sigma)} \right], \tag{5.4}$$

where the double sum is over eluting analyte pairs and use of the Kronecker delta function implies that analytes eluting in different first-dimension samples are counted as fully resolved. Following this approach, the exponential term contributes only to the sum for analytes which are injected into the second dimension in the same sample. The variable ${}^2\sigma$ is the nominal peak width in the second dimension. Using a constant value of ${}^2\sigma$ reflects a gradient elution mode, while making ${}^2\sigma$ a linear function of ${}^2t_i$ would reflect isocratic elution conditions. The scaling factor of 10 used here leads to a slower decay of the exponential term. A small value of S_{2DHSM} is desirable and corresponds to well separated peaks.

To place the 2DHSM approach into context, we compare its performance in finding high-performing column combinations to two often used orthogonality metrics. The first of these approaches, referred to in the present work as the weighted Euclidian distance (WED) approach, was proposed by Gilroy *et al.* [227] an Snyder *et al.* [228] and distinguishes column-pair orthogonality by considering the five-dimensional Euclidian distance between columns in physicochemical HSM space while neglecting analyte/column interactions and

column-pair ordering. The scoring metric is given by:

$$S_{\mathrm{WED}} = \left[(12.5\,\Delta H)^2 + (100\,\Delta S^*)^2 + (30\,\Delta A)^2 + (143\,\Delta B)^2 + (83\,\Delta C)^2 \right]^{0.5}, \qquad (5.5)$$

where ΔH, ΔS^*, ΔA, ΔB, and ΔC are the differences between the corresponding parameters in the HSM model (defined in Equation 5.1) for a given column pair, and the numerical pre-factor assign importance weights to the HSM parameter differences [227, 228]. A high S_{WED} value corresponds to a column pair for which each phase has a very different chemical character and, hence, is likely favorable for a 2D separation. Since the WED approach relies only on HSM column parameters, it can be applied *a priori* to compare column pairs without the need for 2D chromatograms.

The S_{WED} metric accounts for all five HSM parameters. For certain applications, however, when only certain classes of analytes are of interest, then it becomes advantages to adjust the weights in Equation 5.5, as was suggested for nonionized analytes [236] or for acidic and basic polycyclic aromatic compounds [237], or to focus on a restricted number of HSM parameters, such as in selectivity triangles and cubes [229, 230, 238]. It is has also been suggested [237] to use a principal component analysis to replace the weights in Equation 5.5.

One can also describe column-pair orthogonality through information entropy (IE). In this application, the information in question is the 2D chromatogram. The IE metric can only be determined *a posteriori* from existing chromatograms and is applied here to the chromatograms predicted via Equation 5.2 and 5.3. A high-entropy separation yields a chromatogram in which peaks are nearly uniformly dispersed. In order to apply IE theory, the 2D chromatogram is converted into a histogram with grid spacing of 100 in both dimensions, and the probabilities for observing a peak are used to calculate the IE score [239, 240]:

$$S_{\mathrm{IE}} = 100 \left[\frac{-\sum_{\mathrm{peaks},A} \sum_{\mathrm{peaks},B} P(x,y) \log_2 P(x,y) + \sum_{\mathrm{peaks},A} P(x) \log_2 P(x)}{-\sum_{\mathrm{peaks},B} P(y) \log_2 P(y)} \right], \qquad (5.6)$$

where $P(x)$, $P(y)$, and $P(x,y)$ are the probabilities to observe an analyte peak within temporal bin x on column A, within temporal bin y on column B, and within temporal bin x, y on columns A and B, respectively. The double sum in the numerator is information gained from the 2D chromatogram, the second term in the numerator is the negative of the

information gained from applying only the first-dimension column, and the denominator represents the information gained from applying only the second-dimension column. The S_{IE} value ranges from zero when all peaks fall along the diagonal (*i.e.*, the two columns yield identical retention order and relative peak positions) to 100 for a uniform distribution (*i.e.*, the probabilities for *all* temporal bins have the same value). The IE method [239, 240] employs a linear division of the retention time for the binning. In addition to the IE metric, there are several other metrics for measuring the orthogonality or surface coverage of 2D chromatograms [241, 242, 243, 244] that were recently reviewed by Davis and Schure [226].

5.3 Computational Details

This work utilizes the experimentally determined HSM parameters (at pH = 2.8) for 565 columns and 90 analytes (intended to represent a wide spectrum of possible analytes) that are tabulated by Boswell and Stoll [245]. Since the number of analytes is somewhat limited, HSM parameters are generated for an additional 99 910 analytes through 'virtual synthesis' involving the coupling of fragments taken from two of the 90 original analytes as follows:

$$M_{\mathrm{new}} = \chi_a M_a + \chi_b M_b, \tag{5.7}$$

where M_{new}, M_a and M_b are the HSM parameter of the new analyte, and the corresponding HSM parameters from each of the two analytes randomly selected from the pre-existing 90 analytes; χ_a and χ_b are random numbers uniformly distributed between 0.5 and 1.0. Linear combinations with $50 - 100\%$ contributions from existing analyte parameters are used in order to represent large analytes with multiple functionalities.

In order to compute the 2DHSM score, the input parameters used in Equations 5.2, 5.3, and 5.4 need to be specified. The parameters t^*_{EB}, ${}^1t_{\mathrm{max}}$, and ${}^2\sigma$ are set to 99 (*i.e.*, $t_{\mathrm{EB}}/t_0 = 100$), 10 000 time units, and 1 time unit, respectively. For the majority of the data presented, the values of n_{sample} and k'_{max} are set to 100 (*i.e.*, ${}^2t_{\mathrm{max}} = {}^1t_{\mathrm{max}}/n_{\mathrm{sample}} = 100$ time units and a peak capacity of 100 in the second dimension) and 1.375. Using $k'_{\mathrm{max}} = 1.375$ results in at least 85% of the analyte molecules eluting from each column during the specified operating time. To assess the sensitivity of the 2DHSM ranking to these input parameters, two other combinations are also explored: (i) $k'_{\mathrm{max}} = 2.0$, $n_{\mathrm{sample}} = 100$, and ${}^2\sigma = 1$, and (ii) $k'_{\mathrm{max}} = 1.375$, $n_{\mathrm{sample}} = 200$, and ${}^2\sigma = 1$. For *each* of the three sets of

column parameters, we now compute the 2D chromatograms for 100 different sets of 1 000 diverse analytes for all 319 225 column pairs, *i.e.*, in total the positions and widths are determined for $3 \times 100 \times 1\,000 \times 319\,225 = 1.2769 \times 10^{11}$ peaks.

The average number of resolved peaks for each column pair is used to to evaluate the performance of the 2DHSM, WED, and IS approaches. A peak is considered resolved if its resolution with all other peaks is at least 1.2. Common definitions of multi-dimensional resolution consider the distance between peaks in continuous multi-dimensional space [246], however, retention times in the first dimension are discretized during sample injection in the second dimension and, as mentioned above, we assume that peaks cannot elute into multiple samples. We therefore use the following definition of 2D resolution:

$$R_{\text{res}} = \left\lfloor \frac{|{}^2t_i - {}^2t_j|}{4\,({}^2\sigma)} \right\rfloor \qquad \text{if} \qquad \delta\left({}^1t_i^{\text{bin}} - {}^1t_j^{\text{bin}}\right) = 1, \tag{5.8}$$

where the variables are as defined for Equation 5.4. The above equation implies that peaks eluting in different first-dimension samples are fully resolved.

In order to visualize the results obtained herein, the peaks for the 2D chromatogram surfaces are computed by:

$$z = \sum_{i=1}^{N_{\text{analytes}}} \left(\sum_{j=1}^{1500} \sum_{k=1}^{1500} \exp\left[- \left(\frac{\frac{j}{1500}\,{}^1t_{\max} - {}^1t_i^{\text{bin}}}{{}^1\sigma\sqrt{2}} \right)^2 \right] \exp\left[- \left(\frac{\frac{k}{1500}\,{}^2t_{\max} - {}^2t_i}{{}^2\sigma\sqrt{2}} \right)^2 \right] \right), \tag{5.9}$$

where N_{solutes} is the number of eluting analytes and ${}^1\sigma$ is the base width in the first dimension, and is set to 20.

5.4 Results and Discussion

5.4.1 Performance of Scoring Methods

The performances of the WED, IE, and 2DHSM methods for column pair selections are graphically compared in Figure 5.1. A large average number of resolved peaks, $\langle N_{\text{res}} \rangle$, and a large average number of eluted analyte compounds, $\langle N_{\text{elu}} \rangle$, are desirable traits for a good column pair. Focusing first on the dependence of $\langle N_{\text{res}} \rangle$ on the average column rank, $\langle R \rangle$, it is evident that the three metrics perform quite differently. Ideally one would desire a distribution of data that exhibits a negative slope and is narrow with possibly also

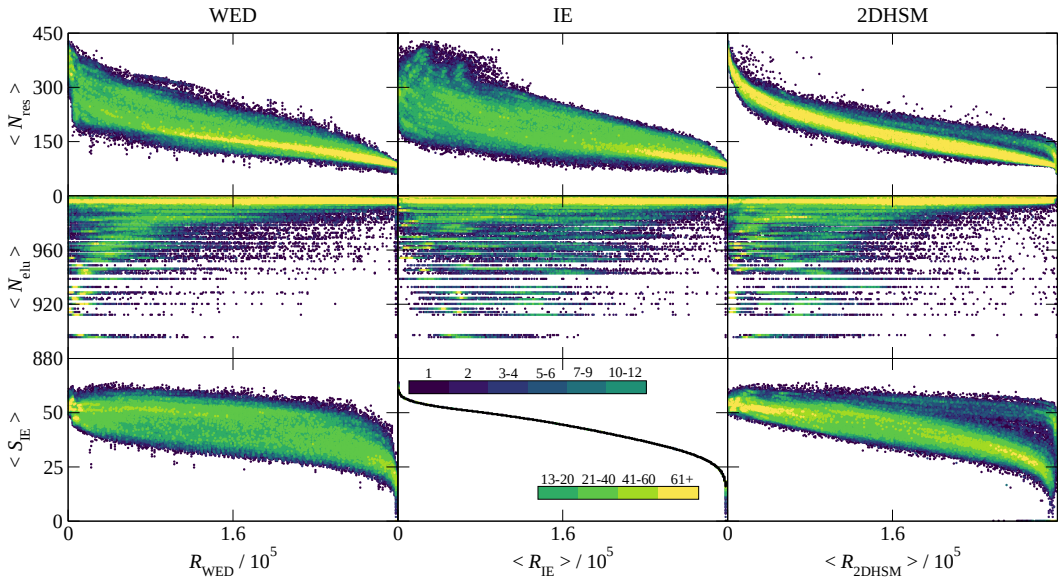

Figure 5.1: Overall performance of the WED, IE, and 2DHSM methods. Rows from top to bottom provide the average number of resolved peaks, average number of eluted analytes, and average orthogonality score, respectively, as a function of the average rank (following the 1^{st} place convention) assigned to a column pair by each metric. The data for each of the 319,225 column pairs are shown with the color denoting the point density (from dark purple to yellow for low density to yellow for high density).

a tapering of its width toward the upper left corner of the graphs; these factors indicate a metric that indeed results in lower ranks (following the 1^{st} place convention) for better performing columns and is highly selective in indicating suitable column choices.

For $\langle N_{\text{res}} \rangle$ versus $\langle R \rangle$, we see that all three methods yield a distribution with an overall negative slope. Looking in more detail, however, it is evident that the width of the distribution and the behavior at low ranks are quite different. The WED method yields a distribution that broadens at low ranks and tapers at high ranks (i.e, the highest point density shown in yellow in Figure 5.1 is found for high ranks). This result is not entirely surprising given that the original purpose of the WED metric is to help users identify similar columns for replacement purposes. Thus, the WED metric is more informative at high ranks, where columns should be most similar. The data generated using the IE approach is very diffuse, exhibiting little correlation between rank and performance for columns with lower rank. For the IE method, the highest $\langle N_{\text{res}} \rangle$ values are found for ranks near 25 000.

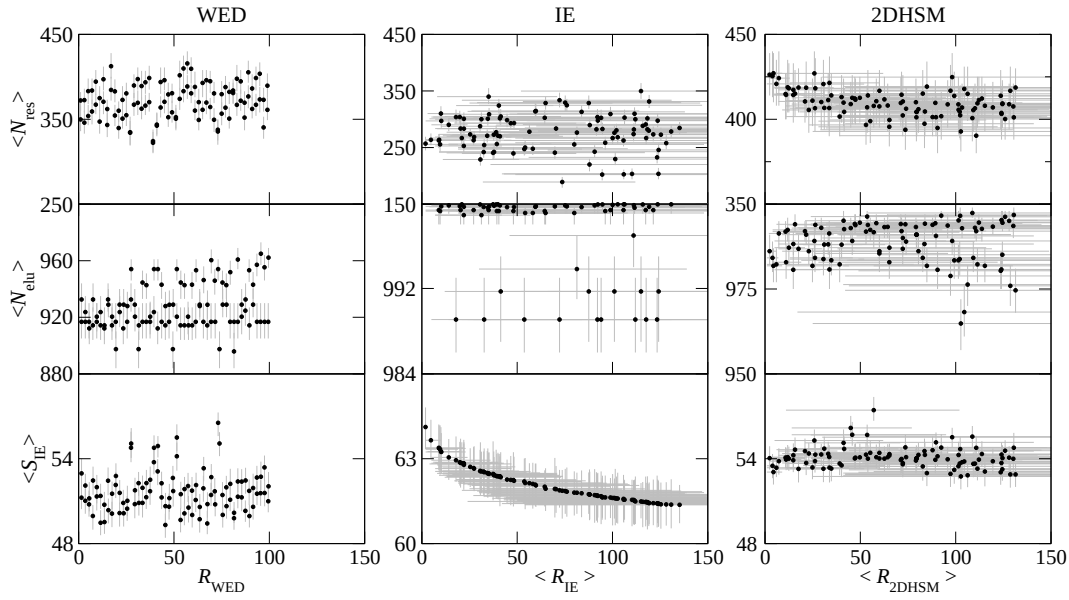

Figure 5.2: Performance of top-100 columns selected by the WED, IE, and 2DHSM methods. Rows from top to bottom provide the average number of resolved peaks, average number of eluted analytes, and average orthogonality score, respectively, as a function of the average rank (following the 1^{st} place convention) assigned to a column pair by each metric. Error bars indicate the 10–90 intervals.

The data for the 2DHSM approach, on the other hand, exhibit all of the desired traits, with a negative slope that rapidly increases in magnitude and a distribution that tapers as low ranks are approached. Furthermore, the spread of the $\langle N_{\text{res}} \rangle$ values for columns with similar ranks is much smaller for the 2DHSM method than for the WED and IE methods.

When examining the $\langle N_{\text{res}} \rangle$ values for the 100 top-ranked column pairs for the three methods, as is shown in Figure 5.2, then one observes that only the 2DHSM method yields a negative slope and more than 90% of these column pairs yield $\langle N_{\text{res}} \rangle$ values greater than 400. In contrast, the WED yield data clustered around $\langle N_{\text{res}} \rangle \approx 370$ and only seven of the 100 top-ranked column pairs yield $\langle N_{\text{res}} \rangle > 400$. The 100 top-ranked columns for the IE method perform much worse with $\langle N_{\text{res}} \rangle$ values falling into the range from 190 to 350. Note that the rank for the WED method is not averaged, because this metric does not consider analyte/column interactions and, hence, gives the same rank order for all 100 analyte sets. The WED method also does not account for column order and, hence, two column pairs

are associated with each rank. When considering the better performing column pairs with lower ranks, then the distributions of the R_{IE} and R_{2DHSM} values are quite asymmetric. In almost all cases, a tail to higher R values is found and, hence, the column ranked with the 100th highest average score yields $\langle R \rangle > 100$. In contrast, the distributions for N_{res} are fairly symmetric for most column pairs. Given the asymmetric distributions for R, the uncertainties in all properties are given here as the 10-90 interval, *i.e.*, the range into which 80 out of the 100 independent 2D chromatograms (each for 1 000 solutes) fall.

With respect to the number of eluted analyte compounds, $\langle N_{\mathrm{elu}} \rangle$, the overall distributions are quite similar for the three column pair scoring methods. As the rank improves, the fraction of pairs that elute less than 990 compounds increases. The highest point densities are found for $\langle N_{\mathrm{elu}} \rangle \approx 995$, but these are concentrated for $\langle R \rangle > 200\,000$ for all three methods. That is, eluting all 1 000 analytes is not necessarily beneficial for resolving the most peaks. Zooming in on the 100 top-ranked columns, the $\langle N_{\mathrm{elu}} \rangle$ values for the WED method are broadly distributed with a median at only 920 analytes. The IE method yields top-ranked columns with the highest $\langle N_{\mathrm{elu}} \rangle$ values, and 80% of them elute on avearge more than 998 analytes. For the 2DHSM method, 90% of the top-ranked columns elute more than 980 analytes.

Considering the information entropy score, $\langle S_{\mathrm{IE}} \rangle$, the distributions for the three methods are quite different. For the IE method, one observes a nearly monotonic decrease in $\langle S_{\mathrm{IE}} \rangle$ with increasing $\langle R_{\mathrm{IE}} \rangle$. It is not strictly monotonic because these averages are computed for 100 analyte sets, where sometimes a small change in S_{IE} can lead to a large change in R_{IE}. However, as indicated by the 10-90 intervals for the top-100 column pairs for this metric, the deviations from monotonic behavior are significantly smaller than the statistical uncertainties. The distribution for the WED method is considerably broader than for the 2DHSM method, *i.e.*, a high S_{WED} value is not well correlated with a high S_{IE} value. The column pairs with the highest $\langle S_{\mathrm{IE}} \rangle$ values are found for R_{WED} and $\langle R_{\mathrm{2DHSM}} \rangle$ greater than 20 000.

The 2DHSM method requires two input parameters that describe the preferred operating conditions: n_{sample} and k'_{max} that give the number of samples taken from the first dimension and the retention factor compared to ethylbenzene beyond which solutes are considered as not eluting from the column in the given operating time. To investigate the reliability of

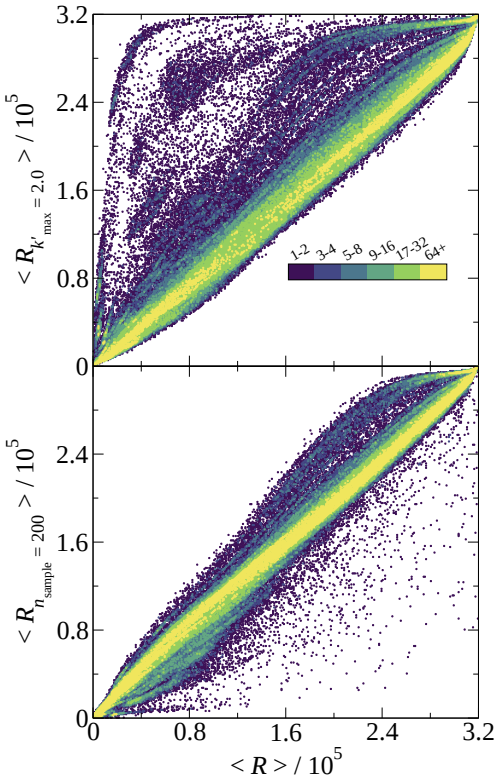

Figure 5.3: Influence of system parameters on the 2DHSM ranking: Correlation of the ranking obtained with $[n_{\text{sample}} = 100; k'_{\text{max}} = 2.0$ (top), $[n_{\text{sample}} = 200; k'_{\text{max}} = 1.375]$ (bottom) versus those obtained with $[n_{\text{sample}} = 100; k'_{\text{max}} = 1.375]$.

the 2DHSM method, two other combinations of $[n_{\text{sample}}; k'_{\text{max}}]$ parameters are also applied for to rank column pairs. As can be seen from the data presented in Figure 5.3, increasing k'_{max} to 2.0 yields a fair amount of scatter, with a correlation coefficient and slope of 0.939 and 0.926, respectively. On the other hand, increasing n_{sample} to 200 yields data that are very well correlated with those for $n_{\text{sample}} = 100$(the correlation coefficient and slope are both 0.989). While increasing n_{sample} to 200 has the potential to increase the significant burden on the ability of the second dimension column to yield good resolution within a short operating time, it can also substantially decrease the number of solutes that need to be separated in the second dimension, and can thus yield an overall separation which is comparable to the case with n_{sample} to 100. With k'_{max} to 2.0, on the other hand, a larger

number of solutes elute, leading to chromatograms which are more crowded than those obtained for $k'_{\max} = 1.375$. Note that analogues to Figures 5.2 and 5.2 for data sets using $k'_{\max} = 2$ and $n_{\text{sample}} = 200$ can be found in Appendix B.

5.4.2 Top Column-Pair Characteristics

Information regarding the chemistry of the top-10 column pairs obtained from the WED, IE, and 2DHSM approaches can be found in Table 5.1 (Note that analogues to Table 5.1 for data sets using $k'_{\max} = 2$ and $n_{\text{sample}} = 200$ can be found in Appendix B). The WED method does not distinguish column pair order and, hence, a given pair receives the same rank irrespective of which of the two is used for the first dimension. However, the performance data clearly indicate that column order matters, and the first listed entry for the top 10 columns pairs yields an average resolving power ($\langle N_{\text{res}} \rangle$) that is about 7% larger than for the same column pair but with reversed order. The first listed entry also yields an average orthogonality score that is 3% larger.

The $\langle N_{\text{res}} \rangle$ values for the top-10 column pairs range from $346-394$, $256-310$, and $414-427$ for the WED, IE, and 2DHSM methods, respectively; that is, the 2DHSM approach outperforms the other two metrics on average by factors of 1.50 and 1.14. Comparison of the ranks assigned for each column pair by each method (*i.e.* the first three columns of Table 5.1) shows that there is no overlap between top-10 column pairs for the three methods, suggesting that each respective approach describes column pair performance in a distinct fashion. Ranking by the WED method yields top-10 column pairs that elute on average only 920 analytes, whereas the top-10 columns for the 2DHSM and IE methods elute on average 986 and 999 analytes. The top-10 columns for the 2DHSM method also yield higher S_{IE} values than those from the WED method.

Interestingly, the top-10 columns for each method exhibit a very strong preference for including an embedded polar (EP) column in at least one dimension; that is an EP column is found for all of the 30 column pairs list in Table 5.1. The high WED score of for the combination of alkyl and embedded polar column pairs has been noted previously [247]. An EP column is found in the first dimension for all 10 column pairs obtained from the IE approache and in the second dimension for all 10 column pairs obtained from the 2DHSM approach. Furthermore, given that 565 columns were investigated here, it is noteworthy that

all three methods yield combinations with specific columns that occur multiple times among the top-10 column pairs. The EC Nucleosil 100-5 Protect 1 column is found eight times for the WED method, the Vydac 218MS column (an octadecyl phase) and the Cosmicsil Adore 100 CN column are included 5 and 3 times, respectively, among the top-10 pairs for the IE method, and the Zorbax Bonus RP column is found in all top-10 pairs for the 2DHSM method. Focusing on pairs obtained from the 2DHSM method that yields much higher $\langle N_{\text{res}} \rangle$ values than the IE method, only five specific EP columns were found in the top-100 column pairs: 72 Zorbax Bonus RP, 19 EC Nucleosil 100-5 Protect 1, 5 BetaMax Acid, 3 Hypersil Prism C18 RP, and 1 ZirChrom-PS. That is, an EP phase is present in 99 out of the top-100 columns. It has been shown that embedded polar stationary phase chains allow for multi-modal analyte siting and increased solvent penetration in the bonded phase [48, 74]. The embedded polar groups and the solvent penetration allow for enhanced retention of polar compounds, while at the same time preventing overly strong lipophilic analyte/bonded phase interactions. Thus, the multi-modal character enables elution of a diverse set of analytes on time scales consistent with a fast second dimension. The preference for specific columns in the first dimension is much weaker, and 73 unique columns are found in the top-100 column pairs for the 2DHSM method including 59 octadecyl, 25 phenyl, 9 triacontyl, 3 octyl, and 3 and 1 with EP and CN functionalities, respectively.

5.4.3 Top Pair Chromatograms

As a final test of method performance, 2D chromatograms are provided in Figure 5.4 for the top columns arising from the WED, IE, and 2DHSM methods. Co-elution of multiple analytes is prevalent for each column pair selection method. This is not surprising because the ratios of $\langle N_{\text{elu}} \rangle / \langle N_{\text{res}} \rangle$ are 2.5, 2.6, 3.9, and 2.3 for the top column pair suggested by the WED (both orders), IE, and 2DHSM methods (see Table 5.1). As expected from these ratios, the most compact chromatogram is obtained for the top-performing column suggested by the IE method for which the maximum number of co-eluting analytes is seven. The corresponding numbers are five and four for the top column pairs suggested by the WED and 2DHSM methods.

Table 5.1: Rank and performance scores of the top column pairs obtained with the WED, IE, and 2DHSM methods. The abbreviations EP, C18, C30, and PBD refer to embedded polar, octadecyl, triacontyl, polybutadiene phases, respectively. The super- and subscripts for each property denote the 10-90 interval.

R_{WED}	$\langle R_{\mathrm{IE}} \rangle$	$\langle R_{\mathrm{2DHSM}} \rangle$	Column 1	Column 2	Chem 1	Chem 2	$\langle N_{\mathrm{res}} \rangle$	$\langle N_{\mathrm{elu}} \rangle$	$10\langle S_{\mathrm{IE}} \rangle$
1	63529^{80550}_{47537}	6708^{10917}_{3993}	ZirChrom-PBD	EC Nucleosil 100-5 Protect 1	PBD	EP	372^{382}_{360}	933^{944}_{924}	530^{540}_{520}
1	37588^{48717}_{26142}	20527^{31476}_{9645}	EC Nucleosil 100-5 Protect 1	ZirChrom-PBD	EP	PBD	350^{366}_{336}	917^{930}_{905}	512^{521}_{504}
3	67155^{83196}_{50948}	5024^{7946}_{3013}	ZirChrom-EZ	EC Nucleosil 100-5 Protect 1	Other	EP	373^{386}_{358}	924^{935}_{913}	521^{529}_{511}
3	49798^{65812}_{36779}	29743^{54613}_{12063}	EC Nucleosil 100-5 Protect 1	ZirChrom-EZ	EP	other	346^{360}_{333}	917^{930}_{905}	510^{518}_{501}
5	64289^{82002}_{47895}	41891^{78295}_{14810}	EC Nucleosil 100-5 Protect 1	Apex II C18	EP	C18	383^{398}_{368}	912^{924}_{901}	512^{523}_{502}
5	71319^{87015}_{54413}	2940^{4568}_{1593}	Apex II C18	EC Nucleosil 100-5 Protect 1	C18	EP	354^{367}_{343}	917^{930}_{905}	508^{517}_{499}
7	44494^{58155}_{31492}	17997^{28168}_{9030}	Zorbax Bonus RP	ZirChrom-PBD	EP	PBD	384^{398}_{368}	933^{944}_{924}	525^{534}_{514}
7	84785^{99401}_{68053}	4647^{7064}_{2566}	ZirChrom-PBD	Zorbax Bonus RP	PBD	EP	359^{373}_{344}	914^{927}_{904}	500^{507}_{489}
9	62103^{78605}_{47010}	2963^{5063}_{1396}	Resolve C18	EC Nucleosil 100-5 Protect 1	C18	EP	394^{408}_{379}	920^{930}_{910}	520^{528}_{508}
9	51580^{67695}_{36698}	23285^{43479}_{8436}	EC Nucleosil 100-5 Protect 1	Resolve C18	EP	C18	367^{380}_{356}	917^{930}_{905}	513^{521}_{503}
46719	$\mathbf{1.94}^{3}_{1}$	86973^{97146}_{75879}	Cosmicsil Adore 100 CN	Vydac 218MS	EP	C18	257^{269}_{245}	1000^{1000}_{1000}	641^{648}_{633}
43979	$\mathbf{4.75}^{10}_{2}$	68765^{77461}_{60403}	Fortis Cyano	Vydac 218MS	EP	C18	264^{275}_{254}	1000^{1000}_{1000}	637^{643}_{630}
61561	$\mathbf{8.92}^{22}_{2}$	78169^{91338}_{65566}	ProntoSIL CN	Discovery HS PEG	EP	EP	262^{274}_{249}	999^{1000}_{998}	634^{642}_{626}
41227	$\mathbf{9.11}^{23}_{2}$	75576^{86549}_{66207}	Cosmicsil Adore 100 CN	Cogent UDC Cholesterol	EP	other	264^{274}_{252}	1000^{1000}_{1000}	633^{641}_{625}
53755	$\mathbf{9.58}^{19}_{2}$	94818^{107127}_{82846}	Pinnacle II Cyano	Discovery HS PEG	EP	EP	256^{266}_{243}	999^{1000}_{998}	633^{641}_{624}
36667	$\mathbf{10.0}^{20}_{3}$	29671^{35423}_{24392}	Discovery CN	Vydac 218MS	EP	C18	297^{308}_{285}	1000^{1000}_{999}	632^{639}_{624}
23197	$\mathbf{10.2}^{23}_{2}$	28322^{33435}_{23766}	Cosmicsil Adore 100 CN	Purospher RP-18	EP	C18	310^{323}_{298}	1000^{1000}_{1000}	633^{641}_{625}
38781	$\mathbf{14.1}^{24}_{5}$	35482^{41118}_{29680}	Genesis CN 300A	Vydac 218MS	EP	C18	290^{302}_{279}	1000^{1000}_{999}	630^{636}_{622}
60859	$\mathbf{17.9}^{36}_{5}$	25931^{31988}_{20689}	Discovery HS PEG	Microsorb-MV 100 CN	EP	EP	304^{317}_{292}	989^{993}_{986}	629^{637}_{621}
32645	$\mathbf{19.9}^{40}_{6}$	25206^{29941}_{21088}	ACE 5 CN	Vydac 218MS	EP	C18	304^{315}_{292}	1000^{1000}_{999}	629^{635}_{620}

Table 5.2: Continued from Table 5.1: Rank and performance scores of the top column pairs obtained with the WED, IE, and 2DHSM methods. The abbreviations EP, C18, C30, and PBD refer to embedded polar, octadecyl, triacontyl, polybutadiene phases, respectively. The super- and subscripts for each property denote the 10-90 interval.

R_{WED}	$\langle R_{\text{IE}} \rangle$	$\langle R_{\text{2DHSM}} \rangle$	Column 1	Column 2	Chem 1	Chem 2	$\langle N_{\text{res}} \rangle$	$\langle N_{\text{elu}} \rangle$	$10\langle S_{\text{IE}} \rangle$
599	24618^{31119}_{17820}	$\mathbf{2.42}^{6}_{1}$	Prontosil 60 Phenyl	Zorbax Bonus RP	phenyl	EP	426^{439}_{414}	986^{991}_{982}	540^{548}_{532}
477	30640^{38600}_{22046}	$\mathbf{3.93}^{7}_{1}$	Ultra II Aqueous C18	Zorbax Bonus RP	C18	EP	425^{440}_{413}	984^{989}_{980}	535^{543}_{526}
403	36049^{46471}_{26550}	$\mathbf{4.43}^{9}_{1}$	Prontosil 120-3-C30	Zorbax Bonus RP	C30	EP	427^{439}_{410}	982^{987}_{977}	531^{539}_{521}
553	32247^{39757}_{24621}	$\mathbf{5.96}^{14}_{1}$	Ultra II Aromax	Zorbax Bonus RP	phenyl	EP	421^{435}_{408}	982^{987}_{977}	534^{540}_{527}
999	26911^{34175}_{19881}	$\mathbf{7.27}^{13}_{2}$	Allure Organic Acids	Zorbax Bonus RP	C18	EP	424^{440}_{408}	991^{995}_{987}	538^{547}_{528}
1455	24147^{30334}_{17885}	$\mathbf{10.8}^{21}_{4}$	Ultra Aromax	Zorbax Bonus RP	phenyl	EP	415^{427}_{400}	992^{995}_{988}	541^{548}_{533}
629	25699^{32706}_{18099}	$\mathbf{11.1}^{19}_{4}$	Ultra II Biphenyl	Zorbax Bonus RP	phenyl	EP	414^{427}_{400}	983^{988}_{978}	539^{547}_{531}
1019	23907^{30779}_{17479}	$\mathbf{11.9}^{20}_{4}$	Ultra Aqueous C18	Zorbax Bonus RP	C18	EP	419^{435}_{404}	992^{996}_{989}	541^{549}_{532}
545	24987^{31916}_{17676}	$\mathbf{12.5}^{21}_{5}$	HSS C18 SB	Zorbax Bonus RP	C18	EP	418^{433}_{404}	987^{992}_{982}	540^{548}_{531}
569	26218^{33559}_{19228}	$\mathbf{14.8}^{29}_{5}$	Pinnacle II Biphenyl	Zorbax Bonus RP	phenyl	EP	414^{427}_{398}	981^{986}_{975}	539^{546}_{530}

5.5 Conclusions and Applications

The problem of selecting columns for analytical separations becomes substantially more complex when a second dimension is added. In this work, we show that the existing solutions to this approach which rely on the concept of column orthogonality or Euclidian distance are insufficient for specifying optimal pairs/pair ordering. We instead develop the 2DHSM method to rank column pairs that accounts for column-pair order, number of resolved peaks, and resolution between peaks. This method is relatively forgiving with respect to input parameter choice, more selective, and yields top-10 column pairs which on average resolve 14–50% more analytes than competing methods. It is found that columns pairs that resolve the largest number of analytes within a given time constraint are not necessarily those which elute the most analytes. Beyond suggesting high-performing 2DLC column pairs, the 2DHSM method also provides insight into the column chemistries contributing to performance. An exceedingly strong preference for including an EP column in the second dimension is found, and the top-100 column pairs are comprised mostly of C18 $\times$ EP (59%) and phenyl $\times$ EP (25%) column pairs.

From a practical standpoint, the 2DHSM method would be most useful as a selection tool for column pairs from a user-defined pool (instead of selecting from all 565 columns). For instance, one may desire to run this analysis on a set of columns provided by the same manufacturer to ensure compatibility of fittings, etc. As with the WED and IE methods, it is straightforward to extend the 2DHSM method to higher dimensional HPLC-based separations. It should be noted that Equation 5.4 or an analogue allowing for analytes to elute in more than one first-dimension sample could be used to quantify the extent of separation for experimentally measured multi-dimensional chromatograms.

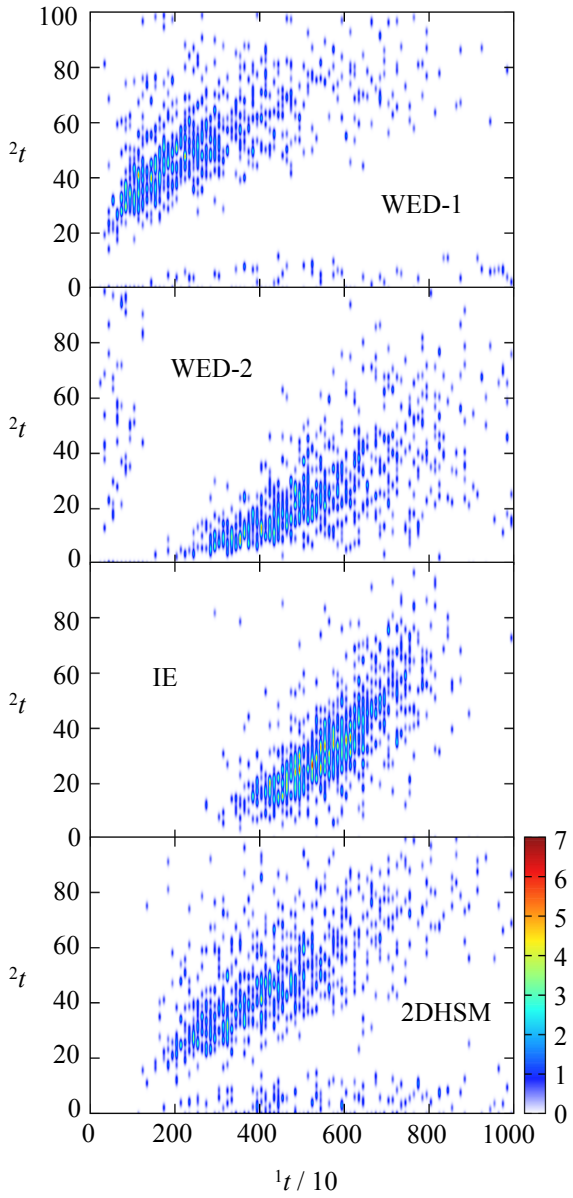

Figure 5.4: Two-dimensional chromatograms for the top-ranked column pairs: ZirChrom-PBD × EC Nucleaosil 100-5 Protect 1 (top) EC Nucleaosil 100-5 Protect 1 × ZirChrom-PBD (top middle) suggested by the WED method, Cosmicsil Adore 100 CN × Vydac 218MS (middle bottom) suggested by the IE method, and Prontosil 60 Phenyl × Zorbax Bonus RP (bottom) suggested by the 2DHSM method.

Chapter 6

Simulation of Solute Uptake in Surfactant Bilayers

6.1 Introduction

Solute uptake in micro-structured soft materials is a recurring theme across several fields. In drug design, for example, one may be interested in development of an optimal vehicle to transport a relatively non-polar pharmaceutical compound through an aqueous medium [95, 248, 249]. Alternatively, one may be concerned with enhancing uptake of fragrant compounds for use in personal care or household cleaning products [6, 7]. Regardless of the immediate area of application, however, several questions arise about the central challenge of tuning solute uptake, including: *i.* how do solute uptake trends relate to micellar structure? *ii.* how do these trends vary with surfactant and solute functionality? and *iii.* how does introduction of additional components influence solute uptake and bilayer stability?

Extensive work has been conducted to answer these challenging questions and has largely confirmed that no single factor governs solute uptake, as solubilization is determined by interactions with each substance (water, surfactant, etc.) and the interfaces they create, for example, between oil and surfactant or water and surfactant. Furthermore, the mesophases these micelles form can modify the extent to which favorable solute sorption domains exist, as well as physical properties of the solution [92, 96]. These findings are

complicated by the fact that solute uptake can also influence structure, and ultimately, loading; for example, it has been shown that use of mixed surfactants and solutes can either enhance or suppress solute uptake, and for polar solutes, subsequent water penetration into micellar structures can result in decreased loading of polar solutes, due to competitive effects [96, 250, 251]. Despite the multitude of experimental studies on surfactant systems, [80, 83, 94, 97, 252, 253, 254, 255] determination of the corresponding molecular level structure is a difficult task. Methods such as NMR, spectroscopy, and scattering can provide information on the structural aspects such as the mesophase structure, head group area, and bilayer repeat distance, however, many of the finer details are left to be extracted through models containing numerous assumptions, or are too difficult to measure. For example, the structure of water in the vicinity of ether groups like those found in CxEy surfactants, such as triethylene glycol mono-n-decyl ether (C10E3), has been probed by vibrational spectroscopy [87, 89], scattering [86], NMR [90, 91], and thermodynamic models [256], yet their precise structure is still debated.

Molecular simulations provide an alternative approach for investigating the microscopic structure of complex chemical systems, thus, the present work aims to study surfactant systems through the lens of molecular simulation. Numerous simulation studies have been conducted for such systems [81, 98, 116, 117, 257, 258, 259, 260, 261], with the majority having relied on molecular dynamics or other closed system methods, which are limited to pre-specified system compositions. Given that many problems of interest involve loading of highly hydrophobic compounds into a surfactant carrier, closed system approaches are quite problematic. Indeed, one is faced with a limited set of options, either simulating the system in the infinite dilution regime (*i.e.*, only a single solute molecule is present in the system), which has limited application in surfactant research, or simulating a system with enough water that a non-trivial concentration of solute can be obtained. While the latter approach would offer more useful insights, the necessary system sizes are prohibitively large.

Others have examined structured surfactant materials using open system approaches such as grand canonical Monte Carlo, however the applications were primarily focused on coalescence of surfactant-coated droplets [262] prediction of the critical micelle concentration [116], or used simplified models which did not directly relate to a specific system [98].

In the present work, we are concerned with simulation of solute uptake in structured micellar materials. In departure from what has been previously done, we leverage osmotic Gibbs ensemble Monte Carlo, and use models of varied resolution. In our approach, a "unit cell" of the target surfactant system is contained in one box, and is in thermodynamic contact with a second box containing a vapor phase of the solute of interest, between which the solute can transfer. As pressure of the second box is varied, we are able to predict solute loading and the resultant structural changes, and in this way, shed light on how micellar structure and solute uptake are interrelated. Specifically, loading of *n*-nonane and 1-hexanol, in a lamellar C10E3 system in united-atom (UA) and coarse-grained (CG) descriptions will be presented, with the goals of *i.* establishing simulation protocols, *ii.* understanding how solute functionality influences loading and structural changes, and *iii.* comparing the performance of force fields having different resolutions (*i.e.*, UA versus CG) to assess their utility for describing quantitative and qualitative trends.

6.2 Simulation Details

6.2.1 Support Calculations

Prior to conducting loading simulations for the surfactant system, a series of support simulations are performed, the first of which being determination of the vapor pressure, p_i^*, of each solute i, predicted by the CG and UA models. This vapor pressure allows the reservoir pressure, p_{res}, to instead be described in terms of a scaling value, $a_i = p_{\text{res}}/p_i^*$. The primary benefit of operating in terms of a_i rather than p_{res} is that it allows results to be compared directly between different models and solutes (which might predict different p_i^* for the same solute) and with experiment (which might yield a different p_i^* than the model).

In order to compute solute vapor pressures, configurational-bias Monte Carlo [104, 105, 106, 134] simulations in the NVT Gibbs ensemble [107, 108] are run for each solute at 300 K. Results for these support calculations are provided in Table 6.1. For the UA cases, it is important to note that, in order to aid in simulation efficiency, the present calculations have been run with protocols that differ slightly than those used to parameterize the force field, thus the predicted p_i^*'s do not represent the true accuracy of the model.

Transfer free energies at infinite dilution, ΔG_{Transf}, for each solute between a vapor and

Table 6.1: Simulation details and vapor pressures.

	N^a_{CG}	N^a_{UA}	${p^*_{\mathrm{CG}}}^b$	Z^c_{CG}	${p^*_{\mathrm{UA}}}^b$	Z^c_{UA}	${p^*_{\mathrm{exp}}}^{b,d}$
C10E3	100	54	–	–	–	–	–
Water[e]	1614	872	–	–	–	–	–
Nonane	200	108	0.657_6	1.00_1	1.24_2	0.998_3	0.603 [263]
Hexanol	400	216	3.15_1	0.997_2	0.0325_4	1.000_2	0.136 [264]

[a]Values correspond to simulations containing surfactant.
[b]All pressure are given in kPa.
[c]Compressibility factor.
[d]References are given in brackets, next to values.
[e]The number of CG water beads is given by $N_{\mathrm{CG}}/3$.

bulk water phase serve as an additional support calculation. These data allow for evaluation of model performance by way of comparison to experimental values and are related to the Ostwald solubility at infinite dilution through:

$$\Delta G_{\mathrm{Transf}} = -RT \ln K_{\mathrm{O}}, \tag{6.1}$$

where R is the gas constant, T is the temperature, and K_{O} is the Ostwald coefficient, defined as the ratio of number densities between the liquid and vapor phases. In order to obtain $\Delta G_{\mathrm{Transf}}$, configurational-bias Monte Carlo simulations in the NpT Gibbs ensemble are run for systems containing a box of liquid water and a vapor box which contains both the solute of interest and helium at 300 K and 1 atm; helium is added to ensure the vapor box does not collapse during the simulation. Solute and water molecules are allowed to transfer between the two boxes, allowing for computation of a transfer free energy.

Table 6.2 provides the simulation results along with ranges over which experimental values have been reported [265]. Note that the experimental data have been converted from Henry's Law constants. The data show that, for both the CG and UA models, nonane $\Delta G_{\mathrm{Transf}}$ results fall above the range of experimentally measured values, though the UA results are closer. For hexanol, results for the CG model are in agreement with experiment, within uncertainty, while the UA values are slightly smaller.

Table 6.2: Transfer free energies.

	CG[a]	UA[a]	Experiment,[a] [265]
Nonane	22.0_4	16.8_4	7.1 to 13.6
Hexanol	-16.3_7	-14.0_5	−19.5 to −17.0

[a]Free energies are given in units of kJ/mol.

6.2.2 Simulations of Surfactant Systems

Configurational-bias Monte Carlo, in the $N_{\text{wat}}N_{\text{C10E3}}p_{\text{bil}}\mu_i T$ osmotic Gibbs ensemble, serves as the driver for simulations of surfactant containing systems, where the subscripts "wat", "C10E3", "bil" and "i" refer to water, surfactant, the bilayer box, and the solute of interest, i. A two-box setup is employed in which the first box contains a C10E3 bilayer solvated by water, yielding a 50/50 weight percent mixture, at 300 K and 1 atm. When replicated in all directions this "unit cell" yields a lamellar structure, which is found to be stable in the present simulations, in agreement with experiment [83]. The second box contains an ideal gas of either n-nonane or 1-hexanol at 300 K and a given a_i, referred to as either a_{nona} or a_{hexOH}, respectively. The compressibility factors for solutes from each model, which are listed in Table 6.1, confirm that the solutes are indeed in the ideal regime at the simulation conditions.

6.2.3 Simulation Parameters and Force Fields

All calculations discussed in the present work are run using the MCCCS-MN software suite [176]. Molecules in each box are allowed to translate and rotate, and with the exception of water, undergo conformational changes. The probability for a volume move is set to $2/N_{\text{molec}}$, where N_{molec} is the total number of molecules in the system, yielding ≈ 1 accepted move per cycle. In two-box systems, volume moves, in addition to particle swap and identity exchange [66, 131, 132] move probabilities are chosen to have a cumulative value of 0.3. Remaining probabilities are distributed among configurational-bias (CBMC), translation, and rotation moves, on the basis of relevant degrees of freedom. This distribution of

probabilities yields a balance between accepted particle transfers (where applicable) and the remaining move types. For UA surfactant molecules, 40% of CBMC moves were conducted through use of the self-adapting fixed-endpoint variant of the CBMC algorithm [135], in order to enhance conformational sampling of the molecule's interior segments. Molecule identity exchange moves are only conducted for solute-type molecules, and are introduced to compensate for the low acceptance rate associated with transferring molecules between a gas and relatively dense liquid phase. These moves involve the use of two impurity molecules for each solute, which represent one third and two thirds of the full molecule, respectively. Only the smallest impurity is allowed to transfer between boxes directly through swaps, while the larger impurity and the solute itself can only move between boxes by exchanging identities with either one of the impurities or the solute. A uniform biasing potential is applied to the impurity molecules in the condensed-phase box in order to ensure that they spend an equivalent amount of the simulation in each of the boxes [66]. The number of molecules used for surfactant-containing simulations are listed in Table 6.1. For support calculations, 1100/5250, 2/2 and 25/25 molecules of water, solute, and helium are used in the UA/CG transfer free energy simulations, while 500/1000 molecules are used for UA/CG vapor pressure calculations. In all simulations, only one molecule of each impurity type is present.

The Shinoda-DeVane-Klein (SDK) force field, which has been used extensively to study surfactant systems, is employed for the CG simulations, generally utilizing a three-to-one mapping of heavy atoms to a single interaction site (for example, propane would be represented by a single "bead"), and mapping three water molecules to a single CG site [109, 110]. These simulations use a 15 Å cutoff and no tail corrections. The Transferable Potentials for Phase Equilibria united-atom (TraPPE-UA) force field [137, 140, 141], which merges hydrogen atoms not belonging to polar functional groups with their parent heavy atom, is used in conjunction with the four-site TIP4P water model [199], for the UA representation. Here, a 14 Å cutoff with tail corrections is employed, and Coulomb energies are computed through a neutral-group-based cutoff approach. Unless otherwise stated, all calculations use four independent simulations and at least 3×10^5 and 1.5×10^5 cycles of equilibration and production, respectively, where one cycle is comprised of N_{molec} Monte Carlo steps. Errors are reported as a subscripted uncertainty in the last digit on the computed value, and

correspond to standard errors of the mean for support calculations, and the 95% confidence interval for all other data.

6.3 Equilibration of Surfactant Systems

It is important to evaluate whether initial system setup or sampling methods influence final simulation results. In the present work, this is addressed by probing the influence of *i.* initial system loading, *ii.* initial system geometry, and *iii.* cell volume/shape sampling approach on loading in the nonane-containing CG system. To do so, eight independent simulations are run, half of which were initialized with a slab of N_{C10E3} nonane molecules present within the bilayer, and the remaining nonane molecules in the reservoir (a "full bilayer" setup) while the other half were initialized with a "bare bilayer", having all nonane molecules in the reservoir. For a given setup (*i.e.*, "bare" or "full bilayer"), two initial geometry types having the same total volume were tested: *i.* elongated in the z-direction, which is perpendicular to the bilayer interface, and *ii.* compressed in the z-direction. Finally, two bilayer box cell sampling approaches were evaluated for each initial loading and geometry *via* a move which changes the volume of the system by altering either x and y box lengths in a coupled fashion or the z box length independently (a "coupled xy volume move"), and a move in which coupled changes in xy are compensated by changes in z, thereby maintaining the overall cell volume (a "coupled xy constant volume cell move"). Figure 6.1 shows the volume equilibration trajectories as well as the final density profile for each simulation, and demonstrates that choice of cell sampling approach and initial geometry has no influence on the final equilibrated structure. Thus, all remaining equilibration simulations (*i.e.*, CG hexanol, UA nonane, and UA hexanol) use a single initial geometry, only coupled xy volume moves to sample cell volume/shape, and are initialized with half the systems having a "full bilayer," and the other half "bare".

Following volume equilibration of the initial system, particle transfer between the bilayer and reservoir boxes is introduced for prediction of loading. Five replicate simulations are launched for each system type (CG or UA and nonane or hexanol), differing by a_i, which ranges from 0.25 to 1.25. Trajectories corresponding to particle transfer equilibration are monitored for each system, and convergence of loadings between the "bare" and "full" setups serve as an indicator that equilibrium has been achieved. Figure 6.2 shows an example of

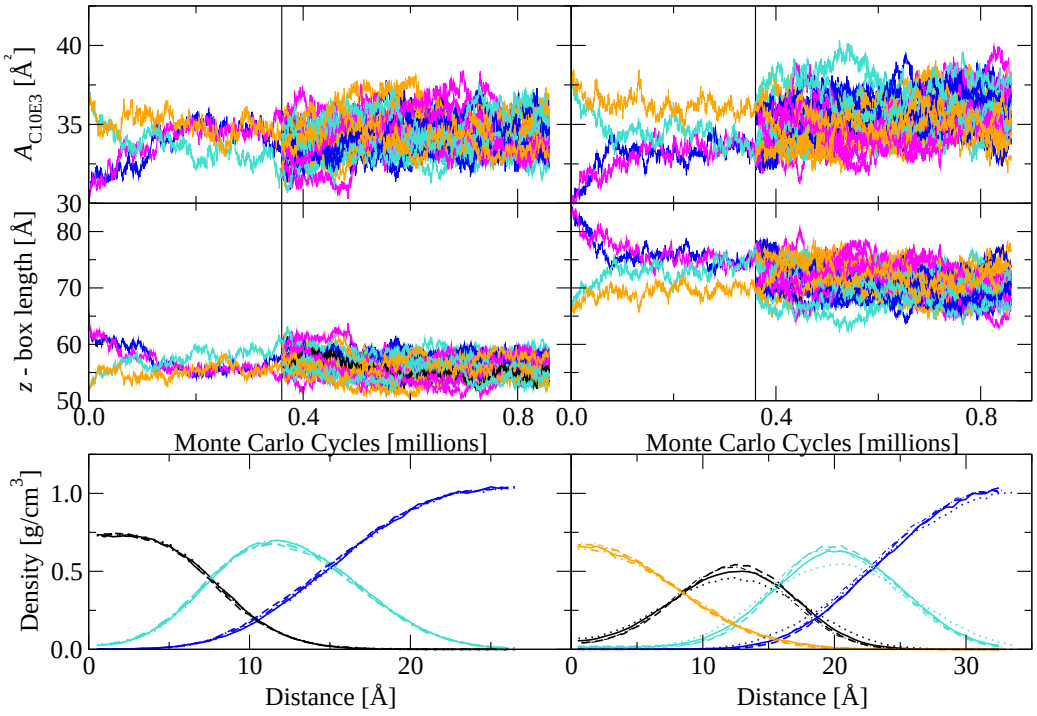

Figure 6.1: Top: Volume equilibration trajectory plots for CG systems with nonane as the solute. The left and right columns correspond to "bare" and "full" bilayer setups, respectively. The vertical black line indicates the point at which multiple independent simulations were launched. Bottom: Density profiles for the volume-equilibrated systems. Columns are as given for the top figure. Orange, black, teal, and blue lines correspond to nonane, the C10 portion of the surfactant, the E3 portion of the surfactant, and water, respectively. Solid, dashed, dotted, and dot-dashed lines correspond to simulations initialized with an elongated z-dimension and using constant volume cell moves, a compressed z-dimension and using constant volume cell moves, an elongated z-dimension and coupled volume moves, and a compressed z-dimension and coupled volume moves, respectively.

particle transfer equilibration trajectories for the CG nonane system. These plots also serve as one way to distinguish system stability; for example, as is shown in Figure 6.2, head group area, A_{C10E3} remains largely unaffected as a_{nona} is increased, however an abrupt change in the loading isotherm and z-box length trends at $a_{\mathrm{nona}} \geq 1.0$ is observed, indicative of nucleation of a bulk-like solute layer between monolayer sheets. Note that this behavior is not entirely unexpected, as the vapor reservoirs in these simulations are at and above the solute vapor pressure. The implication of such behavior is that the surfactant bilayer serves to lower the nucleation barrier for the bulk-like liquid solute phase. Thus, only simulations below $a_i = 1.0$, which more rigorously correspond to a lamellar phase, are discussed in detail. For all systems, loading equilibration was run for at least $5 \times 10^6/6 \times 10^5$ cycles, and production was run for at least an additional $5 \times 10^6/4 \times 10^5$, for the CG/UA systems, respectively. Once systems were equilibrated, 8 to 16 independent production runs were launched at each a_i.

6.4 Results and Discussion

6.4.1 Nonane Loading

Unary loading isotherm data, constructed for stable systems ($a_{\mathrm{nona}} \leq 0.75$), are provided in Table 6.3. Both CG and UA nonane undergo monotonic increases in uptake with increasing reservoir pressure, however both simulation types also exhibit small positive deviation from ideal loading behavior. This behavior suggests that nonane loading helps stabilize the bilayer, likely by enhancing favorable hydrophobic interactions within the bilayer interior. Note that experimental studies for the ternary C10E3/water/docecane system show stability of the lamellar phase at similar compositions [83]. Snapshots of the CG and UA systems, provided in Figure 6.3, show that the two systems look very similar to one another, with nonane exhibiting a preference to load between monolayer leaflets, and water generally being excluded from the bilayer interior; visually, the primary difference between the two systems is a larger degree of disorder in the UA system, which, to some extent, is due to the increase in model resolution. Density profiles, shown in Figure 6.4, provide a more detailed comparison of the structures of each system. With the exception of the CG profiles being more extended in the z-direction than those for the UA systems, trends are

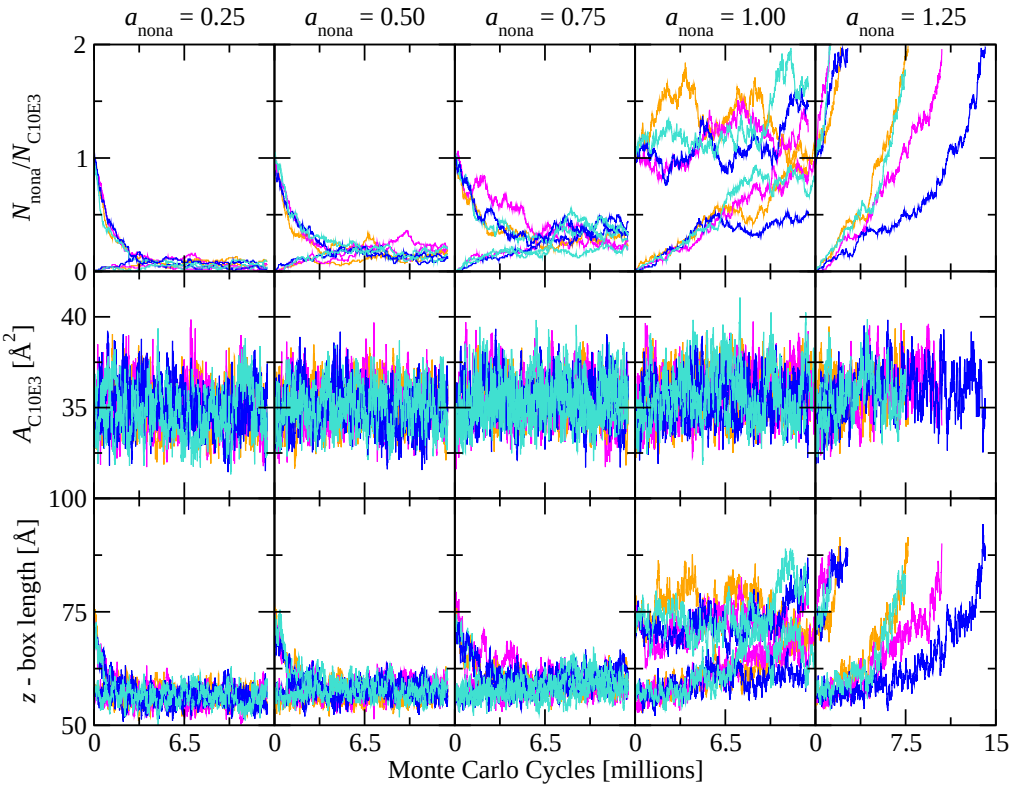

Figure 6.2: Loading trajectories for equilibration of CG bilayer systems with nonane solutes. The zero of the x-axis corresponds to the point at which particle transfer and identity exchange moves were introduced. Blue and magenta lines correspond to systems initialized with a compressed z-dimension while cyan and orange lines are provided for systems initialized with an elongated z-dimension. Blue and cyan lines also correspond to systems which utilize coupled xy constant volume cell moves, while magenta and orange lines correspond to systems using only coupled xy volume moves. Two lines of the same color in any given plot correspond to simulations using a "bare" and "full" bilayer setup, but otherwise running with the same protocols.

Table 6.3: Nonane loading properties.

	CG			UA				
a_{nona}	loading $N_{\text{sol}}/N_{\text{C10E3}}$	A_{C10E3} [Å^2]	t_{bil} [Å]	loading $N_{\text{sol}}/N_{\text{C10E3}}$	A_{C10E3} [Å^2]	t_{bil} [Å]	r_{ete} [Å]	$\langle S \rangle$
0.00	0.0	34.8_6	28.2_2	0.0	41_1	24.7_3	15.86_5	0.13_1
0.25	0.066_1	34.8_1	29.1_1	0.10_2	42_1	25.5_3	15.90_3	0.13_1
0.50	0.156_5	35.2_1	30.4_1	0.21_3	43_1	26.7_5	15.91_2	0.14_1
0.75	0.29_1	35.5_1	32.3_2	0.33_6	43_1	28.3_4	16.08_2	0.14_1

consistent between the two models. In both cases, the nonane density profile exhibits a single peak at $z = 0$, which increases in height with loading. The C10 density at the center of the bilayer is decreased and the profile is shifted to the right, as a consequence. The E3 and WAT profiles maintain a roughly constant maximum density, but are also shifted to the right with increased nonane loading. Dissipative particle dynamics (DPD) simulations of a lamellar C10E6 system containing octadecane solutes has yielded similar results; surfactant leaflets were found to be pushed away from each other at higher solute concentrations, decreasing the amount surfactant at the interior of the bilayer [259]. Additionally, the solutes were found to load between monolayers, exhibiting a single broad peak.

The observed trend of increased C10E3 separation upon nonane loading is also reflected in increasing bilayer thickness, t_{bil}, as is shown in Table 6.3. Note that, in the present work, bilayer thickness is taken as two times the length corresponding to the midpoint between the 10 and 90% points along the water density profile. Water, rather than C10E3, is used to define the location of the interface to avoid inconsistency in computation introduced if/when solutes intercalate into the surfactant head-groups, and/or when an uneven number of surfactant molecules are contained in each leaflet. Small-angle neutron scattering experiments of a C10E3/water lamellar system at comparable conditions yield a bilayer thickness of 27.6 Å, based on the change of interlayer spacing as C10E3 volume fraction is varied [252]. Considering the multiple ways to define bilayer thickness from simulations data (*i.e.*, based on the solvent or C10E3 profile, using a 10/90% or Gibbs dividing surface definition of interface location, etc.), as well as the assumptions that enter in the

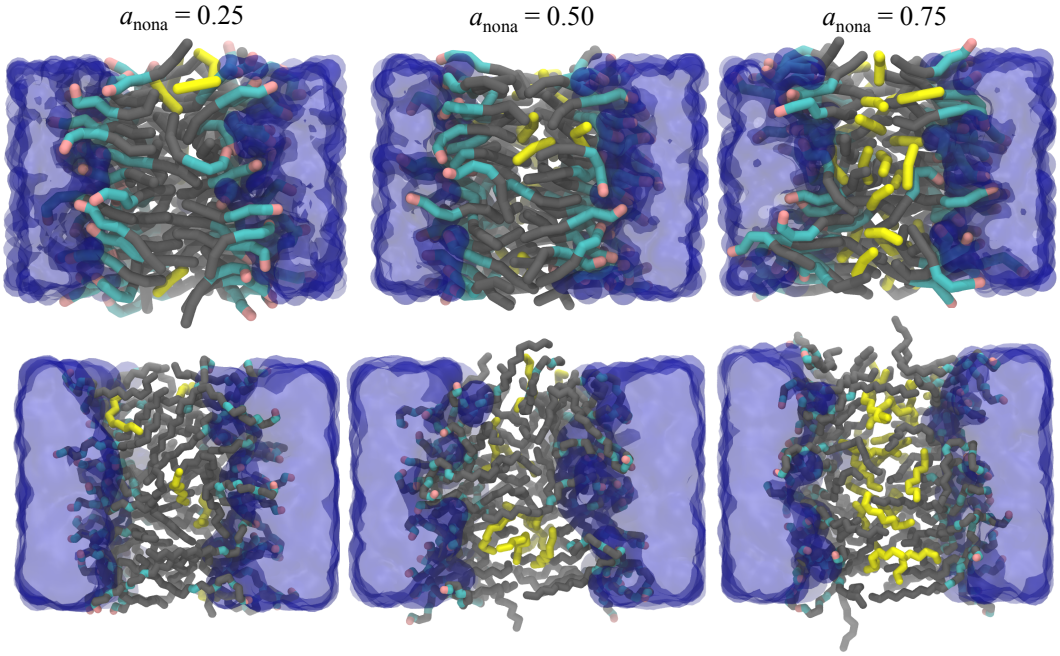

Figure 6.3: Snapshots of nonane-containing bilayer systems at varied a_{nona}, in CG (top) and UA (bottom) representations. Water is rendered as a blue transparent surface. For CG representations, alkyl, ether, and methoxy beads are given in gray, cyan, and pink, respectively, while solute alkyl groups are given in yellow for clarity. In the UA snapshots, CH_x, oxygen, and hydrogen units are given in gray, cyan, and pink, while solute CH_x beads are given in yellow.

experimental measurement (*i.e.*, that bilayer thickness is independent of surfactant concentration), the experimental and simulation values are in good agreement. Head group area, A_{C10E3}, is found to undergo negligible changes, further demonstrating the tendency for nonane molecules to preferentially partition into the center of the bilayer, rather than to intercalate into the monolayer/water interface. Values for the CG and UA system bound the experimental value for generic $A_{\text{C}x\text{E3}}$ of 36.8 Å^2 [80].

The angle formed between a vector along the z-axis (perpendicular to the interface) and one formed between two atoms separated by two bonds (a "1–3" vector), can be utilized to compute an order parameter according to $S_{1-3} = 0.5\langle 3\cos^2\theta - 1\rangle$, where the average is over the 1–3 vector of interest, across all molecules. This value, which can also be determined experimentally, varies between 1 and -0.5, corresponding to alignment perpendicular to

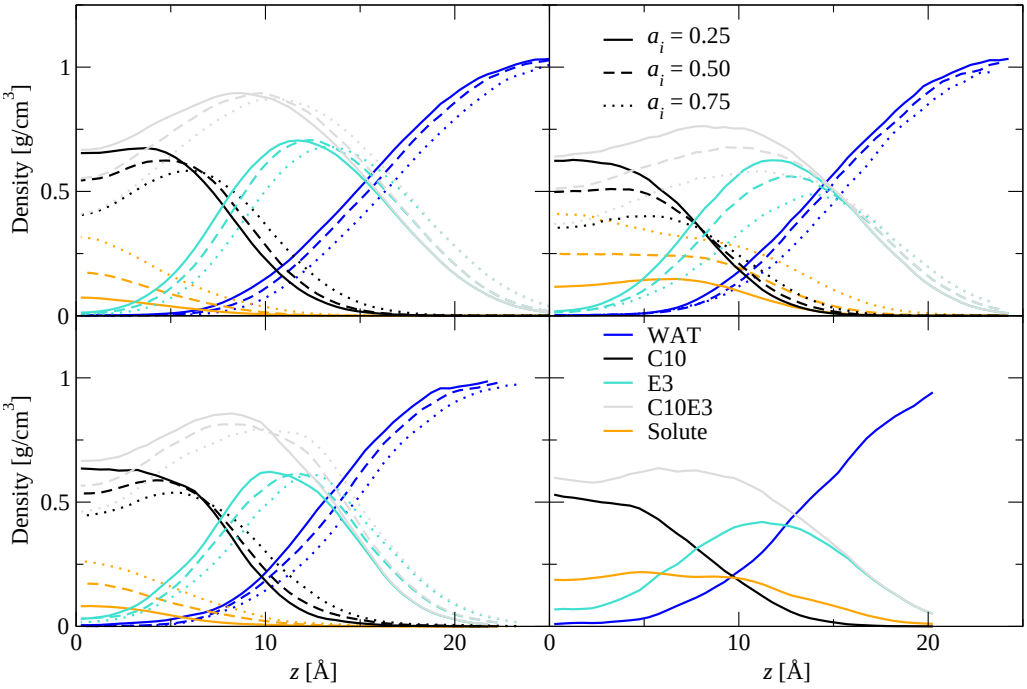

Figure 6.4: Bilayer density profiles. The top and bottom rows correspond to CG and UA systems, respectively, and the left and right columns correspond to nonane and hexanol containing systems, respectively. Colors and line styles are described in legends. Solutes and the full C10E3 molecule are given in orange and gray, respectively.

and parallel to the interface respectively, and is zero for a randomly ordered system. The average of all S_{1-3} along molecules yields the average order parameter, $\langle S \rangle$, and provides a description of overall chain order. Figure 6.5 demonstrates how S_{1-3} varies from the terminal alkyl portion of the surfactant to the polar end. As the interior of the chain is approached from the alkyl terminus, the preferred orientation becomes more perpendicular with respect to the interface. A maximum is reached near the 6$^{\text{th}}$ vector (counting from zero) From there, the chains displays more disorder, stabilizing around a value of 0.1. The last 1–3 vector, which corresponds to the terminal C–O–H group, exhibits no orientational preference, on average. In general, Figure 6.5 and the $\langle S \rangle$ values given in Table 6.3 show that changes in order upon loading of nonane are negligible. This result is somewhat expected as nonane is found to primarily load between leaflets, near the relatively disordered part of

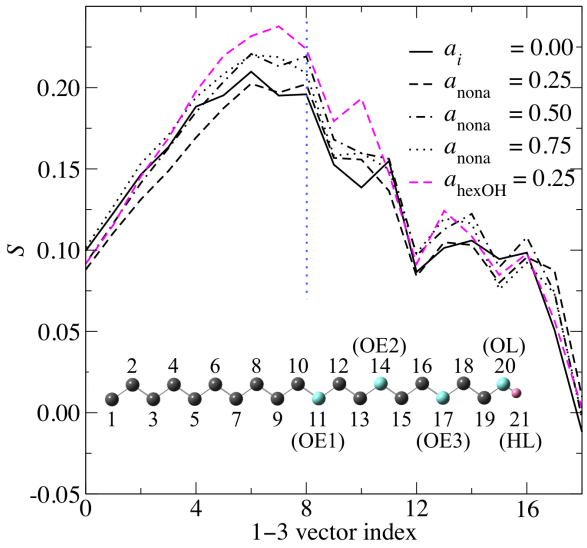

Figure 6.5: Order parameter moving down the UA C10E3 molecule, where the x-axis gives the index of the vector formed by the i^{th} and $(i+2)^{\text{th}}$ backbone atoms. Line colors and style are described in the legend. The vertical blue line at $x = 8$ indicates the index of the order parameter centered on the α carbon. The inset figure shows a C10E3 molecule with atoms indexed, where gray, cyan, and pink is utilized to represent CH_x, oxygen, and hydrogen atoms. Names for oxygen and hydrogen atoms are also given in the inset figure, where "E" and "L" refer to ether and hydroxyl types, respectively.

the surfactant alkyl tail. The presence of nonane is found to lead to slightly more extended chains, as is shown by increased surfactant end-to-end distances, r_{ete}, upon nonane loading. This extension, which only becomes appreciable at $a_{\text{nona}} = 0.75$, is likely caused by the ability of loaded nonane molecules to "screen" surfactant tails from alkyl/ether interactions between opposing leaflets.

A similar S shape is observed experimentally for CxEy lamellar surfactants, where order is shown to increase when moving from the alkyl terminus to the vicinity of the polar/non-polar interface within the molecule, and diminishes towards zero at the terminal polar portions; however, the values along the chain vary significantly with surfactant composition. Experimentally measured values for S_{1-3} at the α carbon position (*i.e.* atom number 10 in Figure 6.5) range from ≈ 0.085 for C12E5 surfactants to ≈ 0.11 for C12E4 surfactants [253, 254, 255]. Generally, an increase in order is observed when decreasing the length of either

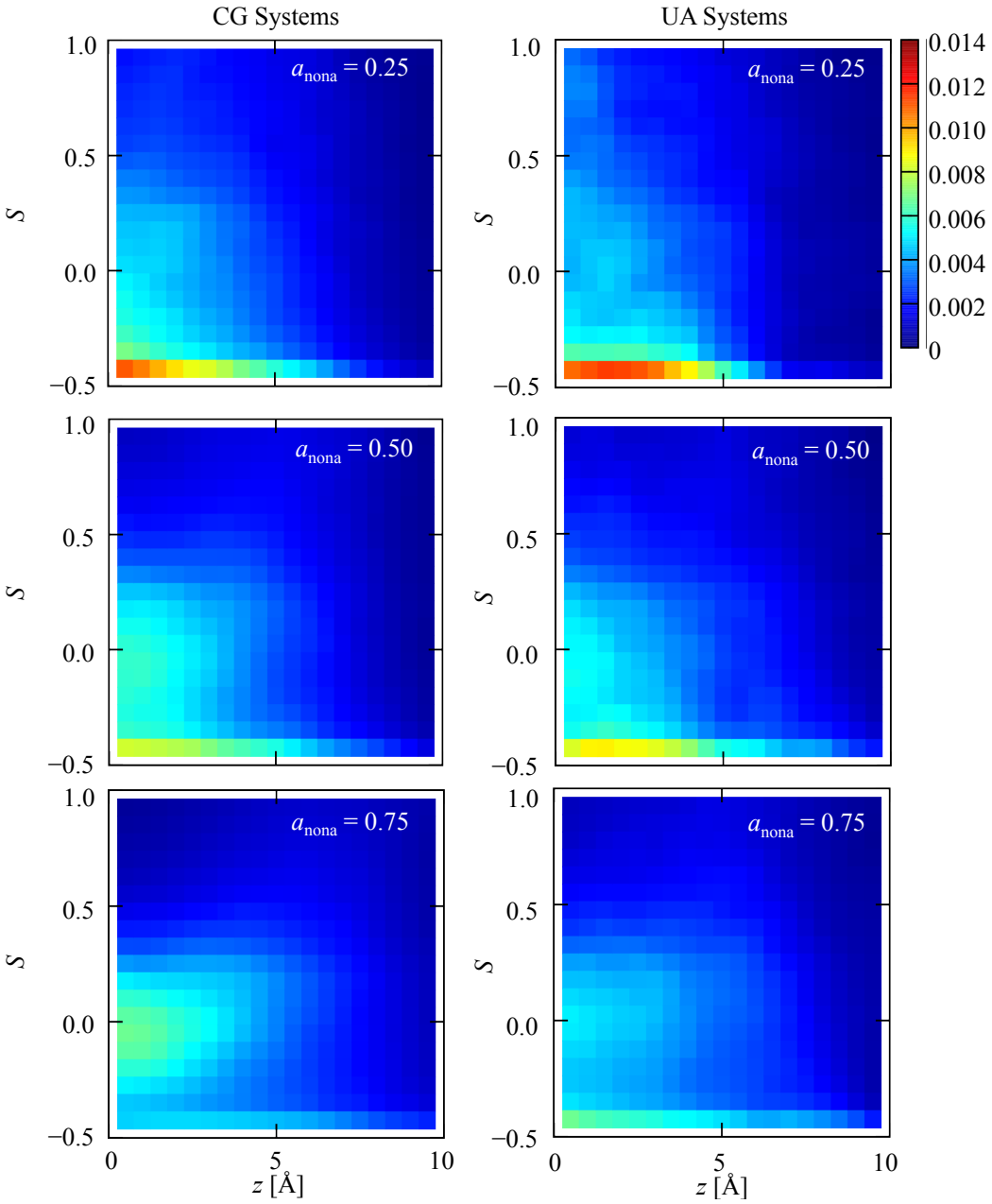

Figure 6.6: Heat maps showing the probability to find a nonane solute with a specific z-COM exhibiting a given order parameter (S), at a_{nona} of 0.25, 0.50, and 0.75.

the head or tail group, which is consistent with what is seen in the present work, where S_{1-3} values at this location are larger than those mentioned above. Other features that can increase ordering include the presence of curved interfaces and ionic surfactants.

The order parameter distribution between nonane end-to-end vectors and the interface normal was also evaluated, as a function of solute z-center-of-mass position (z-COM). The results, shown in Figure 6.6, indicate that nonane exhibits very similar behavior in both the UA and CG systems. At $a_{\text{nona}} = 0.25$, solutes prefer to orient parallel to the interface; as a_{nona} increases, this preference becomes less prominent, with disorder ($S = 0$) emerging as a secondary location of moderate probability, at higher a_{nona}. The emergence of disorder at the interior region of the bilayer is likely caused by crowding of nonane at higher loadings. The primary difference between the CG and UA heat maps is the stronger preference for the parallel orientation that is exhibited by the UA systems. This higher preference is likely due to the relatively larger ratio of A_{C10E3} to t_{bil} for the UA system, meaning that solutes have more room to orient themselves parallel to the interface than in the CG system. It should be noted that, while non-zero probabilities exist for z-COM values approaching 10 Å, the density profile in Figure 6.4 indicates that this is not a common location for the solute.

Hydrogen bonding to and from the surfactant for nonane containing systems is provided in Figure 6.7. The criteria for a hydrogen bond [266] is an O$\cdots$H distance $\leq$ 2.5 Å, an O$\cdots$O distance $\leq$ 3.3 Å, and an O$\cdots$H–O angle cosine ≤ -0.1. A trend in the C10E3 oxygen data emerges, where hydrogen bonding (particularly to water) increases when moving from the "OE1" oxygen to the "OL" oxygen for any given system type, which can be attributed to the proximity of the atoms to the water slab. Hydrogen bonding with the surfactant is not found to vary with nonane loading, consistent with the tendency of nonane to load between surfactant leaflets, rather than in the vicinity of the surfactant/water interface.

There has been significant debate surrounding the structure of water in the vicinity of ether groups in molecules like C10E3 [87, 89, 90, 91, 256]. Many experiments have attempted to elucidate this structure, however there are diverging opinions, with numbers ranging from 3 to 9 bound water molecules per surfactant. Furthermore, different measuring techniques (*i.e.*, NMR, FT-IR, etc.) define "bound" water differently, which can contribute further to inconsistent answers. Additionally, these studies do not distinguish how these water molecules are distributed across the surfactant. The hydrogen bond analysis in the

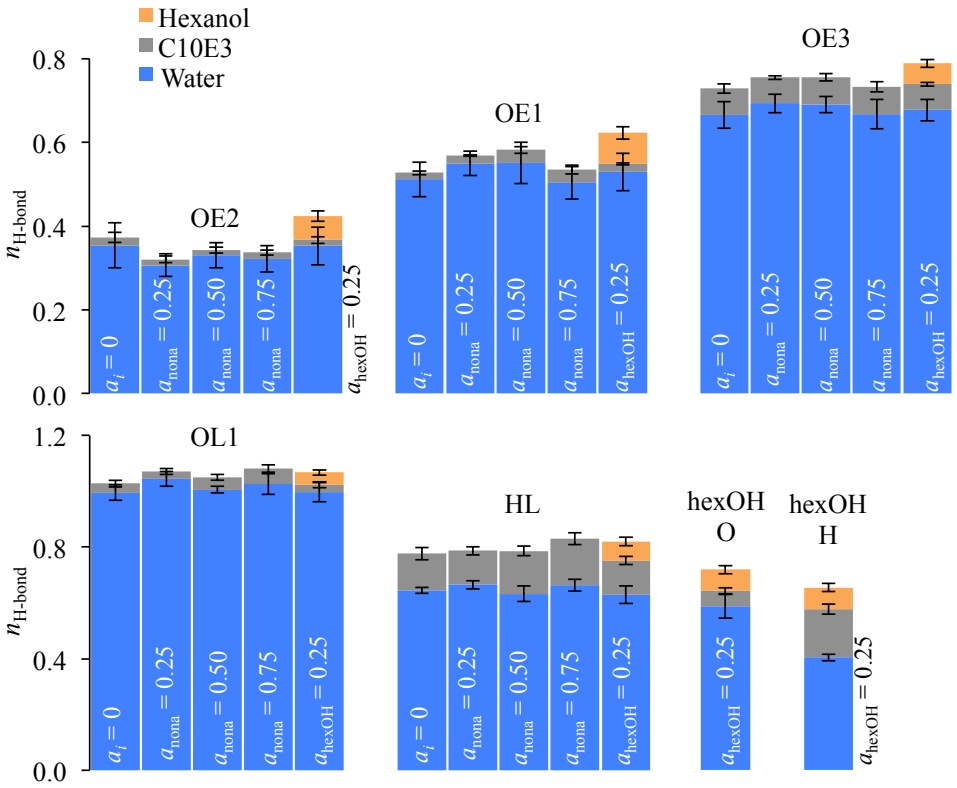

Figure 6.7: Hydrogen bonding in the UA systems to C10E3 from C10E3, water, and hexanol. Atom names are given above each cluster of bars and correspond to the naming convention given in Figure 6.5. The simulation type corresponding to each set of data is inset within each bar for all but OE1 with $a_{\text{hexOH}} = 0.25$, and hexOH H with $a_{\text{hexOH}} = 0.25$, where labels are given to the right of the corresponding column.

present work also sheds light on the number of water molecules per surfactant. Here, we describe the number of bound water molecules as the sum of $n_{\text{H-bond}}$ for OE1, OE2, OE3, OL, and HL with water, and find a value of 3.2_1 for the system at $a_i = 0$, where the number of water molecules bound to atoms in C10E3 varies from 0.354_3 at OE1 to 1.00_1 at OL1, and 0.645_1 at HL1.

Table 6.4: Hexanol loading properties.

	CG			UA				
a_{hexOH}	loading $N_{\text{sol}}/N_{\text{C10E3}}$	A_{C10E3} [Å^2]	t_{bil} [Å]	loading $N_{\text{sol}}/N_{\text{C10E3}}$	A_{C10E3} [Å^2]	t_{bil} [Å]	r_{ete} [Å]	$\langle S \rangle$
0.00	0.0	34.8_6	28.2_3	0.0	41_1	24.7_3	15.86_5	0.13_1
0.25	0.36_1	39.2_2	27.9_2	0.9_1	51_2	25.2_6	16.33_3	0.14_1
0.50	0.76_4	43.8_4	29.3_3	–	–	–	–	–
0.75	1.28_5	48.7_5	30.4_3	–	–	–	–	–

6.4.2 Hexanol Loading

The hexanol-containing systems exhibit significantly different trends. Table 6.4 shows that, for both UA and CG systems, hexanol loads into the surfactant in substantially larger amounts compared to nonane. Furthermore, snapshots for these systems, provided in Figure 6.8, demonstrate that the mechanism of hexanol loading is significantly different than that of nonane. In both the UA and CG cases, hexanol exhibits water/surfactant interfacial activity, driven by the amphiphilic nature of the solute. The snapshots also indicate that, in the CG systems, water remains excluded from the bilayer interior, and hexanol is relatively disordered at all loadings, with hydroxyl/methoxy groups not showing a strong preference to point toward the water layers (particularly at high loading).

Similar to the nonane results, hexanol loading exhibits positive deviations from ideal behavior. Interestingly, other simulation work, which used a hybrid grand canonical Monte Carlo (GCMC) DPD approach found that certain model alcohols exhibit negative deviation from ideality [98]. In the aforementioned system, the alcohol solute was much smaller than the surfactant, containing only one polar and one non-polar bead, compared to the surfactant, which contained three polar beads and two non-polar tails, comprised of five beads each. The alcohol solute was found to sit almost exclusively at the interface, though alcohol molecules would occasionally diffuse into the water. The difference in loading trends between the present work and those coming from the GCMC/DPD simulations stem largely from the fact that small alcohols, such as those simulated in the GCMC/DPD work, have high interfacial activity and posses hydrophobic portions which are too small to contribute to bilayer stability. In contrast, the alkyl chain in hexanol is significantly longer, allowing

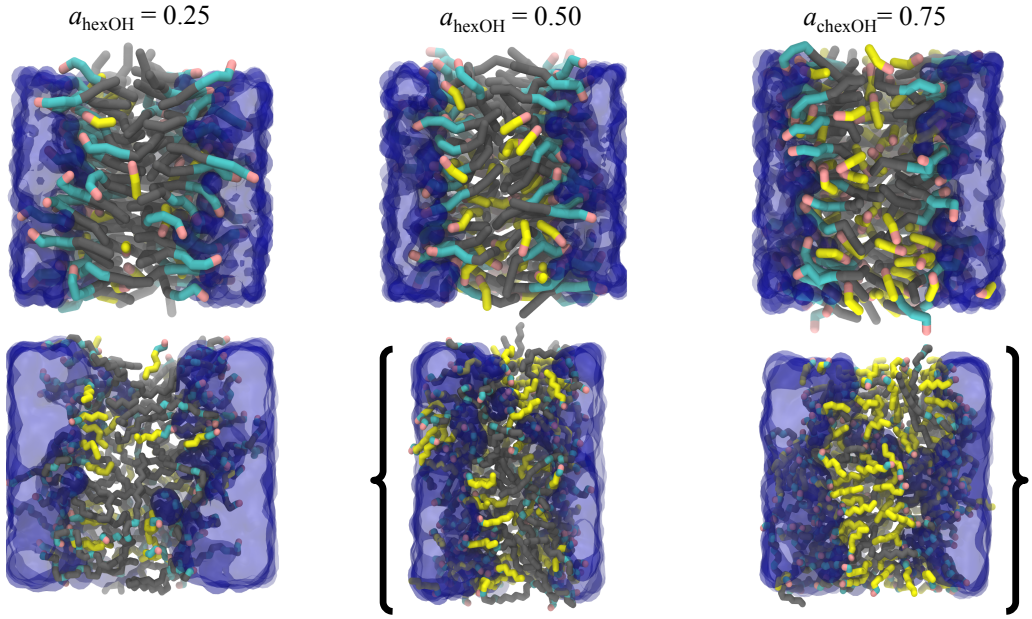

Figure 6.8: Snapshots of hexanol-containing bilayer systems at varied a_{hexOH} in CG (top) and UA (bottom) representations. Colors are as given in Figure 6.3. Systems within brackets were found to have become unstable.

the molecule to interact favorably with a larger portion of the surfactant. As a consequence, hexanol is found to distribute more homogeneously across the bilayer than the small alcohol studied in the GCMC/DPD work.

Different behavior is observed for the UA systems. At $a_{\mathrm{hexOH}} = 0.25$, the solvent shows higher penetration than the equivalent CG system, introducing a higher degree of disorder, and at higher a_{hexOH}, the system becomes unstable despite being below $a_{\mathrm{hexOH}} = 1$. Indicators of instability are shown in the snapshots of the hexanol containing UA systems at $a_{\mathrm{hexOH}} = 0.50$ and 0.75, which were taken at the point where solvent was found to bridge bilayer leaflets; in contrast to the system at $a_{\mathrm{hexOH}} = 0.25$, at higher a_{hexOH}, hexanol clusters begin forming which attract water to the interior region of the bilayer, attracting the polar portion of the surfactant into the interior. This result suggests a few possibilities, the first being that the UA model may predict that the 50/50% volume mixture of water/surfactant studied here is closer to a phase boundary than the CG model. An alternative possibility is that, for small, highly associating solutes which can act as secondary surfactants, the ability

of the solute to entrain water within the bilayer is enhanced and promotes instability. Thus, this behavior may not be observed in systems which do not strongly aggregate (*i.e.*, do not form hydrogen bonds). Finally, the differing results could be a consequence of system size effects.

Density profiles for hexanol-containing systems are provided in Figure 6.4, and demonstrate that the solute is distributed across the bilayer. In the CG systems at $a_{\mathrm{hexOH}} = 0.25$, a single small peak emerges at the point where the density profiles for the C10 and E3 segments of the surfactant intersect. As a_{hexOH} increases, a second hexanol peak emerges at the bilayer interior, and becomes larger than the interfacial peak at $a_{\mathrm{hexOH}} = 0.75$. The presence of hexanol most significantly perturbs the height of the surfactant density profiles; in contrast to the CG nonane systems and despite significantly higher loadings, surfactant monolayers are not pushed apart as strongly. This behavior, reflected in t_{bil} values which are equivalent or smaller for the hexanol-containing systems to those containing nonane, despite having relatively higher loadings, is largely due to the interfacial activity of hexanol; systems containing hexanol exhibit significantly larger A_{C10E3} values. When A_{C10E3} is increased, the volume in a given z-slice is increased, which leads to the decreasing overall surfactant density profiles shown for hexanol in Figure 6.4.

The density profiles for the UA systems provide a slightly different picture. Hexanol exhibits two small peaks, one at the C10/WAT interface, and the second at $z \approx 5$ Å. A more significant presence of the E3 portion of the surfactant at the bilayer interior are is also observed for the UA system. When compared to UA systems containing nonane at equivalent a_i, systems loaded with hexanol exhibit a slightly larger higher r_{ete}, but roughly equivalent $\langle S \rangle$. Figure 6.5 demonstrates that the only significant difference between S_{1-3} values for hexanol and nonane containing systems occurs for vectors which correspond to atoms 9 an 12, which are two atoms away from OE1 and OE2, respectively; note that atoms OE1 and OE2 are the surfactant oxygen atoms to which hexanol forms the largest number of hydrogen bonds, as is shown in Figure 6.7.

The order parameter distribution for hexanol in C10E3 is provided in Figure 6.9. Focusing first on the CG data at $a_{\mathrm{hexOH}} = 0.25$, two regions of relatively high probability density are apparent, the first at a constant value of $S = -0.5$, and the second emanating from $S = 0$ for $z = 0$, and going to slightly larger S values at larger z. As loading is

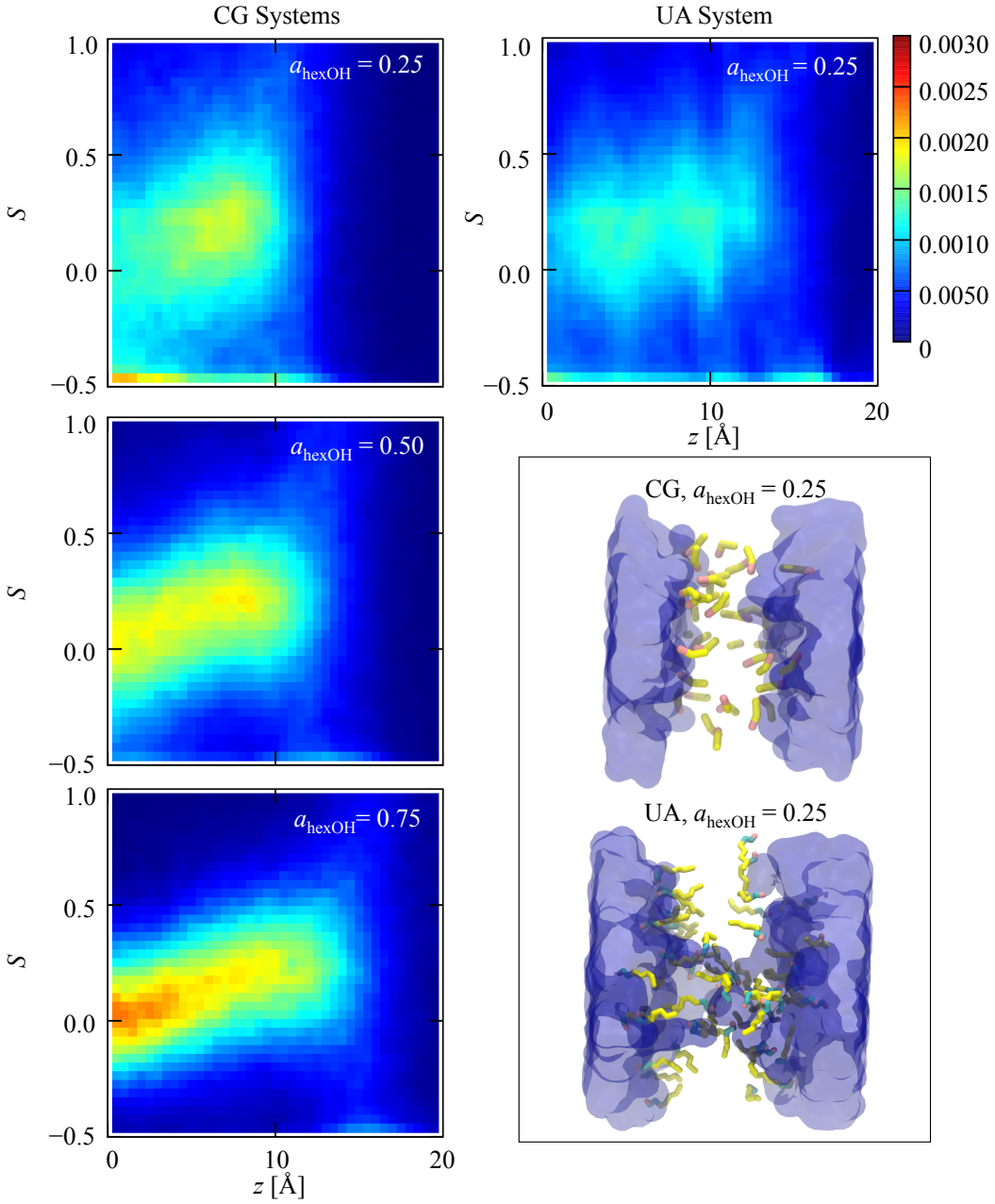

Figure 6.9: Heat maps showing the probability to find a hexanol solute with a specific z-COM exhibiting a given order parameter (S), at a_{hexOH} of 0.25, 0.50, and 0.75.

increased, the first region vanishes, while the second becomes less diffuse. As was seen for the nonane systems, at the bilayer interior, preference for the parallel alignment is replaced with a preference for disorder as a_{hexOH} is increased. Once again, this behavior is attributed to crowding at higher loadings. The larger tendency towards disorder observed for hexanol when compared nonane can likely be explained by the tendency of hexanol to load in greater quantities than nonane at any given activity as well as the ability of hexanol to form clusters, leading more significant disorder. The small but non-zero tendency for hexanol to load in arrangements parallel to the interface in the vicinity of of water is somewhat unintuitive, but snapshots for the CG and UA systems at $a_{\mathrm{hexOH}} = 0.25$ show that these molecules are indeed present, and generally are those which do not have hydroxyl groups embedded in the water. Instead, the hydroxyl groups of these molecules point towards surfactant polar groups or other hexanol molecules in the vicinity. The order parameter distribution for UA hexanol at $a_{\mathrm{hexOH}} = 0.25$ exhibits the same two regions as for the CG system, though they are less defined.

Hydrogen bond analysis for the surfactant in the hexanol containing system, shown in Figure 6.7, indicates that overall hydrogen bonding between the surfactant and water is comparable to that observed for the bare and nonane containing systems. The presence of hexanol does not modify existing H-bonds to C10E3, but rather adds to what is already present. The largest number of hydrogen bonds between hexanol and the surfactant are observed for "OE1" and "OE2," which are closest to the bilayer interior and thus experience the least competition with water. Hydrogen bond analysis of the hexanol molecule indicates that the largest fraction of H-bonds to hexanol oxygen atoms come from water, with the remainder divided nearly equivalently between the surfactant and hexanol. In contrast, hexanol donates more H-bonds to itself, rather than C10E3, though the largest amount are donated to water. These data agree with density profiles, which show that hexanol is distributed throughout the bilayer

6.5 Conclusions

Monte Carlo simulations in the osmotic Gibbs ensemble have been conducted to study nonane and hexanol loading and structure in a lamellar C10E3/water system. Overall, solute loading was not found to have a large influence on chain ordering, however the

solutes themselves exhibited a greater preference for disordered arrangements at the bilayer interior as loading was increased. Nonane loading was shown to exhibit positive deviations from ideal loading, and solutes exhibited a preference to load between monolayers, primarily perturbing the bilayer by increasing its thickness. At corresponding activities, hexanol was found to yield much higher loading, and undergo uptake in a substantially different manner. The amphiphilic nature of the solute lead to interfacial activity, and thus surfactant head group area was found to increase significantly with hexanol loading. Hexanol was also found to exhibit a more significant presence near the interface at higher loading. The UA hexanol loading calculations were found to become unstable before $a_{\mathrm{hexOH}} = 1$. It was found that this instability was encouraged by formation of hexanol clusters between bilayer leaflets, which lead to increased water penetration and entrainment within aggregates, and inclusion of C10E3 in hexanol/water clusters. Differences in results between the CG and UA systems could be caused by differences in each model's phase diagram (*i.e.*, the UA system may be closer to a phase boundary), system size effects, or the ability of the UA system to form hydrogen bonds.

In general, however, the coarse-grained representation is in good agreement with the UA on the system on the relevant trends and features of the nonane-containing systems, including loading and structure. The CG force field also captures the enhancement of hexanol loading found by the UA simulations when compared to nonane, however, in general, does not agree as closely as nonane CG/UA systems. Nevertheless, the CG hexanol simulation data provide useful insights into loading characteristics of polar solutes.

Chapter 7

Simultaneous Loading of Non-Polar and Amphiphilar Solutes in Surfactant Bilayers

7.1 Introduction

In fields ranging from drug design [95, 248, 249] to home and personal care product formation[7, 267], solubilization of species in structured surfactant materials is an extremely important phenomenon. As such, numerous studies have been conducted with the aim of understanding and predicting solute loading and system structure. The majority of these studies rely on thermodynamic models and correlations to glean insights into such systems [94, 96, 97, 268]. Unfortunately, however, these models are not without short comings. For example, linear solvation free energy relationships (LSER's) require the solute to be at infinite dilution, which is seldom the case for practical application [94, 98], and thus, comes at the expense of understanding concentration dependent loading effects. Other thermodynamic models either rely on *a priori* knowledge of system structure, using the location of the solute within the system to determine appropriate contributions to solubilization free energy [93], or attempt to accomplish the reverse, using thermodynamics to extrapolate

The reference "[Ch.6]" refers to the contents of Chapter 6.

structural changes [94, 97, 268].

Experimental studies of polar solute loading in various trimethyl ammonium bromide and sodium dodecyl sulfate micelles, [94] for example, have yielded a somewhat idealized view of system structure upon loading. Based on the observation that transfer free energies for primary alcohols ranging in length from 1-butanol to 1-heptanol from water into micelles were more favorable than for *n*-dodecane, it was concluded that alkanols mainly locate at the surfactant interface. In contrast, molecular simulation studies of loading of 1-hexanol in lamellar mono-*n*-decyl ether (C10E3) [Ch.6] have yielded a more complex view of micellar structure, showing that primary alcohols can be distributed from the bilayer interior to the surfactant/water interface, showing only small enhancements near the interface, at low loadings. Other studies [269, 270, 271, 272, 273] have concluded that ordering of monolayer alkyl tails in bilayer systems is the primary reason that solute solubility can be substantially different than in bulk oil, based on thermodynamic measurements. While surfactant ordering may play a significant role, there are additional factors that can influence solute uptake. For example, it has been shown that entrainment of water within micellar structures can lead to supression of solute loading, even for solutes that can form hydrogen bonds [96].

Matters only become more complicated when additional components are introduced. Many studies [92, 96, 251] have shown that both enhancement and suppresion in drug loading is possible depending on the nature of the cosolvent, however, detailed structural analysis was not conducted, leading authors to simply conclude that use of cosolvent can alter micellar structure. Light scattering studies of drug solubilization in oil-in-water (o/w) type micelles [96, 274, 275, 276] have demonstrated that, when oils are small and interfacially active, drug solubilization is small. From this behavior it was concluded that these small oils tend to exist at the surfactant/water interface rather than micelle interior, where the drug molecules would prefer to partition. Larger oils, on the other hand, were shown to partition more considerably into the micelle interior, and thus lead to enhancement in drug solubility. Synergistic loading (*i.e.* enhanced solubility of a given compound type in the presence of an additive) has also been observed when non-polar solutes are loaded simultaneously with quadrupolar solutes in various ionic surfactants [268]; experiments studying the uptake of hexane in the presence of benzene lead the authors to the conclusion that, contrary to the

earlier belief that benzene loaded exclusively at the micelle core, at least some benzene is present at the interface [268].

Molecular simulation provides a straight-forward and complementary means to investigate fundamental aspects of the aforementioned phenomena, particularly because the coordinates of all molecules within the system are a direct observable. Indeed, a significant number of simulation studies have been conducted for solute/surfactant systems [114, 116, 118, 260, 277] yielding valuable insights into system structure and stability, and even using results to inform thermodynamic-based solubilization predictions [93]. For example, simulations studying the loading of coarse-grained, small alcohols (intended to represent ethanol) in non-ionic surfactant bilayers confirm that sufficiently small interfacially active molecules can indeed preferentially locate at the surfactant water interface [98]. Others have studied the influence of small quantities of polycyclic aromatic hydrocarbons on the structure and dynamics of ionic surfactant micelles, and found that micellar structure was largely invariant to the presence of solutes, and that solutes spent time in both the interior and surface regions of the micelle [278]. In the present work, coarse-grained osmotic Gibbs ensemble configurational-bias Monte Carlo simulations are leveraged to study synergistic loading of 1-hexanol and *n*-nonane in a lamellar C10E3 and water system. Detailed structural analysis is provided along with discussion of possible factors contributing to synergistic and competitive loading.

7.2 Simulation Details

The loading simulation setup involves use of three boxes which are in thermodynamic equilibrium with one another but do not share explicit interfaces. The first box contains a "unit cell" of the lamellar system, comprised of a 50/50 weight percent mixture of triethylene glycol mono-*n*-decyl ether (C10E3)/water, while the remaining two boxes contain gas phase nonane and hexanol, respectively. Nonane and hexanol molecules are allowed to transfer between their respective gas phase and the lamellar C10E3/water box, thus, by controlling the pressure of each solute reservoir box, one controls the effective concentration of solute in the surfactant system. All simulations in the present work are conducted through the use of the MCCCS-MN software suite [176] and the Shinoda-DeVane-Klein (SDK) coarse-grained force field [109, 110].

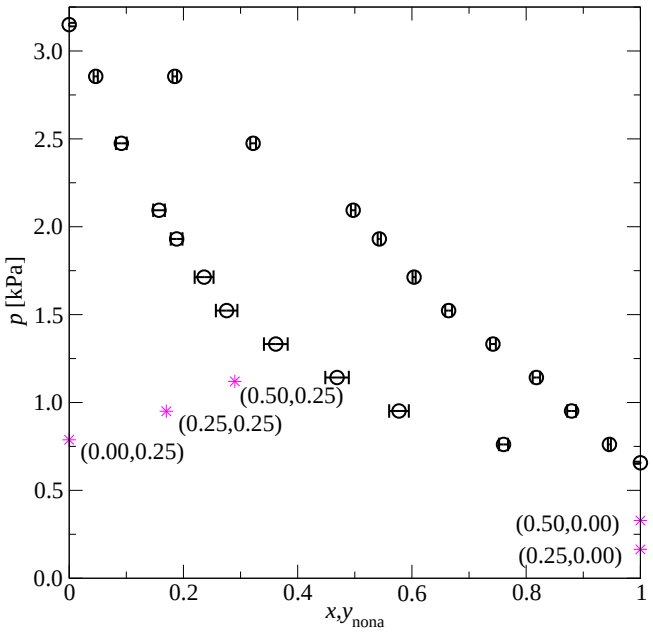

Figure 7.1: Circles: binary VLE computed for hexanol and nonane mixtures at 300 K using the SDK [109, 110] force field. Data for nonane molfractions of 1 and 0 are taken from previous work [Ch.6]. Asterisks: $(a_{\mathrm{nona}}, a_{\mathrm{hexOH}})$ values investigated in the present work. Data for simulations at (0.25, 0.00), (0.50, 0.00), and (0.00, 0.25) are also taken from previous work [Ch.6].

7.2.1 Support Calculations

As described previously [Ch.6], the pressure of each solute reservoir is controlled in terms of a scaling value, $a_i = p_{i,\mathrm{res}}/p_i^*$, where $p_{i,\mathrm{res}}$ is the solute reservoir pressure and p_i^* is the vapor pressure of solute i, predicted by the simulation force field; vapor pressures for SDK hexanol and nonane are taken from previous work [Ch.6]. In the case where two solutes are utilized, such as in the present work, one also needs to know the phase diagram for the binary solute system system, in order to ensure that the solute matrix is in the vapor phase region of the phase diagram at the given reservoir a_i values. Note that the notation $(a_{\mathrm{nona}}, a_{\mathrm{hexOH}})$ will be used henceforth to describe the a_i values for simulation, where subscripts "nona" and "hexOH" refer to the solutes nonane and hexanol, respectively.

Binary VLE data are obtained through 2-box Gibbs [107, 108] NpT ensemble simulations at 300 K and pressures ranging between the previously computed vapor pressures of nonane and hexanol, determined by the SDK force field [Ch.6]. A total of 2000 molecules are used, where the ratio of nonane to hexanol molecules is adjusted at each pressure to ensure an adequate amount of molecules in the vapor phase. Volume move probability is set to $2/N_{\text{molec}}$ where N_{molec} is the total number of molecules in the system, resulting in approximately one accepted move per cycle. Swap move probabilities were set to 0.3, and the remaining probability is distributed across configurational bias [104, 105, 106, 134], translation, and rotation moves according to the relevant number of degrees of freedom. At least 50,000 cycles of equilibration is performed at each pressure, where 1 cycle is equivalent to N_{molec} steps, and production was run for at least 200,000 additional cycles. Reported errors correspond to the standard error of the mean over eight independent simulations. The resulting binary phase diagram is given in Figure 7.1, along with the location of $(a_{\text{nona}}, a_{\text{hexOH}})$ combinations discussed in the present work.

7.2.2 Surfactant Containing Systems

Solute loading calculations are performed through the use of configurational-bias Monte Carlo simulations in the osmotic $N_{\text{wat}}N_{\text{C10E3}}p_{\text{bil}}\mu_{\text{nona}}\mu_{\text{hexOH}}T$ Gibbs ensemble, where N, μ, and T correspond to number of molecules, chemical potential, and temperature, and the subscripts "C10E3," "bil", and "wat," refer to the C10E3 surfactant, the bilayer simulation system and water, respectively. The surfactant containing box is held at 300 K and 1 atm, where the 50/50 weight percent C10E3/water system has been shown to form a stable lamellar phase [83] [Ch.6], while the solute reservoir boxes are at 300 K and a given $(a_{\text{nona}}, a_{\text{hexOH}})$. With the exception of the additional solute reservoir box, simulations are conducted using protocols described previously [Ch.6]. To briefly summarize, 8 to 16 independent simulations are conducted at each $(a_{\text{nona}}, a_{\text{hexOH}})$, where two independent simulations are initialized with a slab containing a mixture of 50 molecules of each solute type, between bilayer leaflets, and the remaining two independent simulations are initialized with no solutes in the bilayer. Once equilibrated through translational, rotational, configurational-bias, and volume moves, solutes are allowed to transfer between boxes. Solute transfer is achieved through a combination of molecule swap and particle identity

Table 7.1: Bilayer properties.

a_{nona}	a_{hexOH}	source[a]	$N_{\text{nona}}/N^b_{\text{C10E3}}$	$N_{\text{hexOH}}/N^b_{\text{C10E3}}$	A_{C10E3} [Å^2]	t_{bil} [Å]
0.00	0.00	previous	–	–	34.8_6	28.2_2
0.25	0.00	previous	0.066_1	–	34.8_1	29.1_1
0.50	0.00	previous	0.156_5	–	35.2_1	30.4_1
0.00	0.25	previous	–	0.36_1	39.2_2	27.9_2
0.25	0.25	present	0.082_3 (25_5 %)	0.39_1 (10_5%)	39.9_2	29.8_2
0.50	0.25	present	0.21_1 (35_{10}%)	0.41_1 (15_5%)	40.0_3	31.4_2

[a]Previous data are taken from [Ch.6]

[b]Values in parenthesis give the percent enhancement compared to systems with a single solute type, rounded to the nearest 5. Subscripted values give the error in the average enhancement, also rounded to the nearest 5.

exchange [66, 131, 132] moves. Particle exchange moves involve the use of two "impurity" molecules for each solute type, which represent one- and two-thirds of the full solute, respectively. A uniform biasing potential is applied to each impurity in the C10E3/water box to ensure an even distribution between condensed and vapor phases [66]. Simulations are considered equilibrated once loading trajectories for each independent simulation converge. Upon equilibration, production is run for an additional 10^6 cycles. Unless otherwise stated, errors correspond to the 95% confidence interval, across independent runs. Simulations consist of 100 C10E3 molecules, 1614 water molecules (538 coarse-grained beads), and 200 molecules of each solute type.

7.3 Results

Data related to solute loading are provided in Table 7.1 and demonstrate that there is significant loading enhancement of at least one solute, at both $(a_{\text{nona}}, a_{\text{hexOH}})$ values investigated. Experimentally, synergistic loading has been observed for hexane in cetylpyridinium chloride/water spherical micelles in the presence of benzene [268]. The curved nature of the cetylpyridinium chloride/water spherical micelles makes it such that significant uptake of solutes which do not lower surface tension requires an increase in micellar interfacial area,

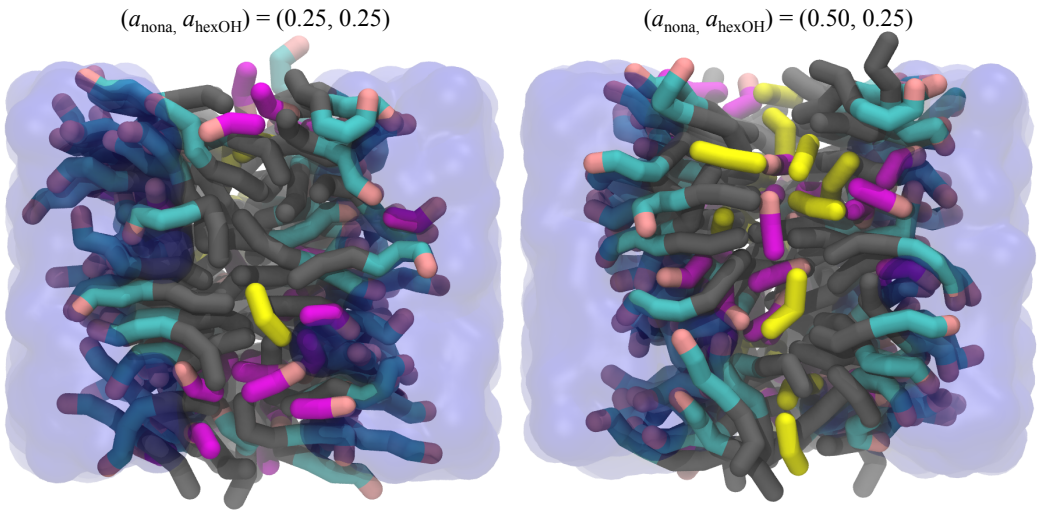

Figure 7.2: Snapshots of systems containing binary solutes. Water is represented as a transparent blue surface while gray, cyan, and pink units correspond to alkyl, ether, and methoxy groups, respectively. For clarity, alkyl groups in nonane are given in yellow while alkyl groups in hexanol are magenta.

which has an associated free energy penalty. Based on thermodynamic arguments, the authors suggests that, in order for hexane uptake to be enhanced, benzene must be present at the interface to lower interfacial tension and thereby allow for creation of addition surface area [268]. In the present work, however, the surfactant/water interface does not exhibit curvature on average, thus the aforementioned arguments do not necessarily apply.

To understand the effects motivating solute loading enhancement or the current system, structure is investigated. Snapshots for systems containing binary solutes are given in Figure 7.2, and show that solutes load in an intuitive fashion; consistent with what is observed for loading in the $(a_{\text{nona}}, a_{\text{hexOH}}) = (0.25,\ 0.00)$, $(0.50,\ 0.00)$, and $(0.00,\ 0.25)$ systems [Ch.6], nonane appears to preferentially locate in the bilayer interior while hexanol is distributed from the bilayer interior to the interface. The system density profiles given in Figure 7.3 along with the surfactant head group areas (A_{C10E3}) and bilayer thicknesses (t_{bil}) provide a quantitative view of structural changes. Focusing first on the surfactant density profiles, the following can be seen: at $(a_{\text{nona}}, a_{\text{hexOH}}) = (0.25,\ 0.25)$, both the non-polar (C10) and polar (E3) portions of the surfactant exhibit a decrease in density when

compared to the profiles for the "unary" solute cases at $(a_{\text{nona}}, a_{\text{hexOH}}) = (0.25, 0.00)$ and $(0.00, 0.25)$. Concomitantly, the density of the C10 profile exhibits a small dip at $z = 0$, indicating increased separation of monolayer leaflets, while the E3 profile becomes slightly more narrow. The overall decrease in the C10E3 profile can be attributed to a combination of increased A_{C10E3}, compared to (0.25, 0.00), and an increase in the amount of solute present within the bilayer compared to either of the unary cases. Bilayer thickness is not found to change significantly for the $(a_{\text{nona}}, a_{\text{hexOH}}) = (0.25, 0.25)$ case. Surfactant density profiles for the system with $(a_{\text{nona}}, a_{\text{hexOH}}) = (0.50, 0.25)$ exhibit the same trends, including an equivalent A_{C10E3}, though t_{bil} is found to be slightly greater, due to the larger quantity of solubilized compounds. Despite significant changes in the density profiles for the systems containing a binary solute matrix when compared to those with unary solutes, Figure 7.4 indicates that the overall system density profile exhibits minimal differences between the various $(a_{\text{nona}}, a_{\text{hexOH}})$ values; changes in height and location of C10E3 profiles are largely compensated by the presence of additional solutes, and subsequent changes in A_{C10E3}.

Figure 7.5 provides the density profiles for solutes in each system. The increase in resolution afforded by the change in range of the y-axis allows for the subtle differences between solute profiles to be observed. It is immediately evident that, other than the height of the nonane density profiles, no significant changes of shape occur when moving from systems with a unary solute matrix to those with binary solutes. Changes in the hexanol density profile at $(a_{\text{nona}}, a_{\text{hexOH}}) = (0.25, 0.25)$ are small, and primarily indicate a slight shift towards the surfactant/water interface. The shift in hexanol density profile is more substantial for the $(a_{\text{nona}}, a_{\text{hexOH}}) = (0.50, 0.25)$ case, however, for both systems with a binary solute matrix, a slight decrease in hexanol density is observed at between $z \approx 3$ and 7 Åcompared to the case with $(a_{\text{nona}}, a_{\text{hexOH}}) = (0.00, 0.25)$.

Another useful quantity to consider is the solute order parameter, S_i:

$$S_i = 0.5\langle 3\cos^2\theta - 1\rangle, \tag{7.1}$$

where θ is the angle formed between the solute end-to-end vector and the interface normal, and the angled brackets indicate an average over all molecules of type i. A values of 1 indicates perpendicular alignment with respect to the interface, while values of -0.5 and zero indicate parallel alignment, and no preferential orientation, respectively. Figures 7.6 and 7.7 provide heat maps which indicate the probability for a solute of a given type to be at

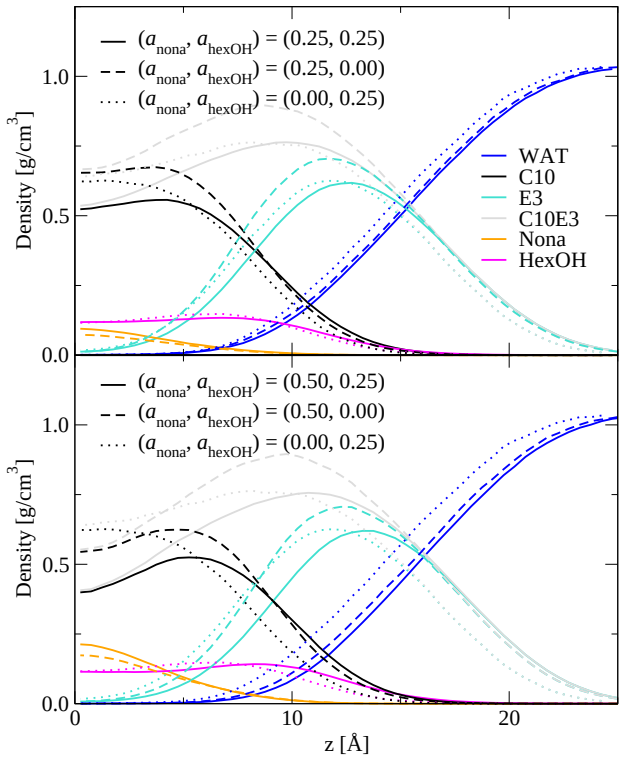

Figure 7.3: System density profiles. Data for systems containing a single solute type are adapted from [[Ch.6]. Line colors given in the top legend correspond to both plots.

a given z-center-of-mass location, with a given S_i value. Both figures indicate that changes are minimal for systems with binary solutes, when compared to systems with unary solute matrices. A small decrease in preference for slightly perpendicular arrangements ($S \approx 0.2$) is observed for hexanol in both cases, when moving from systems with unary to binary solute matrices, while nonane exhibits a slightly stronger preference for disordered arrangements. Worded otherwise, the presence of additional solutes appears to only slightly influence S values for each solute, generally by decreasing the relative importance of order. These changes, albeit subtle, are not entirely surprising, as increasing the amount of solute can lead to more crowding within the bilayer system.

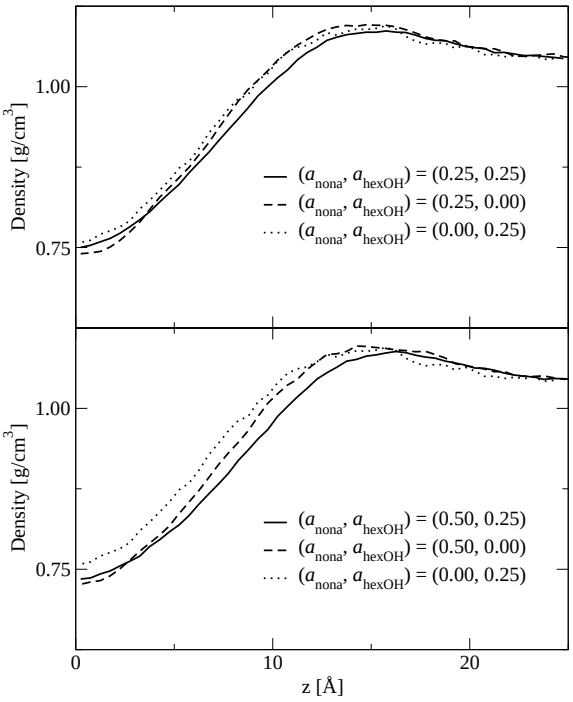

Figure 7.4: Total system density profiles, obtained by summing profiles for all solutes, surfactant, and water in a given system. Data for systems containing a single solute type are adapted from [Ch.6].

7.4 Discussion

Returning then, to the question of why synergistic loading is observed for hexanol and nonane in the present system, the above analysis has demonstrated that most changes in system structure upon introduction of additional solute types are subtle. Enhancement of nonane loading is most readily attributed to the increase in A_{C10E3} which occurs upon introduction of hexanol into the system; this increase leads to additional volume, particularly between bilayer leaflets, for nonane to partition into. This explanation is consistent with experimental studies of the factors influencing solute partitioning in lipid bilayer membranes [279]. The authors found that, regardless of how the bilayer membrane was perturbed (*i.e.* through changes in temperature, cholesterol length, or phospholipid length),

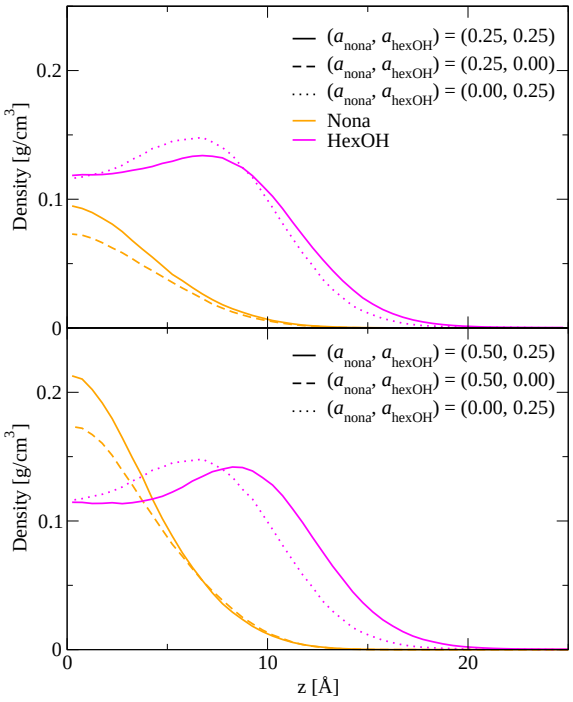

Figure 7.5: Solute density profiles. Data for systems containing a single solute type are adapted from [Ch.6]. Line colors given in the top legend correspond to both plots.

perturbations that resulted in lower head group areas lead to decreased solubilization of benzene, while changes that increased head group area resulted in increased solubilization. The authors attributed this behavior to corresponding changes in disorder of the phospholipid molecules [279], which is believed to modify the entropy associated with solute uptake [269, 270, 271, 272, 273], however the changes in A_{C10E3} dealt with in the present work are much smaller (≈ 5 compared to ≈ 15 Å^2) and are not found to significantly alter chain order [Ch.6].

The factors driving the small enhancements of hexanol in the $(a_{\mathrm{nona}}, a_{\mathrm{hexOH}}) = (0.25, 0.25)$ and (0.50, 0.25) cases are more elusive, however the observed changes in hexanol density profile offer a possible explanation. Hexanol, when part of a unary solute matrix, is found to distribute throughout the bilayer, at both the interior and interfacial regions. Intuitively, one might have expected the amphiphilar solute to exhibit a much stronger

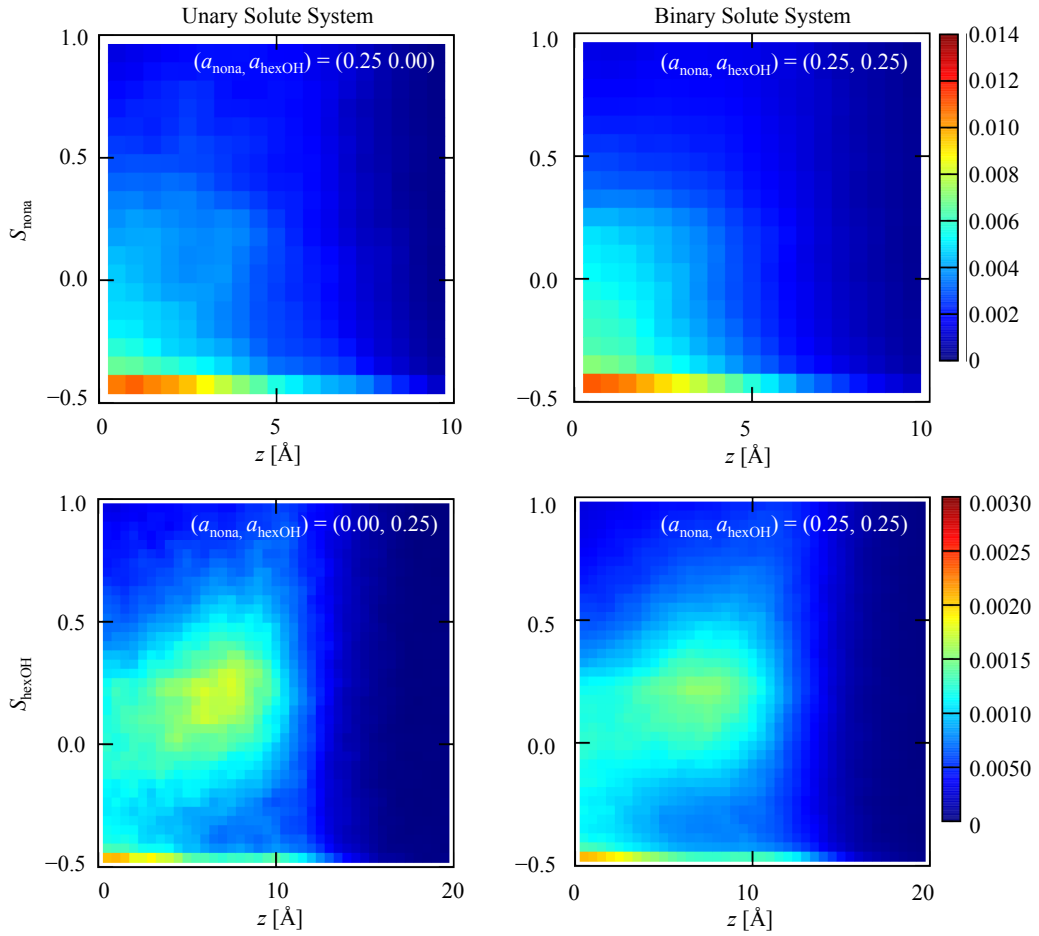

Figure 7.6: Heat maps comparing the system with $(a_{\text{nona}}, a_{\text{hexOH}}) = (0.25, 0.25)$ to those with (0.25, 0.00) and (0.00, 0.25), showing the probability for a solute of type i of given z-center-of-mass location, with a given S_i value. Data for systems containing a single solute type are adapted from [Ch.6].

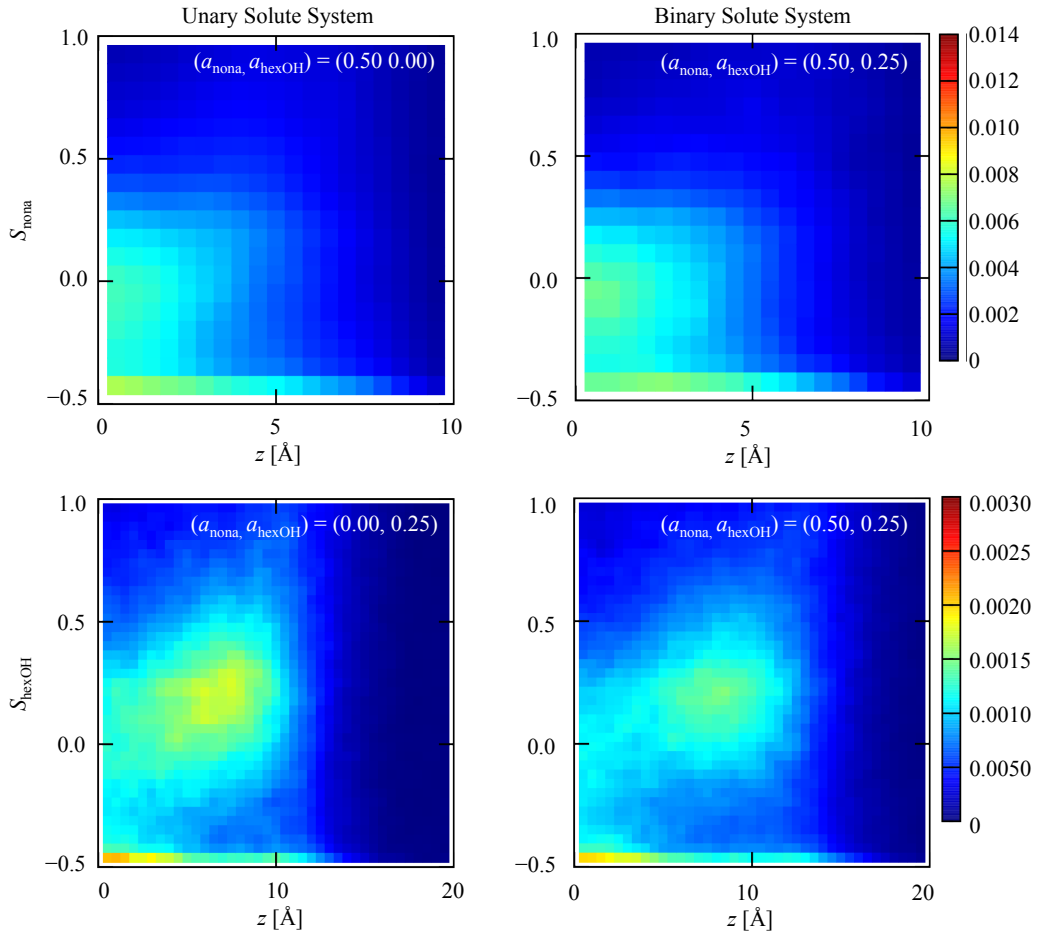

Figure 7.7: Heat maps comparing the system with $(a_{\text{nona}}, a_{\text{hexOH}}) = (0.50, 0.25)$ to those with (0.50, 0.00) and (0.00, 0.25), showing the probability for a solute of type i of given z-center-of-mass location, with a given S_i value. Data for systems containing a single solute type are adapted from [Ch.6].

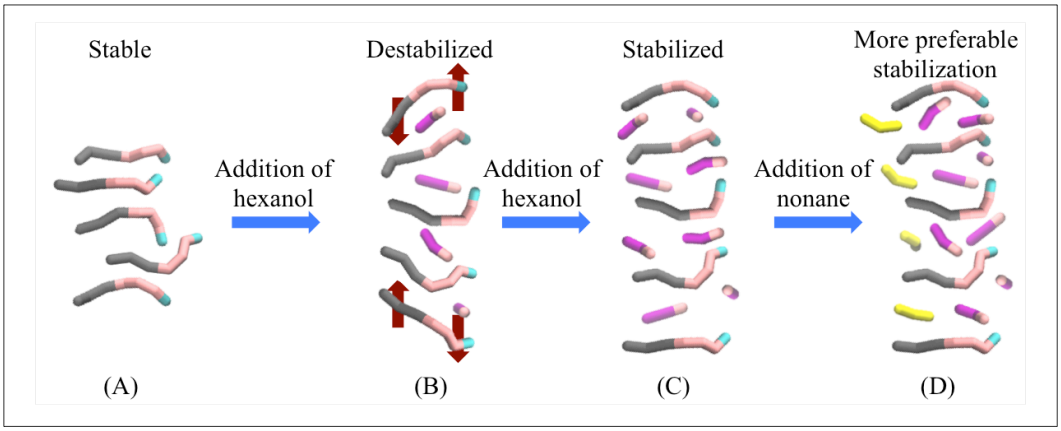

Figure 7.8: Cartoon depicting possible factors leading to enhancement in hexanol loading.

preference for an interfacial location, however images (A) and (B) of Figure 7.8 depict a possible factor contributing to the more homogeneous hexanol distribution seen in the present work. Image (A) represents a surfactant monolayer within a stable bilayer system. If hexanol were to exhibit a stronger preference to load at the interface, the surfactant head groups would be pushed apart by solute intercalation. A large enough increase in the space between surfactants would likely lead to interfacial curvature, as is depicted in image (B) of Figure 7.8, destabilizing the monolayer. This suggests that the need to offset bilayer curvature induced by increased head group area may contribute to the extent of hexanol partitioning at the bilayer interior, as depicted in image (C) of Figure 7.8. The shift of the hexanol density profile away from the bilayer interior in combination with the increased nonane density near $z = 0$ for the system at $(a_{\mathrm{nona}}, a_{\mathrm{hexOH}}) = (0.50,\ 0.25)$ compared to (0.00, 0.25), then, could be interpreted as competition for solubilization at the bilayer interior, however if this were the case, enhancements in loading of both solute types would be unexpected. Thus, it is more likely that observed changes in system structure and enhanced hexanol loading is due to contributions from nonane to stabilization of the bilayer at the interior region, freeing hexanol to load closer to the polar portion of the bilayer, as depicted in image (D) of Figure 7.8.

7.5 Conclusions

Monte Carlo simulations in the osmotic Gibbs ensemble were used to investigate simultaneous loading of hexanol and nonane in a lamellar C10E3/water system. Significant loading enhancements relative to systems containing a single solute type are observed for nonane at $(a_{\mathrm{nona}}, a_{\mathrm{hexOH}}) = (0.25, 0.25)$ and $(0.50, 0.25)$, and only the latter for hexanol. When loaded in tandem, characteristic perturbations of each solute type are observed (*i.e.* increases in t_{bil} by nonane and increases in A_{C10E3} by hexanol). The loading location of nonane is not significantly altered by the presence of hexanol, however hexanol exhibits a slightly higher tendency to load near the polar portion of the bilayer system in the presence of nonane. Changes in solute order are very small, and manifest as a subtle decreased in preference for ordered arrangements. The data suggest that enhancements in loading of nonane are due to increased bilayer volume, induced by the presence hexanol, and that the small enhancements for hexanol are possibly enabled by back-filling of nonane into the bilayer interior upon increased A_{C10E3} due to hexanol loading.

References

[1] Dorsey, J. G. and Dill, K. A. The Molecular Mechanism of Retention in Reversed-Phase Liquid Chromatography. *Chem. Rev. 89* **1989**, 331–346.

[2] Lewis, T., Faubel, M., Winter, B., and Hemminger, J. C. CO_2 Capture in Amine-Based Aqueous Solution: Role of the Gas-Solution Interface. *Angew. Chem. Int. Ed. 50* **2011**, 10178–10181.

[3] Shah, D. Improved Oil Recovery by Surfactant and Polymer Flooding. Elsevier **2012**.

[4] Weissbuch, I., Lahav, M., and Leiserowitz, L. Toward Stereochemical Control, Monitoring, and Understanding of Crystal Nucleation. *Cryst. Growth Des. 3* **2003**, 125–150.

[5] Tarascon, J.-M. and Armand, M. Issues and Challenges Facing Rechargeable Lithium Batteries. *Nature 414* **2001**, 359–367.

[6] Somasundaran, P., Chakraborty, S., Qiang, Q., Deo, P., Wang, J., and Zhang, R. Surfactants, Polymers, and Thier Nanoparticles for Personal Care Applications. *J. Cosmet. Sci. 55* **2004**, S1–S17.

[7] Lukowicz, T., Maldonado, R. C., Molinier, V., Aubry, J.-M., and Nardello-Rataj, V. Fragrance Solubilization in Temperature Insensitive Aqueous Microemulsions Based on Synergistic Mixtures of Nonionic and Anionic Surfactants. *Colloids Surf., A 458* **2014**, 85–95.

[8] Neue, U. D. HPLC Columns: Theory, Technology, and Practice. Wiley **1997**.

[9] Pasch, H. and Trathnigg, B. HPLC of Polymers. Springer Science & Business Media **1999**.

[10] Braithwaite, A. and Smith, J. Chromatographic Methods. Springer Science & Business Media **2012**.

[11] Carr, P. W., Li, J., Dallas, A. J., Eikens, D. I., and Tan, L. C. Revisionist Look at Solvophobic Driving Forces in Reversed-Phase Liquid Chromatography. *J. Chromatogr. 656* **1993**, 113–133.

[12] Carr, P. W., Tan, L. C., and Park, J. H. Revisionist Look at Solvophobic Driving Forces in Reversed-Phase Liquid Chromatography II. Comparison of the Behavior of Nonpolar and Polar Solutes. *J. Chromatogr. A 724* **1996**, 1–12.

[13] Carr, P. W., Tan, L. C., and Park, J. H. Revisionist Look at Solvophobic Driving Forces in Reversed-Phase Liquid Chromatography III. Comparison of the Behavior of Nonpolar and Polar Solutes. *J. Chromatogr. A 724* **1996**, 1–12.

[14] Park, J. H., Lee, Y. K., Weon, Y. C., Tan, L. C., Li, J., Li, L., Evans, J. F., and Carrg, P. W. Revisionist Look at Solvophobic Driving Forces in Reversed-Phase Liquid Chromatography IV. Partitioning vs. Adsorption Mechanism on Various Types of Polymeric Bonded Phases. *J. Chromatogr. A 767* **1997**, 1–10.

[15] Jaroniec, M. and Martire, D. A General Model of Liquid and Solid Chroamtography with Mixed Mobile Phases Involving Concurrent Adsorption and Partition Effects. *J. Chromatogr. A 351* **1986**, 1–16.

[16] Dill, K. A. The Mechanism of Solute Retention in Reversed-Phase Liquid Chromatography. *J. Phys. Chem. 91* **1987**, 1980–1988.

[17] Dorsey, J. G. and Dill, K. A. The Molecular Mechanism of Retention in Reversed-Phase Liquid Chromatography. *Chem. Rev. 89* **1989**, 331–346.

[18] Sentell, K. B. and Dorsey, J. G. Retention Mechanisms in Reversed-Phase Liquid Chromatography. Stationary-Phase Bonding Density and Partitioning. *Anal. Chem. 61* **1989**, 930–934.

[19] Sentell, K. B. and Dorsey, J. G. Retention Mechanisms in Reversed-Phase Chromatography: Stationary Phase Bonding Density and Solute Selectivity. *J. Chromatogr. A 461* **1989**, 193–207.

[20] Tijssen, R., Schoenmakers, P. J., Böhmer, M. R., Koopal, L. K., and Billiet, H. A. Lattice Models for the Description of Partitioning/adsorption and Retention in Reversed-Phase Liquid Chromatography, Including Surface and Shape Effects. *J. Chromatogr. A 656* **1993**, 135–196.

[21] Jaroniec, M. Partition and Displacement Models in Reversed-Phase Liquid Chromatography with Mixed Eluents. *J. Chromatogr. A 656* **1993**, 37–50.

[22] Tchapla, A., Heron, S., Lesellier, E., and Colin, H. General View of Molecular Interaction Mechanisms in Reversed-Phase Liquid Chromatography. *J. Chromatogr. A 656* **1993**, 81–112.

[23] Sander, L. and Wise, S. Shape Selectivity in Reversed-Phase Liquid Chromatography for the Separation of Planar and Non-Planar Solutes. *J. Chromatogr. A 656* **1993**, 335–351.

[24] Vailaya, A. and Horváth, C. Retention in Reversed-Phase Chromatography: Partition or Adsorption? *J. Chromatogr. A 829* **1998**, 1–27.

[25] Kaczmarski, K., Prus, W., and Kowalska, T. Adsorption/Partition Model of Liquid Chromatography for Chemically Bonded Stationary Phases of the Aliphatic Cyano, Reversed-Phase C8 and Reversed-Phase C18 Types. *J. Chromatogr. A 869* **2000**, 57–64.

[26] Nikitas, P., Pappa-Louisi, A., and Agrafiotou, P. New Insights on the Retention Mechanism of Non-Polar Solutes in Reversed-Phase Liquid Chromatographic Columns. *J. Chromatogr. A 1034* **2004**, 41–54.

[27] Taft, R. W., Abraham, M. H., Doherty, R. M., and Kamlet, M. J. The Molecular Properties Governing Solubilities of Organic Nonelectrolytes in Water. *Nature 313* **1985**, 384–386.

[28] Abraham, M. H., Whiting, G. S., Doherty, R. M., and Shuely, W. J. Hydrogen Bonding: XVI. A New Solute Solvation Parameter, π_H^2, from Gas Chromatographic Data. *J. Chromatogr. A 587* **1991**, 213–228.

[29] Abraham, M. H., Duce, P. P., Grellier, P. L., Prior, D. V., Morris, J. J., and Taylor, P. J. Hydrogen-Bonding. Part 5. A Thermodynamically-Based Scale of Solute Hydrogen-Bond Acidity. *Tetrahedron Lett. 29* **1988**, 1587–1590.

[30] Abraham, M. H., Greillier, P. L., Prior, D. V., Morris, J. J., Taylor, P. J., Laurence, C., and Berthelot, M. Hydrogen-Bonding. Part 6. A Thermodynamically-Based Scale of Solute Hydrogen-Bond Basicity. *Tetrahedron Lett. 30* **1989**, 2571–2574.

[31] Abraham, M. H. Hydrogen Bonding. 31. Construction of a Scale of Solute Effective or Summation Hydrogen-Bond Basicity. *J. Phys. Org. Chem. 6* **1993**, 660–684.

[32] Hemström, P. and Irgum, K. Hydrophilic Interaction Chromatography. *J. Sep. Sci 29* **2006**, 1784–1821.

[33] Smith, R. M. Functional Group Contributions to the Retention of Analytes in Reversed-Phase High-Performance Liquid Chromatography. *J. Chromatogr. A 656* **1993**, 381–415.

[34] Horváth, C., Melander, W., and Molnár, I. Solvophobic Interactions in Liquid Chromatography with Non-Polar Stationary Phases. *J. Chromatogr. A 125* **1976**, 129–156.

[35] Horvath, C. and Lipsky, S. Column Design in High Pressure Liquid Chromatography. *J. Chromatogr. Sci. 7* **1969**, 109–116.

[36] Alpert, A. J., Shukla, M., Shukla, A. K., Zieske, L. R., Yuen, S. W., Ferguson, M. A., Mehlert, A., Pauly, M., and Orlando, R. Hydrophilic-Interaction Chromatography of Complex Carbohydrates. *J. Chromatogr. A 676* **1994**, 191–202.

[37] Henderson, G. M. and Rule, H. G. A New Method of Resolving a Racemic Compound. *J. Chem. Soc.* **1939**, 1568–1573.

[38] Cuatrecasas, P., Wilchek, M., and Anfinsen, C. B. Selective Enzyme Purification by Affinity Chromatography. *Proc. Natl. Acad. Sci. USA 61* **1968**, 636–643.

[39] Taylor, T. I. and Urey, H. C. Fractionation of the Lithium and Potassium Isotopes by Chemical Exchange with Zeolites. *J. Chem. Phys. 6* **1938**, 429–438.

[40] Lathe, G. and Ruthven, C. The Separation of Substances and Estimation of Their Relative Molecular Sizes by the Use of Columns of Starch in Water. *Biochem. J. 62* **1956**, 665.

[41] Svec, F. and Huber, C. G. Monolithic Materials: Promises, Challenges, Achievements. *Anal. Chem. 78* **2006**, 2100–2107.

[42] Mazzeo, J. R., D.Neue, U., Kele, M., and Plumb, R. S. Advancing LC Performance with Smaller Particles and Higher Pressure. *Anal. Chem. 77* **2005**, 460–467.

[43] DeStefano, J., Langlois, T., and Kirkland, J. Characteristics of Superficially-Porous Silica Particles for Fast HPLC: Some Performance Comparisons with Sub-2-μm Particles. *J. Chromatogr. Sci. 46* **2008**, 254–260.

[44] Nguyen, D. T.-T., Guillarme, D., Heinisch, S., Barrioulet, M.-P., Rocca, J.-L., Rudaz, S., and Veuthey, J.-L. High Throughput Liquid Chromatography with Sub-2μm Particles at High Pressure and High Temperature. *J. Chromatogr. A 1167* **2007**, 76–68.

[45] Nawrocki, J., Rigney, M., McCormick, A., and Carr, P. Chemistry of Zirconia and its Use in Chromatography. *J. Chromatogr. A 657* **1993**, 229–282.

[46] Li, J. and Carr, P. W. Effect of Temperature on the Thermodynamic Properties, Kinetic Performance, and Stability of Polybutadiene-Coated Zirconia. *Anal. Chem. 69* **1997**, 837–843.

[47] Stoll, D. R., Wang, X., and Carr, P. W. Comparison of the Practical Resolving Power of One- and Two-Dimensional High-Performance Liquid Chromatography Analysis of Metabolomic Samples. *Anal. Chem. 80* **2008**, 268–278.

[48] Lindsey, R. K., Rafferty, J. L., Eggimann, B. L., and Siepmann, J. I. Molecular Simulation Studies of Reversed-Phase Liquid Chromatography. *J. Chromatogr. A 1287* **2013**, 60–82.

[49] Klatte, S. J. and Beck, T. L. Molecular Dynamics of Tethered Alkanes: Temperature-Dependent Behavior in a High-Density Chromatographic System. *J. Phys. Chem. 97* **1993**, 5727–5734.

[50] Yarovsky, I., Aguilar, M.-I., and Hearn, M. T. High-Performance Liquid Chromatography of Amino Acids, Peptides and Proteins: CXXV. Molecular Dynamics Simulation of N-Butyl Chains Chemically Bonded to Silica-Based Reversed-Phase High-Performance Liquid Chromatography Sorbents. *J. Chromatogr. A 660* **1994**, 75–84.

[51] Klatte, S. J. and Beck, T. L. Molecular Dynamics Simulations of Tethered Alkane Chromatographic Stationary Phases. *J. Phys. Chem. 99* **1995**, 16024–16032.

[52] Yarovsky, I., Aguilar, M.-I., and Hearn, M. T. Influence of the Chain Length and Surface Density on the Conformation and Mobility of *n*-Alkyl Ligands Chemically Immobilized onto a Silica Surface. *Anal. Chem. 67* **1995**, 2145–2153.

[53] Ban, K., Saito, Y., and Jinno, K. Characterization of the Microscopic Surface Structure of the Octadecylsilica Stationary Phase Using a Molecular Dynamics Simulation. *Anal. Sci. 20* **2004**, 1403–1408.

[54] Braun, J., Fouqueau, A., Bemish, R. J., and Meuwly, M. Solvent Structures of Mixed Water/Acetonitrile Mixtures at Chromatographic Interfaces from Computer Simulations. *Phys. Chem. Chem. Phys. 10* **2008**, 4765–4777.

[55] Gupta, P. K. and Meuwly, M. Dynamics of Water/Methanol Mixtures at Functionalized Chromatographic Interfaces. *J. Phys. Chem. B 116* **2012**, 10951–10959.

[56] Schure, M., Pesek, J., and Leigh, I. Chemically Modified Surfaces. Royal Society of Chemistry **1994**.

[57] Rafferty, J. L., Siepmann, J. I., and Schure, M. R. Mobile Phase Effects in Reversed-Phase Liquid Chromatography: A Comparison of Acetonitrile/Water and Methanol/Water Solvents As Studied by Molecular Simulation. *J. Chromatogr. A 1218* **2011**, 2203–2213.

[58] Zhang, L., Rafferty, J. L., Siepmann, J. I., Chen, B., and Schure, M. R. Chain Conformation and Solvent Partitioning in Reversed-Phase Liquid Chromatography: Monte

Carlo Simulations for Various Water/Methanol Concentrations. *J. Chromatogr. A 1126* **2006**, 219–231.

[59] Sadeghi, R. R. and Cheng, H.-P. The Dynamics of Proton Transfer in a Water Chain. *J. Chem. Phys. 111* **1999**, 2086–2094.

[60] McDonald, P. D. Improving Our Understanding of Reversed-Phase Separations for the 21st Century. *Adv. Chromatogr. 42* **2003**, 323–376.

[61] Melnikov, S. M., Höltzel, A., Seidel-Morgenstern, A., and Tallarek, U. Influence of Residual Silanol Groups on Solvent and Ion Distribution at a Chemically Modified Silica Surface. *J. Phys. Chem. C 113* **2009**, 9230–9238.

[62] Li, Z., Rutan, S. C., and Dong, S. Wetting of Octadecylsilylated Silica in Methanol-Water Eluents. *Anal. Chem. 68* **1996**, 124–129.

[63] Walter, T. H., Iraneta, P., and Capparella, M. Mechanism of Retention Loss When C8 and C18 HPLC Columns are Used with Highly Aqueous Mobile Phases. *J. Chromatogr. A 1075* **2005**, 177–183.

[64] Rafferty, J. L., Siepmann, J. I., and Schure, M. R. A Molecular Simulation Study of the Effects of Stationary Phase and Solute Chain Length in Reversed-Phase Liquid Chromatography. *J. Chromatogr. A 1223* **2012**, 24–34.

[65] Rafferty, J. L., Siepmann, J. I., and Schure, M. R. Influence of Bonded-Phase Coverage in Reversed-Phase Liquid Chromatography via Molecular Simulation: I. Effects on Chain Conformation and Interfacial Properties. *J. Chromatogr. A 1204* **2008**, 11–19.

[66] Rafferty, J. L., Siepmann, J. I., and Schure, M. R. Retention Mechanism for Polycyclic Aromatic Hydrocarbons in Reversed-Phase Liquid Chromatography with Monomeric Stationary Phases. *J. Chromatogr. A 1218* **2011**, 9183–9193.

[67] Mansfield, E. R., Mansfield, D. S., Patterson, J. E., and Knotts IV, T. A. Effects of Chain Grafting Positions and Surface Coverage on Conformations of Model Reversed-Phase Liquid Chromatography Stationary Phases. *J. Phys. Chem. C 116* **2012**, 8456–8464.

[68] Lippa, K. A., Sander, L. C., and Mountain, R. D. Molecular Dynamics Simulations of Alkylsilane Stationary-Phase Order and Disorder. 1. Effects of Surface Coverage and Bonding Chemistry. *Anal. Chem. 77* **2005**, 7852–7861.

[69] Melnikov, S. M., Höltzel, A., Seidel-Morgenstern, A., and Tallarek, U. Composition, Structure, and Mobility of Water- Acetonitrile Mixtures in a Silica Nanopore Studied by Molecular Dynamics Simulations. *Anal. Chem. 83* **2011**, 2569–2575.

[70] Melnikov, S. M., Höltzel, A., Seidel-Morgenstern, A., and Tallarek, U. A Molecular Dynamics Study on the Partitioning Mechanism in Hydrophilic Interaction Chromatography. *Angew. Chem. Int. Ed. 51* **2012**, 6251–6254.

[71] Rafferty, J. L., Siepmann, J., and Schure, M. the Effects of Chain Length, Embedded Polar Groups, Pressure, and Pore Shape on Structure and Retention in Reversed-Phase Liquid Chromatography: Molecular-Level Insights from Monte Carlo Simulations. *J. Chromatogr. A 1216* **2009**, 2320–2331.

[72] Ban, K., Saito, Y., and Jinno, K. Molecular Dynamics Simulation for the Characterization of Liquid Chromatographic Stationary Phase: Effect of Temperature. *Anal. Sci. 21* **2005**, 397–402.

[73] Lippa, K. A., Sander, L. C., and Mountain, R. D. Molecular Dynamics Simulations of Alkylsilane Stationary-Phase Order and Disorder. 2. Effects of Temperature and Chain Length. *Anal. Chem. 77* **2005**, 7862–7871.

[74] Rafferty, J. L., Siepmann, J. I., and Schure, M. R. Molecular-Level Comparison of Alkylsilane and Polar-Embedded Reversed-Phase Liquid Chromatography Systems. *Anal. Chem. 80* **2008**, 6214–6221.

[75] Klatte, S. J. and Beck, T. L. Microscopic Simulation of Solute Transfer in Reversed Phase Liquid Chromatography. *J. Phys. Chem. 100* **1996**, 5931–5934.

[76] Slusher, J. T. and Mountain, R. D. A Molecular Dynamics Study of a Reversed-Phase Liquid Chromatography Model. *J. Phys. Chem. B 103* **1999**, 1354–1362.

[77] Rafferty, J. L., Siepmann, J. I., and Schure, M. R. A Molecular Simulation Study of the Effects of Stationary Phase and Solute Chain Length in Reversed-Phase Liquid Chromatography. *J. Chromatogr. A 1223* **2012**, 24–34.

[78] Rafferty, J. L., Siepmann, J. I., and Schure, M. R. Mobile Phase Effects in Reversed-Phase Liquid Chromatography: A Comparison of Acetonitrile/Water and Methanol/Water Solvents As Studied by Molecular Simulation. *J. Chromatogr. A 1218* **2011**, 2203–2213.

[79] Alonso, C., Waring, A., and Zasadzinski, J. A. Keeping Lung Surfactant Where It Belongs: Protein Regulation of Two-Dimensional Viscosity. *Biophysical Journal 89* **2005**, 266–273.

[80] Mitchell, D. J., Tiddy, G. J. T., Waring, L., Bostock, T., and Mcdonald, M. P. Phase Behaviour of Polyoxyethylene Surfactants with Water. Mesophase Structures and Partial Miscibility (Cloud Points). *J. Chem. Soc., Faraday Trans. 1 79* **1983**, 975–1000.

[81] Siepmann, J. I., Karaborni, S., and Klein, M. L. Monte Carlo Simulation of the Liquid-Vapor Coexistence in a Langmuir Monolayer of Pentadecanoic Acid. *J. Phys. Chem. 98* **1994**, 6675–6678.

[82] Harder, E., MacKerell, A. D., and Roux, B. Many-Body Polarization Effects and the Membrane Dipole Potential. *J. Am. Chem. Soc. 131* **2009**, 2760–2761.

[83] Ali, A. A. and Mulley, B. A. Formation of Liquid Crystal and Other Non-Fluid Phases in Emulsions Containing Non-Ionic Surfactants. *J. Pharm. Pharmacol. 30* **1978**, 205–213.

[84] Friberg, S. E. Vapor Pressure of Some Fragrance Ingredients in Emulsion and Microemulsion Formulations. *Int. J. Cosmetic Sci. 7* **1997**, 75–86.

[85] Yang, J. Viscoelastic Wormlike Micelles and Their Applications. *Curr. Opin. Colloid Interface Sci. 7* **2002**, 276–281.

[86] Sottmann, T., Strey, R., and Chen, S.-H. A Small-Angle Neutron Scattering Study of Nonionic Surfactant Molecules at the Water-oil Interface: Area per Molecule Microemulsion Domain Size, and Rigidity. *J. Chem. Phys. 106* **1997**, 6483–6490.

[87] Nibu, Y. and Inoue, T. Phase Behavior of Aqueous Mixtures of Some Polyethylene Glycol Decyl Ethers Revealed by DSC and FT-IR Measurements. *J. Colloid Interface Sci. 205* **1998**, 305–315.

[88] Nagarajan, R. and Ruckenstein, E. Theory of Surfactant Self-Assembly: A Predictive Molecular Themodynamic Approach. *Langmuir 7* **1991**, 2934–2969.

[89] Tyrode, E., Johnson, C. M., Kumpulainen, A., Rutland, M. W., and Claesson, P. M. Hydration State of Nonionic Surfactant Monolayers at the Liquid/Vapor Interface: Structure Determination by Vibrational Sum Spectroscopy. *J. Am. Chem. Soc. 127* **2005**, 16848–16859.

[90] Nilsson, P. G. and Lindman, B. Water Self-Diffusion in Nonionic Surfactant Solutions. Hydration and Obstruction Effects. *J. Chem.Phys. 87* **1983**, 4756–4791.

[91] Klose, G., Eisenblätter, S., Galle, J., Islamov, A., and Dietrich, U. Hydration and Structural Properties of a Homologous Series of Nonionic Alkyl Oligo(ethylene Oxide) Surfactants. *Langmuir 11* **1995**, 2889–2892.

[92] Lawrence, M. J. Surfactant Systems: Their Use in Drug Delivery. *Chem. Soc. Rev. 23* **1994**, 417–424.

[93] Stephenson, B. C., Rangel-Yagui, C. O., Junior, A. P., Tavares, L. C., Beers, K., and Blankschtein, D. Experimental and Theoretical Investigation of the Micellar-Assisted Solubilization of Ibuprofen in Aqueous Media. *Langmuir 22* **2006**, 1514–1525.

[94] Nagarajan, R. Solubilization by Amphiphilar Aggregates. *Curr. Opin. Colloid Interface Sci. 2* **1997**, 282–293.

[95] Narang, A., Delmarre, D., and Gao, D. Stable Drug Encapsulation in Micelles and Microemulsions. *Int. J. Pharm. 345* **2007**, 9–25.

[96] Rane, S. S. and Anderson, B. D. What Determines Drug Solubility in Lipid Vehicles: Is It Predictable? *Adv. Drug. Deliver. Rev. 60* **2008**, 638–656.

[97] Cao, Y., Marra, M., and Anderson, B. D. Predictive Relationships for the Effects of Triglyceride Ester Concentration and Water Uptake on Solubility and Partitioning of Small Molecules into Lipid Vehicles. *J. Pharm. Sci. 93* **2004**, 2768–2779.

[98] Rodgers, J. M., Webb, M., and Smit, B. Alcohol Solubility in a Lipid Bilayer: Efficient Grand-Canonical Simulation of an Interfacially Active Molecule. *J. Chem. Phys. 132* **2010**, 064107–1–064107–10.

[99] Moore, G. E. Cramming More Components into Integrated Circuits. *Electronics 38* **1965**, 1514–1525.

[100] Sumida, K., Rogow, D. L., Mason, J. A., McDonald, T. M., Bloch, E. D., Herm, Z. R., Bae, T.-H., and Long, J. R. Carbon Dioxide Capture in Metal-Organic Frameworks. *Chem. Rev. 112* **2012**, 724–781.

[101] Wang, X., Miller, D. S., Bukusoglu, E., de Pablo, J. J., and Abbott, N. L. Topological Defects in Liquid Crystals As Templates for Molecular Self-Assembly. *Nature Materials 15* **2015**, 106–112.

[102] Borrell, B. Chemistry: Power Play. *Nature 451* **2008**, 240–243.

[103] Laio, A. and Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A. 99* **2002**, 12562–12566.

[104] Siepmann, J. I. and Frenkel, D. Configurational Bias Monte Carlo: A New Sampling Scheme for Flexible Chains. *Mol. Phys. 75* **1992**, 59–70.

[105] Vlugt, T. H., Martin, M., Smit, B., Siepmann, J., and Krishna, R. Improving the Efficiency of the Configurational-Bias Monte Carlo Algorithm. *Mol. Phys. 94* **1998**, 727–733.

[106] Mooij, G. C. A. M., Frenkel, D., and Smit, B. Direct Simulation of Phase Equilibria of Chain Molecules. *J. Phys. Condens. Matter 4* **1992**, L255–L259.

[107] Panagiotopoulos, A. Z. Direct Determination of Phase Coexistence Properties of Fluids by Monte Carlo Simulation in a New Ensemble. *Mol. Phys. 61* **1987**, 813–826.

[108] Panagiotopoulos, A. Z., Qirke, N., Stapleton, and M. Tidesley, D. J. Phase Equilibria by Simulation in the Gibbs Ensemble Alternative Derivation, Generalization and Application to Mixture and Membrane Equilibria. *Mol. Phys. 63* **1988**, 527–545.

[109] Shinoda, W., Devane, R., and Klein, M. L. Coarse-Grained Molecular Modeling of Non-Ionic Surfactant Self-Assembly. *Soft Matter 4* **2008**, 2454–2462.

[110] Shinoda, W., Devane, R., and Klein, M. L. Multi-Property Fitting and Parameterization of a Coarse Grained Model for Aqueous Surfactants. *Mol. Simul. 33* **2007**, 27–36.

[111] Nanoporous Materials Genome Center. http://www.chem.umn.edu/nmgc/. Accessed 2016-02-17.

[112] The Materials Project. https://www.materialsproject.org. Accessed 2016-02-17.

[113] Esselink, K., Hilbers, P., van Os, N., Smit, B., and Karaborni, S. Molecular Dynamics Simulations of Model Oil/Water/Surfactant Systems. *Colloids Surf. A 91* **1994**, 155–167.

[114] Klein, M. L. and Shinoda, W. Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems. *Science 321* **2008**, 798–800.

[115] Larson, R. G. Monte Carlo Simulation of Microstructural Transitions in Surfactant Systems. *J. Chem. Phys. 96* **1992**, 7904.

[116] Floriano, M. A., Caponetti, E., and Panagiotopoulos, A. Z. Micellization in Model Surfactant Systems. *Langmuir 15* **1999**, 3143–3151.

[117] Woodhead, J. L. and Hall, C. K. Simulation of Micelle Formation in the Presence of Solutes. *Langmuir 26* **2010**, 15135–15141.

[118] Smit, B., Hilbers, P. A. J., Esselink, K., Rupert, L. A. M., Os, N. M. V., and Schlijper, A. G. Structure of a Water/Oil Interface in the Presence of Micelles: A Computer Simulation Study. *J. Phys. Chem. 95* **1991**, 6361–6368.

[119] Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B 111* **2007**, 7812–7824.

[120] Marrink, S. J., de Vries, A. H., and Mark, A. E. Coarse Grained Model for Semi-quantitative Lipid Simulations. *J. Phys. Chem. B 108* **2004**, 750–760.

[121] DeVane, R., Jusufi, A., Shinoda, W., cheng Chiu, C., Nielsen, S. O., Moore, P. B., and Klein, M. L. Parametrization and Application of a Coarse Grained Force Field for Benzene/Fullerene Interactions with Lipids. *J. Phys. Chem. B 114* **2010**, 16364–16372.

[122] He, X., Shinoda, W., DeVane, R., Anderson, K. L., and Klein, M. L. Paramaterization of a Coarse-Grained Model for Linear Alkylbenzene Sulfonate Surfactants and Molecular Dynamics Studies of Their Self-Assembly in Aqueous Solution. *Chem. Phys. Lett. 487* **2010**, 71–76.

[123] DeVane, R., Klein, M. L., Chiu, C., Nielsen, S. O., Shinoda, W., and Moore, P. B. Coarse-Grained Potential Models for Phenyl-Based Molecules: I. Parametrization Using Experimental Data. *J. Phys. Chem. B 114* **2010**, 6386–6393.

[124] Faeder, J. and Ladanyi, B. M. Molecular Dynamics Simulations of the Interior of Aqueous Reverse Micelles. *J. Phys. Chem. B 104* **2000**, 1033–1046.

[125] McQuarrie, D. A. Statistical Thermodynamics. Harper Collins Publishers **1973**.

[126] Frenkel, D. and Smit, B. Understanding Molecular Simulation: From Algorithms to Applications, vol. 1. Academic press **2001**.

[127] Allen, M. and Tildesley, D. Computer Simulation of Liquids. Oxford Science Publications. Clarendon Press **1987**.

[128] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys. 21* **1953**, 1087–1092.

[129] Ben-Naim, A. and Marcus, Y. Solvation Thermodynamics of Nonionic Solutes. *J. Chem. Phys. 18*.

[130] Rafferty, J. L., Siepmann, J. I., and Schure, M. R. Molecular Simulations of Retention in Chromatographic Systems: Use of Biased Monte Carlo Techniques to Access Multiple Time and Length Scales. In Multiscale Molecular Methods in Applied Chemistry. Springer **2011**, 181–200.

[131] Siepmann, J. I. and Mcdonald, I. R. Monte Carlo Simulations of Mixed Monolayers. *Mol. Phys. 75* **1992**, 255–259.

[132] Martin, M. G. and Siepmann, J. I. Predicting Multicomponent Phase Equilibria and Free Energies of Transfer for Alkanes by Molecular Simulation. *J. Am. Chem. Soc. 119* **1997**, 8921–8924.

[133] Rosenbluth, M. N. and Rosenbluth, A. W. Monte Carlo Calculation of the Average Extension of Molecular Chains. *J. Chem. Phys. 23* **1955**, 356–359.

[134] Martin, M. G. and Siepmann, J. I. Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes. *J. Phys. Chem. B 103* **1999**, 4508–4517.

[135] Wick, C. D. and Siepmann, J. I. Self-Adapting Fixed-Endpoint Configurational-Bias Monte Carlo Method for the Regrowth of Interior Segments of Chain Molecules with Strong Intramolecular Interactions. *Macromolecules 33* **2000**, 4508–4517.

[136] Trappe: Transferable Potential for Phase Equilibria. http://www.chem-siepmann.oit.umn.edu/siepmann/trappe. Accessed 2015-8-22.

[137] Martin, M. G. and Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of *n*-Alkanes. *J. Phys. Chem. B 102* **1998**, 2569–2577.

[138] Chen, B. and Siepmann, J. I. Transferable Potentials for Phase Equilibria. 3. Explicit-Hydrogen Description of Normal Alkanes. *J. Phys. Chem. B 103* **1999**, 5370–5379.

[139] Wick, C. D., Martin, M. G., and Siepmann, J. I. Transferable Potentials for Phase Equilibria. 4. United-Atom Description of Linear and Branched Alkenes and Alkylbenzenes. *J. Phys. Chem. B 104* **2000**, 8008–8016.

[140] Chen, B., Potoff, J. J., and Siepmann, J. I. Monte Carlo Calculations for Alcohols and Their Mixtures with Alkanes. Transferable Potentials for Phase Equilibria. 5. United-Atom Description of Primary Secondary, and Tertiary Alcohols. *J. Phys. Chem. B 105* **2001**, 3093–3104.

[141] Stubbs, J. M., Potoff, J. J., and Siepmann, J. I. Transferable Potentials for Phase Equilibria. 6. United-Atom Description for Ethers Glycols, Ketones, and Aldehydes. *J. Phys. Chem. B 108* **2004**, 17596–17605.

[142] Wick, C. D., Stubbs, J. M., Rai, N., and Siepmann, J. I. Transferable Potentials for Phase Equilibria. 7. Primary, Secondary, and Tertiary Amines, Nitroalkanes and Nitrobenzene, Nitriles, Amides, Pyridine, and Pyrimidine. *J. Phys. Chem. B 109* **2005**, 18974–18982.

[143] Lubna, N., Kamath, G., Potoff, J. J., Rai, N., and Siepmann, J. I. Transferable Potentials for Phase Equilibria. 8. United-Atom Description for Thiols, Sulfides, Disulfides, and Thiophene. *J. Phys. Chem. B 109* **2005**, 24100–24107.

[144] Rai, N. and Siepmann, J. I. Transferable Potentials for Phase Equilibria. 9. Explicit Hydrogen Description of Benzene and Five-Membered and Six-Membered Heterocyclic Aromatic Compounds. *J. Phys. Chem. B 111* **2007**, 10790–10799.

[145] Rai, N. and Siepmann, J. I. Transferable Potentials for Phase Equilibria. 10. Explicit-Hydrogen Description of Substituted Benzenes and Polycyclic Aromatic Compounds. *J. Phys. Chem. B 117* **2012**, 273–288.

[146] MacKerell Jr., A. D., Bashford, D., Bellott, M., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., T.Nguyen, D., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R.,

Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B 102* **1998**, 3586–3616.

[147] 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., and Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc. 117* **1995**, 5179–5197.

[148] Van Gunsteren, W. and Berendsen, H. Groningen Molecular Simulation (GROMOS) Library Manual. *Biomos, Groningen 24* **1987**, 13.

[149] Jones, J. E. On the Determination of Molecular Fields. I. From the Variation of the Viscosity of a Gas with Temperature. In Mathematical, Physical and Engineering Sciences, vol. 106. Proceedings of the Royal Society of London A **1924**, 441–462.

[150] Shah, M. S., Tsapatsis, M., and Siepmann, J. I. Development of the Transferable Potentials for Phase Equilibria Model for Hydrogen Sulfide. *J. Phys. Chem. B 119* **2015**, 7041–7052.

[151] Mie, G. Zur Kinetischen Theorie Der Einatomigen Körper. *Annalen der Physik 316* **1903**, 657–697.

[152] The Royal Swedish Academy of Sciences: The Nobel Prize in Chemistry 2013 (Press Release). `http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/press.pdf` **2013**. Accessed 2016-4-22.

[153] The American Chemical Society: 251st American Chemical Society National Meeting and Exposition. `http://www.acs.org/content/acs/en/meetings/spring-2016.html` **2016**. Accessed 2016-4-22.

[154] CHARMM Force Field Files. `http://mackerell.umaryland.edu/charmm_ff.shtml` **2016**. Accessed 2016-4-22.

[155] Cerutti, D. S., Swope, W. C., Rice, J. E., and Case, D. A. FF14IPQ: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins. *J. Chem. Theory Comput. 10* **2014**, 4515–4534.

[156] Bai, P., Tsapatsis, M., and Siepmann, J. I. TraPPE-Zeo: Transferable Potentials for Phase Equilibria Force Field for All-Silica Zeolites. *J. Phys. Chem. C 117* **2013**, 24375–24387.

[157] MacDermaid, C. M., Kashyap, H. K., DeVane, R. H., Shinoda, W., Klauda, J. B., Klein, M. L., and Fiorin, G. Molecular Dynamics Simulations of Cholesterol-Rich Membranes Using a Coarse-Grained Force Field for Cyclic Alkanes. *J. Chem. Phys. 143* **2015**, 243144.

[158] de Jong, D. H., Singh, G., Bennett, W. D., Arnarez, C., Wassenaar, T. A., Schäfer, L. V., Periole, X., Tieleman, D. P., and Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput. 9* **2012**, 687–697.

[159] Guvench, O. and MacKerell Jr, A. D. Automated Conformational Energy Fitting for Force-Field Development. *J. Mol. Model. 14* **2008**, 667–679.

[160] Betz, R. M. and Walker, R. C. Paramfit: Automated Optimization of Force Field Parameters for Molecular Dynamics Simulations. *J. Comput. Chem. 36* **2015**, 79–87.

[161] Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput. 10* **2014**, 4497–4514.

[162] Wang, J. and Kollman, P. A. Automatic Parameterization of Force Field by Systematic Search and Genetic Algorithms. *J. Comput. Chem. 22* **2001**, 1219–1228.

[163] Hülsmann, M., Müller, T. J., Ködderman, T., and Reith, D. Automated Force Field Optimisation of Small Molecules Using a Gradient-Based Workflow Package. *Mol. Sim. 36* **2010**, 1182–1196.

[164] Norrby, P.-O. and Liljefors, T. Automated Molecular Mechanics Parameterization with Simultaneous Utilization of Experimental and Quantum Mechanical Data. *J. Comput. Chem. 19* **1998**, 1146–1166.

[165] Huang, L. and Roux, B. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput. 9* **2013**, 3543–3556.

[166] Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem. 17* **1996**, 490–519.

[167] Jorgensen, W. L. and Tirado-Rives, J. The OPLS [optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc. 110* **1988**, 1657–1666.

[168] Toxvaerd, S. Molecular Dynamics Calculation of the Equation of State of Alkanes. *J. Chem. Phys. 93* **1990**, 4290–4295.

[169] Siepmann, J. I., Karaborni, S., and Smit, B. Simulating the Critical Properties of Complex Fluids. *Nature 365* **1993**, 330–332.

[170] Nath, S. K., Escobedo, F. A., and de Pablo, J. J. On the Simulation of Vapor-Liquid Equilibria for Alkanes. *J. Chem. Phys. 108* **1998**, 9905–9911.

[171] Černỳ, V. Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm. *J. Optim. Theory. Appl. 45* **1985**, 41–51.

[172] Kirkpatrick, S., Gelatt Jr., C. D., and Vecchi, M. P. Optimization by Simulated Annealing. *Science 220* **1983**, 671–680.

[173] Ryckaert, J.-P. and Bellemans, A. Molecular Dynamics of Liquid *n*-Butane Near Its Boiling Point. *Chem. Phys. Lett. 30* **1975**, 123–125.

[174] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N.,

Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ã., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. Gaussian09 Revision C.01. Gaussian Inc. Wallingford CT 2009.

[175] Valiev, M., Bylaska, E. J., Govind, N., Kowalski, K., Straatsma, T. P., Van Dam, H. J., Wang, D., Nieplocha, J., Apra, E., and Windus. NWChem: A Comprehensive and Scalable Open-Source Solution for Large Scale Molecular Simulations. *Comput. Phys. Commun. 181* **2010**, 1477–1489.

[176] Siepmann Group Software. `http://chem-siepmann.oit.umn.edu/siepmann/software.html`. Accessed: 2015-8-22.

[177] Anastas, P. and Warner, J. Green Chemistry: Theory and Practice. Oxford University Press **1998**.

[178] Allmon, S. D. and Dorsey, J. G. Properties of Subcritical Water As an Eluent for Reversed-Phase Liquid Chromatography–Disruption of the Hydrogen-Bond Network at Elevated Temperature and Its Consequences. *J. Chromatogr. A 1217* **2010**, 5769–5775.

[179] Vitha, M. and Carr, P. W. The Chemical Interpretation and Practice of Linear Solvation Energy Relationships in Chromatography. *J. Chromatogr. A 1126* **2006**, 143–194.

[180] K. Valkó, L. R. S. and Glajch, J. L. Retention in Reversed-Phase Liquid Chromatography As a Function of Mobile Phase Composition. *J. Chromatogr. A 656* **1993**, 501–520.

[181] Hjerten, S., Liao, J.-L., and Zhang, R. High-Performance Liquid Chromatography on Continuous Polymer Beds. *J. Chromatogr. A 473* **1989**, 273–275.

[182] Minakuchi, H., Nakanishi, K., Soga, N., Ishizuka, N., and Tanaka, N. Octadecylsilylated Porous Silica Rods As Separation Media for Reversed-Phase Liquid Chromatography. *Anal. Chem. 68* **1996**, 3498–3501.

[183] Paek, C., McCormick, A. V., and Carr, P. W. Preparation and Evaluation of Carbon Coated Alumina As a High Surface Area Packing Material for High Performance Liquid Chromatography. *J. Chromatogr. A 1217* **2010**, 6475–6483.

[184] Kirkland, J. J. Superficially Porous Silica Microspheres for the Fast High-Performance Liquid Chromatography of Macromolecules. *J. Chromatogr. A 64* **1992**, 1239–1245.

[185] Gritti, F., Leonardis, I., Abia, J., and Guiochon, G. Physical Properties and Structure of Fine Core–Shell Particles Used As Packing Materials for Chromatography: Relationships Between Particle Characteristics and Column Performance. *J. Chromatogr. A 1217* **2010**, 3819–3843.

[186] Cunningham, G. P., Vidulich, G. A., and Kay, R. L. Several Properties of Acetonitrile-Water Acetonitrile-Methanol, and Ethylene Carbonate-Water Systems. *J. Chem. Eng. Data 12* **1967**, 336–337.

[187] Grant-Taylor, D. F. and Macdonald, D. D. Thermal Pressure and Energy-Volume Coefficients for the Acetonitrile and Water System. *Can. J. Chem. 54* **1976**, 2813–2819.

[188] Dack, M. R. J. Solvent Structure. the Use of Internal Pressure and Cohesive Energy Density to Examine Contributions to Solvent-Solvent Interactions. *Aust. J. Chem. 28* **1975**, 1643–1648.

[189] Hawthorne, S. B., Yang, Y., and Miller, D. J. Extraction of Organic Pollutants from Environmental Solids with Sub- and Supercritical Water. *Anal. Chem. 66* **1994**, 2912–2920.

[190] Marcus, Y. the Structuredness of Water at Elevated Temperatures Along the Saturation Line. *J. Mol. Liq. 79* **1999**, 151–165.

[191] Lu, J., Brown, J. S., Boughner, E. C., Liotta, C. L., and Eckert, C. A. Solvatochromic Characterization of Near-Critical Water As a Benign Reaction Medium. *Ind. Eng. Chem. Res. 41* **2002**, 2835–2841.

[192] Pawlowski, T. M. and Poole, C. F. Solvation Characteristics of Pressurized Hot Water and its Use in Chromatography. *Anal. Commun. 36* **1999**, 71–75.

[193] Allmon, S. D. and Dorsey, J. G. Retention Mechanisms in Subcritical Water Reversed-Phase Chromatography. *J. Chromatogr. A 1216* **2009**, 5106–5111.

[194] Coym, J. W. and Dorsey, J. G. Reversed-Phase Retention Thermodynamics of Pure-Water Mobile Phases at Ambient and Elevated Temperature. *J. Chromatogr. A 1035* **2004**, 23–29.

[195] Rafferty, J. L., Zhang, L., Siepmann, J. I., and Schure, M. R. Retention Mechanism in Reversed-Phase Liquid Chromatography: A Molecular Perspective. *Anal. Chem. 79* **2007**, 6551–6558.

[196] Rafferty, J. L., Sun, L., Siepmann, J. I., and Schure, M. R. Investigation of the Driving Forces for Retention in Reversed-Phase Liquid Chromatography: Monte Carlo Simulations of Solute Partitioning Between *n*-Hexadecane and Various Aqueous–Organic Mixtures. *Fluid Phase Equilib. 290* **2010**, 25–35.

[197] Ranatunga, R. P. J. and Carr, P. W. A Study of the Enthalpy and Entropy Contributions of the Stationary Phase in Reversed-Phase Liquid Chromatography. *Anal. Chem. 72* **2000**, 5679–5692.

[198] Rafferty, J. L., Siepmann, J. I., and Schure, M. R. Molecular Simulations of Retention in Chromatographic Systems: Use of Biased Monte Carlo Techniques to Access Multiple Time and Length Scales. In Multiscale Molecular Methods in Applied Chemistry **2011**, 181–200.

[199] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys. 79* **1983**, 926–935.

[200] Vlugt, T. J. H., Zhu, W., Kapteijn, F., Moulijn, J. A., Smit, B., and Krishna, R. Adsorption of Linear and Branched Alkanes in the Zeolite Silicalite-1. *J. Am. Chem. Soc. 120* **1998**, 5599–5600.

[201] Demiralp, E., Gin, T., and Goddard, W. A. Morse Stretch Potential Charge Equilibrium Force Field for Ceramics: Application to the Quartz-Stishovite Phase Transition and to Silica Glass. *Phys. Rev. Lett. 82* **1999**, 1708–1711.

[202] Mcguffin, V. L. and Evans, C. E. Influence of Pressure on Solute Retention in Liquid Chromatography. *J. Micro. Sep. 3* **1991**, 513–520.

[203] Mcguffin, V. L. and Chen, S. Molar Enthalpy and Molar Volume of Methylene and Benzene Homologues in Reversed-Phase Liquid Chromatography. *J. Chromatogr. A 762* **1997**, 35–46.

[204] Sander, L. C. and Wise, S. A. the Influence of Column Temperature on Selectivity in Reversed-Phase Liquid Chromatography for Shape-Constrained Solutes. *J. Sep. Science 24* **2001**, 910–920.

[205] Shinoda, K. Iceberg Formation and Solubility. *J. Phys. Chem. 81* **1977**, 1300–1302.

[206] Khuwijitjaru, P., Adachi, S., and Matsuno, R. Solubility of Saturated Fatty Acids in Water at Elevated Temperatures. *Biosci., Biotechnol., Biochem. 66* **2002**, 1723–1726.

[207] Claussen, W. F. The Solubility of Aromatic Hydrocarbons in Water. *J. Am. Chem. Soc. 74* **1952**, 3937–3938.

[208] Neue, U. D., Kele, M., Bunner, B., Kromidas, A., Dourdeville, T., Mazzeo, J. R., Grumbach, E. S., Serpa, S., Wheat, T. E., Hong, P., and Gilar, M. Ultra-Performance Liquid Chromatography Technology and Applications. *Adv. Chromatogr 48* **2010**, 99–143.

[209] Cabrera, K. Applications of Silica-Based Monolithic HPLC Columns. *J. Sep. Sci 27* **2004**, 843–852.

[210] Wu, A. H. and French, D. Implementation of Liquid Chromatography/Mass Spectrometry into the Clinical Laboratory. *Clin. Chim. Acta. 420* **2013**, 4–10.

[211] Fatta-Kassinos, D., Meric, S., and Nikolaou, A. Pharmaceutical Residues in Environmental Waters and Wastewater: Current State of Knowledge and Future Research. *Anal. Bioanal. Chem. 399* **2011**, 251–275.

[212] Xu, W., Wang, X., and Cai, Z. Anal. Chem. of the Persistent Organic Pollutants Identified in the Stockholm Convention: A Review. *Anal. Chim. Acta 790* **2013**, 1–13.

[213] Gilar, M., Daly, A. E., Kele, M., Neue, U. D., and Gebler, J. C. Implications of Column Peak Capacity on the Separation of Complex Peptide Mixtures in Single- and Two-Dimensional High-Performance Liquid Chromatography. *J. Chromatogr. A 1061* **2004**, 183–192.

[214] Cohen, S. A. and Schure, M. R. Multidimensional Liquid Chromatography: Theory and Applications in Industrial Chemistry and the Life Sciences. John Wiley & Sons **2008**.

[215] Guillarme, D., Ruta, J., Rudaz, S., and Veuthey, J.-L. New Trends in Fast and High-Resolution Liquid Chromatography: A Critical Comparison of Existing Approaches. *Anal. Bioanal. Chem. 397* **2010**, 1069–1082.

[216] Erni, F. and Frei, R. Two-Dimensional Column Liquid Chromatographic Technique for Resolution of Complex Mixtures. *J. Chromatogr. A 149* **1978**, 561–569.

[217] Cortes, H. J. Multidimensional Chromatography. M. Dekker **1989**.

[218] Bushey, M. M. and Jorgenson, J. W. Automated Instrumentation for Comprehensive Two-Dimensional High-Performance Liquid Chromatography of Proteins. *Anal. Chem. 62* **1990**, 161–167.

[219] Mondello, L., Lewis, A. C., and Bartle, K. D. Multidimensional Chromatography. John Wiley & Sons **2001**.

[220] Stoll, D. R., Li, X., Wang, X., Carr, P. W., Porter, S. E., and Rutan, S. C. Fast, Comprehensive Two-Dimensional Liquid Chromatography. *J. Chromatogr. A 1168* **2007**, 3–43.

[221] Groskreutz, S. R., Swenson, M. M., Secor, L. B., and Stoll, D. R. Selective Comprehensive Multi-Dimensional Separation for Resolution Enhancement in High Performance Liquid Chromatography. Part I: Principles and Instrumentation. *J. Chromatogr. A 1228* **2012**, 31–40.

[222] Cohen, S. A. and Schure, M. R. Multidimensional Liquid Chromatography: Theory and Applications in Industrial Chemistry and the Life Sciences. John Wiley & Sons **2008**.

[223] Bedani, F., Schoenmakers, P. J., and Janssen, H.-G. Theories to Support Method Development in Comprehensive Two-Dimensional Liquid Chromatography-A Review. *J. Sep. Sci 35* **2012**, 1697–1711.

[224] Stoll, D. R. Guidelines for Bioanalytical 2D Chromatography Method Development and Implementation. *Bioanalysis 2* **2010**, 105–122.

[225] Jandera, P., Hájek, T., Staňková, M., Vyňuchalová, K., and Česla, P. Optimization of Comprehensive Two-Dimensional Gradient Chromatography Coupling In-Line Hydrophilic Interaction and Reversed Phase Liquid Chromatography. *J. Chromatogr. A 1268* **2012**, 91–101.

[226] Schure, M. R. and Davis, J. M. Orthogonal Separations: Comparison of Orthogonality Metrics by Statistical Analysis. *J. Chromatogr. A 1414* **2015**, 60–76.

[227] Gilroy, J. J., Dolan, J. W., and Snyder, L. R. Column Selectivity in Reversed-Phase Liquid Chromatography: IV. Type-B Alkyl-Silica Columns. *J. Chromatogr. A 1000* **2003**, 757–778.

[228] Snyder, L., Dolan, J., and Carr, P. The Hydrophobic-Subtraction Model of Reversed-Phase Column Selectivity. *J. Chromatogr. A 1060* **2004**, 77–116.

[229] Zhang, Y. and Carr, P. W. A Visual Approach to Stationary Phase Selectivity Classification Based on the Snyder-Dolan Hydrophobic-Subtraction Model. *J. Chromatogr. A 1216* **2009**, 6685–6694.

[230] Johnson, A. R., Johnson, C. M., Stoll, D. R., and Vitha, M. F. Identifying Orthogonal and Similar Reversed Phase Liquid Chromatography Stationary Phases Using the

System Selectivity Cube and the Hydrophobic Subtraction Model. *J. Chromatogr. A 1249* **2012**, 62–82.

[231] Snyder, L. R., Dolan, J. W., and W., C. P. A New Look at the Selectivity of RPC Columns. *Anal. Chem. 79* **2007**, 3254–3262.

[232] Marchand, D., Snyder, L., and Dolan, J. Characterization and Applications of Reversed-Phase Column Selectivity Based on the Hydrophobic-Subtraction Model. *J. Chromatogr. A 1191* **2008**, 2–20.

[233] Marchand, D., Carr, P., McCalley, D. V., Neue, U., Dolan, J., and Snyder, L. Contributions to Reversed-Phase Column Selectivity. II. Cation Exchange. *J. Chromatogr. A 1218* **2011**, 7110–7129.

[234] Grinberg, N. and Grushka, E. Advances in Chromatogarphy Volume 50. CRC Press **2012**.

[235] Carr, P., Dolan, J., Dorsey, J., Snyder, L., and Kirkland, J. Contributions to Reversed-Phase Column Selectivity: III. Column Hydrogen-Bond Basicity. *J. Chromatogr. A 1395* **2015**, 57–64.

[236] Dolan, J. and Snyder, L. Selecting an Orthogonal Column During High-Performance Liquid Chromatographic Method Development for Samples that may Contain Non-Ionized Solutes. *J. Chromatogr. A 1216* **2009**, 3467–3472.

[237] Græsbøll, R., Nielsen, N. J., and Christensen, J. H. Using the Hydrophobic Subtraction Model to Choose Orthogonal Columns for Online Comprehensive Two-Dimensional Liquid Chromatography. *J. Chromatogr. A 1326* **2014**, 39–46.

[238] Johnson, A. R. and Vitha, M. F. Chromatographic Selectivity Triangles. *J. Chromatogr. A 1218* **2011**, 556–586.

[239] Slonecker, P. J., Li, X., Ridgway, T. H., and Dorsey, J. G. Informational Orthogonality of Two-Dimensional Chromatographic Separations. *Anal. Chem. 68* **1996**, 682–689.

[240] Pourhaghighi, M. R., Karzand, M., and Girault, H. H. Orthogonality of Two-Dimensional Separations Based on Conditional Entropy. *Anal. Chem. 83* **2011**, 7676–7681.

[241] Davis, J. M. and Carr, P. W. Effective Saturation: A More Informative Metric for Comparing Peak Separation in One- and Two-Dimensional Separations. *Anal. Chem. 81* **2008**, 1198–1207.

[242] Semard, G., Peulon-Agasse, V., Bruchet, A., Bouillon, J.-P., and Cardinaël, P. Convex Hull: A New Method to Determine the Separation Space Used and to Optimize Operating Conditions for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A 1217* **2010**, 5449–5454.

[243] Schure, M. R. The Dimensionality of Chromatographic Separations. *J. Chromatogr. A 1218* **2011**, 293–302.

[244] Rutan, S. C., Davis, J. M., and Carr, P. W. Fractional Coverage Metrics Based on Ecological Home Range for Calculation of the Effective Peak Capacity in Comprehensive Two-Dimensional Separations. *J. Chromatogr. A 1255* **2012**, 267–276.

[245] HPLC Columns. www.hplccolumns.org. Accessed: 2015-9-13.

[246] Schure, M. R. Quantification of Resolution for Two-Dimensional Separations. *J. Microcol. Sep. 9* **1997**, 169–176.

[247] Zhang, Y., Qibule, H., Jin, Y., Wang, J., and Ma, W. Simultaneous Determination of Vitamins A, D3 and E in Infant Formula and Adult Nutritions by Online Two-Dimensional Liquid Chromatography. *Se Pu. 33* **2015**, 291–297.

[248] Williams, H. D., Trevaskis, N. L., Charman, S. A., Shanker, R. M., Charman, W. N., Pouton, C. W., and Porter, C. J. H. Strategies to Address Low Drug Solubility in Discovery and Development. *Pharmacol. Rev. 65* **2012**, 315–499.

[249] Gosangari, S. and Dyakonov, T. Enhanced Dissolution Performance of Curcumin with the Use of Supersaturatable Formulations. *Pharm. Dev. Technol. 18* **2013**, 475–480.

[250] Li, P., Ghosh, A., Wagner, R. F., Krill, S., Joshi, Y. M., and Serajuddin, A. T. Effect of Combined Use of Nonionic Surfactant on Formation of Oil-in-Water Microemulsions. *Int. J. Pharm. 288* **2005**, 27–34.

[251] Kawakami, K., Oda, N., Miyoshi, K., Funaki, T., and Ida, Y. Solubilization Behavior of a Poorly Soluble Drug Under Combined Use of Surfactants and Cosolvents. *Eur. J. Pharm. Sci. 28* **2006**, 7–14.

[252] Le, T. D., Olsson, U., Mortensen, K., Zipfel, J., and Richtering, W. Nonionic Amphiphilic Bilayer Structures Under Shear. *Langmuir 17* **2001**, 999–1008.

[253] Ferreira, T. M., Medronho, B., Martin, R. W., and Topgaard, D. Segmental Order Parameters in a Nonionic Surfactant Lamellar Phase Studied with ^{1}H–^{13}C Solid-State NMR. *Phys. Chem. Chem. Phys. 10* **2008**, 6033–3038.

[254] Schönhoff, M., Söderman, O., Li, Z. X., and Thomas, R. K. Internal Dynamics and Order Parameters in Surfactant Aggregates: A ^{2}H NMR Study of Adsorption Layers and Bulk Phases. *Langmuir 16* **2000**, 3971–3976.

[255] Ward, T. J. I., Ku, H., Phillippi, M. A., and Marie, C. Order of Polyoxyethylene Chains in the Lamellar Phase of a Nonionic Surfactant. *Mol. Cryst. Liq. Cryst. 154* **1988**, 55–60.

[256] Goldstein, R. E. on the Theory of Lower Critical Solution Points in Hydrogen-Bonded Mixtures. *J. Chem.Phys. 80* **1983**, 5340–5341.

[257] Karaborni, S., Os, N. M. V., Esselink, K., and Hilbers, P. A. J. Molecular Dynamics Simulations of Oil Solubilization in Surfactant Solutions. *Langmuir 9* **1993**, 1175–1178.

[258] Larson, R. G. Monte Carlo Simulation of Microstructural Transitions in Surfactant Systems. *J. Chem. Phys. 96* **1992**, 7904–7918.

[259] Denham, N., Holmes, M. C., and Zvelindovsky, A. V. The Phases in a Non-Ionic Surfactant $C_{12}E_6$–Water Ternary System: A Coarse-Grained Computer Simulation. *J. Phys. Chem. B 115* **2011**, 1385–1393.

[260] Rane, S. S. and Anderson, B. D. Molecular Dynamics Simulations of Functional Group Effects on Solvation Thermodynamics of Model Solutes in Decane and Tricaprylin. *Mol. Pharmaceutics 5* **2008**, 1023–1036.

[261] Sanders, S. A., Sammalkorpi, M., and Panagiotopoulos, A. Z. Atomistic Simulations of Micellization of Sodium Hexyl, Heptyl, Octyl, and Nonyl Sulfates. *J. Phys. Chem. B. 116* **2012**, 2430–2437.

[262] Rekvig, L. and Frenkel, D. Molecular Simulations of Droplet Coalescence in Oil/Water/Surfactant Systems. *J. Chem. Phys. 127* **2007**, 134701–1–134701–11.

[263] Carruth, G. F. and Kobayashi, R. Vapor Pressure of Normal Paraffins Ethane Through *n*-Decane from Their Triple Points to About 10 Mm Mercury. *J. Chem. Eng. Data 18* **1973**, 115–126.

[264] Nasirzadeh, K., Neuder, R., and Kunz, W. Vapor Pressure Determination of the Aliphatic C5 to C8 1-Alcohols. *J. Chem. Eng. Data 51* **2006**, 7–10.

[265] Sander, R. Compilation of Henry's Law Constants. *Atmos. Chem. Phys. 15* **2015**, 4399–4981.

[266] Stubbs, J. M. and Siepmann, J. I. Elucidating the Vibrational Spectra of Hydrogen-Bonded Aggregates in Solution: Electronic Structure Calculations with Implicit Solvent and First Principles Molecular Dynamics Simulations with Explicit Solvent for 1-Hexanol in *n*-Hexane. *J. Am. Chem. Soc. 127* **2005**, 4722–4729.

[267] Bouton, F., Durand, M., Nardello-Rataj, V., Borosy, A. P., Quellet, C., and Aubry, J.-M. A QSPR Model for the Prediction of the "Fish-Tail" Temperature of C_iE_4/Water/Polar Hydrocarbon Oil Systems. *Langmuir 26* **2010**, 7962–7970.

[268] Nagarajan, R., Chaiko, M. A., and Ruckenstein, E. Locus of Solubilization of Benzene in Surfactant Micelles. *J. Phys. Chem. 88* **1984**, 2916–2922.

[269] Marqusee, J. A. and Dill, K. A. Solute Partitioning into Chain Molecule Interphases: Monolayers, Bilayer Membranes, and Micelles. *J. Chem. Phys. 85* **1986**, 434.

[270] Simon, S. and Gutknecht, J. Solubility of Carbon Dioxide in Lipid Bilayer Membranes and Organic Solvents. *BBA-Biomembranes 596* **1980**, 352–358.

[271] Jain, M. K. and Wray, L. V. Partition Coefficients of Alkanols in Lipid Bilayer/Water. *Biochem. Pharmacol. 27* **1978**, 1294–1295.

[272] Katz, Y. and Diamond, J. M. Thermodynamic Constants for Nonelectrolyte Partition Between Dimyristoyl Lecithin and Water. *J. Membr. Biol. 17* **1974**, 101–120.

[273] Simon, S., Stone, W., and Bennett, P. Can Regular Solution Theory Be Applied to Lipid Bilayer Membranes? *BBA-Biomembranes 550* **1979**, 38–47.

[274] Malcolmson, C., Satra, C., Kantaria, S., Sidhu, A., and Lawrence, M. J. Effect of Oil on the Level of Solubilization of Testosterone Propionate into Nonionic Oil-in-Water Microemulsions. *J. Pharm. Sci. 87* **1998**, 109–116.

[275] Warisnoicharoen, W., Lansley, A. B., and Lawrence, M. J. Light-Scattering Investigations on Dilute Nonionic Oil-in-Water Microemulsions. *AAPS PharmSci. Tech. 2* **2000**, 16–26.

[276] McFann, G. J. and Johnston, K. P. Phase Behavior of Nonionic Surfactant/Oil/Water Systems Containing Light Alkanes. *Langmuir 9* **1993**, 2942–2948.

[277] MacKerell, A. D. Molecular Dynamics Simulation Analysis of a Sodium Dodecyl Sulfate Micelle in Aqueous Solution: Decreased Fluidity of the Micelle Hydrocarbon Interior. *J. Phys. Chem. 99* **1995**, 1846–1855.

[278] Liang, X., Marchi, M., Guo, C., Dang, Z., and Abel, S. Atomistic Simulation of Solubilization of Polycyclic Aromatic Hydrocarbons in a Sodium Dodecyl Sulfate Micelle. *Langmuir 32* **2016**, 3645–3654.

[279] Young, L. R. D. and Dill, K. A. Solute Partitioning into Lipid Bilayer Membranes. *Biochemistry 27* **1988**, 5281–5289.

Appendix A

Acronyms

Table A.1: Acronyms used in the present work.

2DHSM	Two-dimensional hydrophobic subtraction model
2DLC	Two-dimensional liquid chromatography
C10E3	Triethylene glycol mono-n-decyl ether
C16	Hexandecane
CBMC	Configurational-bias Monte Carlo
CG	Coarse-grained
DOF	Degrees of freedom
hexOH	Hexanol
HSM	Hydrophobic subtraction model
ICBSA	Iterative configurational-bial Monte Carlo
IE	Information Entropy
LC	Liquid chromatography
MC	Monte Carlo
MD	Molecular dynamics
MM	Molecular Mechanics
nona	Nonane
ODS	Octadecylsilane

Continued on next page

Table A.1 – continued from previous page

PES	Potential energy surface
PHW	Pressurized hot water
QM	Quantum mechanics
RET	Retentive
RPLC	Reversed-phase liquid chromatography
UA	United-atom
VAP	Vapor
WAT	Water
WED	Weighted Euclidean Distance

Appendix B

Performance of Additional Parameter Sets for Ranking of Column Pair Performance

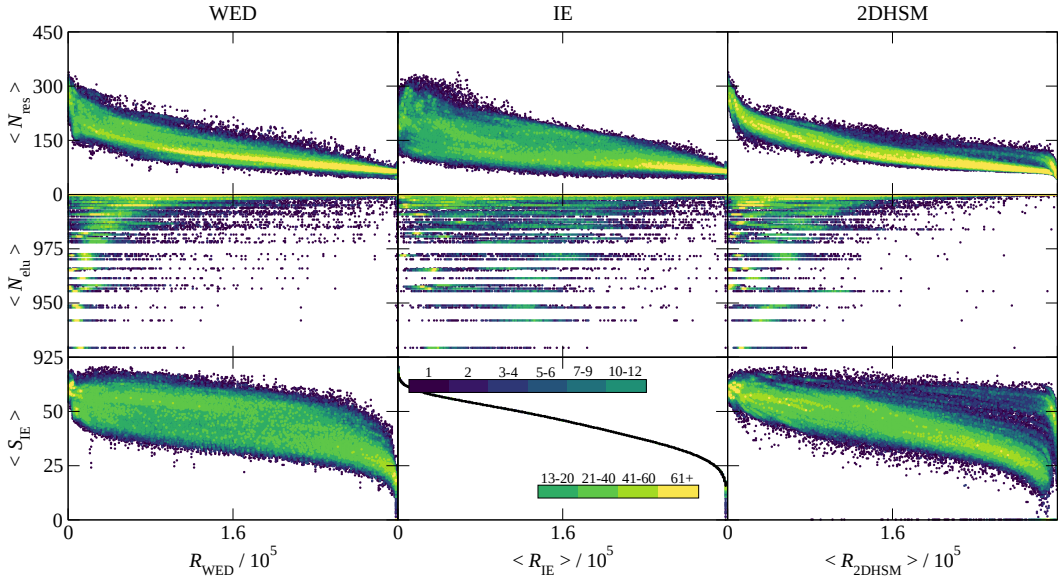

Figure B.1: Overall performance of the WED, IE, and 2DHSM methods for parameter set $k'_{\max} = 2.0$ and $n_{\text{sample}} = 100$. Rows from top to bottom provide the average number of resolved peaks, average number of eluted analytes, and average orthogonality score, respectively, as a function of the average rank (following the 1^{st} place convention) assigned to a column pair by each metric. The data for each of the 319,225 column pairs are shown with the color denoting the point density (from dark purple to yellow for low density to high density).

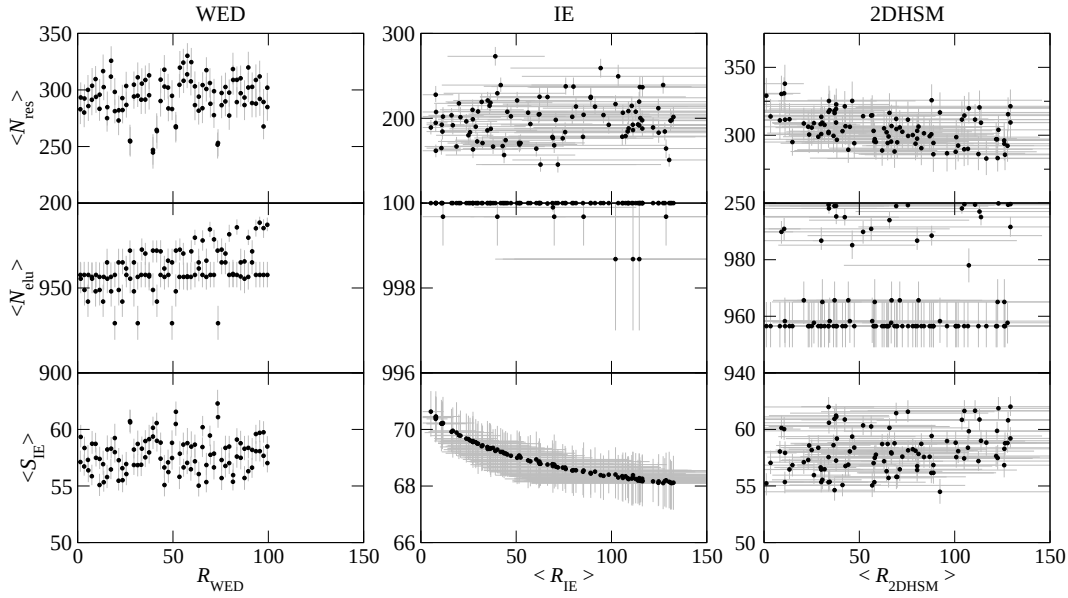

Figure B.2: Performance of top-100 columns selected by the WED, IE, and 2DHSM methods for parameter set $k'_{\max} = 2.0$ and $n_{\text{sample}} = 100$. Rows from top to bottom provide the average number of resolved peaks, average number of eluted analytes, and average orthogonality score, respectively, as a function of the average rank (following the 1^{st} place convention) assigned to a column pair by each metric. Error bars indicate the 10–90 intervals.

Table B.1: Rank and performance scores of the top column pairs obtained with the WED, IE, and 2DHSM methods, for parameter set $k'_{\max} = 2.0$ and $n_{\text{sample}} = 100$. The abbreviations EP, C18, and PBD refer to embedded polar, octadecyl, and polybutadiene phases, respectively. The super- and subscripts for each property denote the 10-90 interval.

R_{WED}	$\langle R_{\text{IE}} \rangle$	$\langle R_{\text{2DHSM}} \rangle$	Column 1	Column 2	Chem 1	Chem 2	$\langle N_{\text{res}} \rangle$	$\langle N_{\text{elu}} \rangle$	$10\langle S_{\text{IE}} \rangle$
1	21164^{28954}_{14796}	6617^{9421}_{4792}	ZirChrom-PBD	EC Nucleosil 100-5 Protect 1	PDB	EP	283^{296}_{269}	958^{965}_{950}	571^{580}_{561}
1	39281^{51432}_{29950}	1024^{1580}_{539}	EC Nucleosil 100-5 Protect 1	ZirChrom-PBD	EP	PBD	293^{307}_{280}	955^{964}_{946}	593^{603}_{583}
3	28358^{37510}_{19917}	5831^{8194}_{3581}	ZirChrom-EZ	EC Nucleosil 100-5 Protect 1	Other	EP	280^{293}_{269}	958^{965}_{950}	567^{575}_{557}
3	43217^{57451}_{31735}	815^{1337}_{354}	EC Nucleosil 100-5 Protect 1	ZirChrom-EZ	EP	Other	292^{306}_{280}	949^{958}_{939}	584^{593}_{574}
5	46737^{59705}_{35676}	314^{543}_{124}	EC Nucleosil 100-5 Protect 1	Apex II C18	EP	C18	300^{312}_{287}	942^{951}_{932}	573^{583}_{560}
5	37541^{49780}_{28071}	5323^{8407}_{2937}	Apex II C18	EC Nucleosil 100-5 Protect 1	C18	EP	286^{299}_{273}	958^{965}_{950}	564^{574}_{552}
7	52095^{63329}_{39515}	322^{511}_{130}	Zorbax Bonus RP	ZirChrom-PBD	EP	PBD	291^{305}_{277}	957^{964}_{948}	559^{567}_{548}
7	25347^{33462}_{18632}	6030^{8276}_{4230}	ZirChrom-PBD	Zorbax Bonus RP	PBD	EP	304^{319}_{288}	955^{964}_{946}	587^{596}_{576}
9	25488^{34158}_{19041}	3207^{5324}_{987}	Resolve C18	EC Nucleosil 100-5 Protect 1	C18	EP	310^{324}_{298}	948^{957}_{938}	587^{595}_{577}
9	35792^{45716}_{26128}	151^{249}_{60}	EC Nucleosil 100-5 Protect 1	Resolve C18	EP	C18	294^{308}_{281}	958^{965}_{950}	575^{583}_{566}
35291	$\mathbf{5.25}^{12}_{1}$	54988^{8319}_{8112}	Discovery CN	Vydac 218MS	EP	C18	189^{198}_{181}	1000^{1000}_{1000}	706^{714}_{697}
20415	$\mathbf{7.80}^{17}_{1}$	21256^{4121}_{3082}	ProntoSIL CN	Purospher RP-18	EP	C18	228^{238}_{219}	1000^{1000}_{1000}	704^{712}_{697}
31429	$\mathbf{7.80}^{14}_{3}$	47433^{7978}_{7885}	ACE 5 CN	Vydac 218MS	EP	C18	195^{203}_{184}	1000^{1000}_{1000}	704^{711}_{695}
44777	$\mathbf{7.87}^{21}_{1}$	139538^{14075}_{1298}	Cosmicsil Adore 100 CN	Vydac 218MS	EP	C18	162^{169}_{154}	1000^{1000}_{1000}	705^{714}_{696}
22443	$\mathbf{7.99}^{17}_{1}$	56896^{9041}_{7414}	Cosmicsil Adore 100 CN	Purospher RP-18	EP	C18	207^{217}_{197}	1000^{1000}_{1000}	704^{711}_{696}
42193	$\mathbf{10.80}^{21}_{3}$	113555^{10004}_{8936}	Fortis Cyano	Vydac 218MS	EP	C18	165^{175}_{157}	1000^{1000}_{1000}	702^{710}_{692}
34185	$\mathbf{10.95}^{23}_{3}$	49263^{7969}_{7724}	Thermo CN	Vydac 218MS	EP	C18	192^{201}_{183}	1000^{1000}_{1000}	702^{709}_{695}
37279	$\mathbf{11.33}^{24}_{3}$	64671^{9422}_{9138}	Genesis CN 300A	Vydac 218MS	EP	C18	184^{195}_{175}	1000^{1000}_{1000}	702^{710}_{693}
57931	$\mathbf{11.61}^{28}_{1}$	47464^{10442}_{9430}	Discovery HS PEG	Microsorb-MV 100 CN	EP	EP	202^{213}_{192}	1000^{1000}_{999}	702^{712}_{695}
29527	$\mathbf{16.09}^{33}_{4}$	40248^{6158}_{6931}	Discovery CN	Cogent UDC Cholesterol	EP	Other	203^{212}_{194}	1000^{1000}_{1000}	699^{707}_{692}

Table B.2: Continued from Table B.1: Rank and performance scores of the top column pairs obtained with the WED, IE, and 2DHSM methods, for parameter set $k'_{\max} = 2.0$ and $n_{\text{sample}} = 100$. The abbreviations EP, C18, and F refer to embedded polar, octadecyl, and fluoro phases, respectively. The super- and subscripts for each property denote the 10-90 interval.

R_{WED}	$\langle R_{\text{IE}} \rangle$	$\langle R_{\text{2DHSM}} \rangle$	Column 1	Column 2	Chem 1	Chem 2	$\langle N_{\text{res}} \rangle$	$\langle N_{\text{elu}} \rangle$	$10\langle S_{\text{IE}} \rangle$
129	58976^{71092}_{46338}	$\mathbf{1.24}^{2}_{1}$	Zorbax Bonus RP	Allure PFP Propyl	EP	F	329^{341}_{316}	957^{964}_{948}	552^{563}_{543}
57	40043^{50946}_{30284}	$\mathbf{3.44}^{5}_{2}$	Zorbax Bonus RP	ProntoSIL SpheriBOND 80 ODS1	EP	C18	314^{326}_{301}	957^{964}_{948}	570^{580}_{560}
219	30736^{38237}_{22592}	$\mathbf{8.32}^{17}_{3}$	Zorbax Bonus RP	Spherisorb ODS-1	EP	C18	311^{325}_{298}	957^{964}_{948}	580^{590}_{571}
219	16259^{20967}_{11520}	$\mathbf{9.13}^{19}_{3}$	Spherisorb ODS-1	Zorbax Bonus RP	C18	EP	330^{343}_{317}	990^{994}_{986}	601^{611}_{590}
205	16838^{21936}_{12289}	$\mathbf{10.56}^{23}_{2}$	Allsphere ODS1	Zorbax Bonus RP	C18	EP	331^{345}_{319}	991^{995}_{987}	600^{610}_{590}
129	57638^{69249}_{46509}	$\mathbf{10.95}^{32}_{1}$	Allure PFP Propyl	Zorbax Bonus RP	F	EP	338^{351}_{324}	958^{966}_{950}	554^{563}_{544}
205	31683^{40124}_{24626}	$\mathbf{11.04}^{23}_{5}$	Zorbax Bonus RP	Allsphere ODS1	EP	C18	311^{324}_{299}	957^{964}_{948}	579^{588}_{569}
17	45753^{57253}_{35958}	$\mathbf{13.35}^{27}_{4}$	Zorbax Bonus RP	Resolve C18	EP	C18	312^{324}_{300}	957^{964}_{948}	565^{573}_{555}
31	42164^{54679}_{30364}	$\mathbf{14.90}^{33}_{4}$	Zorbax Bonus RP	ZirChrom-PS	EP	Other	295^{305}_{285}	957^{964}_{948}	568^{578}_{558}
451	39457^{49243}_{30039}	$\mathbf{20.85}^{37}_{6}$	BetaMax Acid	ProntoSIL SpheriBOND 80 ODS1	EP	C18	309^{321}_{297}	966^{974}_{959}	571^{581}_{560}

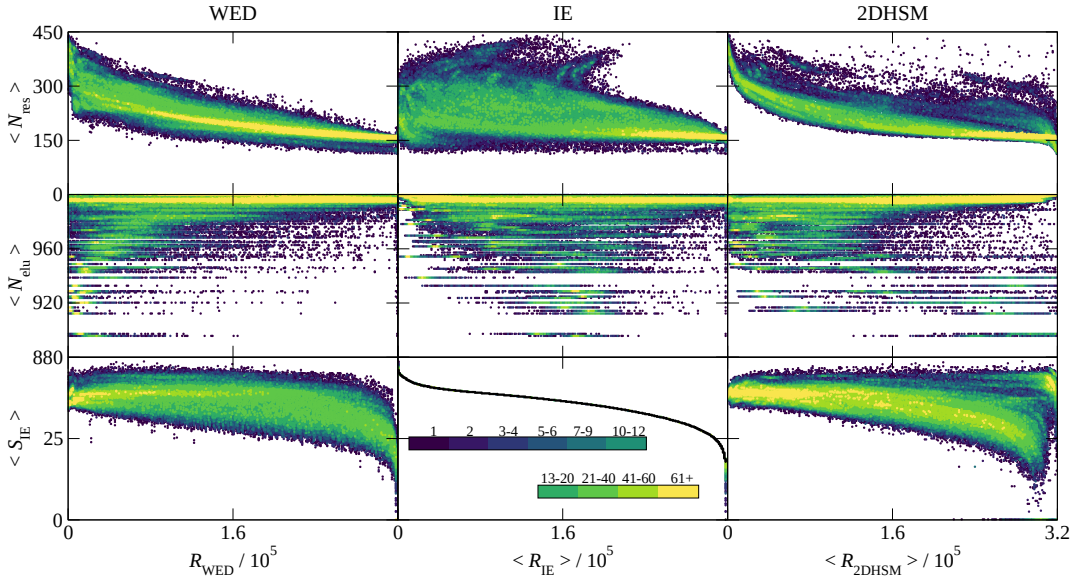

Figure B.3: Overall performance of the WED, IE, and 2DHSM methods for parameter set $k'_{\max} = 1.375$ and $n_{\text{sample}} = 200$. Rows from top to bottom provide the average number of resolved peaks, average number of eluted analytes, and average orthogonality score, respectively, as a function of the average rank (following the 1^{st} place convention) assigned to a column pair by each metric. The data for each of the 319,225 column pairs are shown with the color denoting the point density (from dark purple to yellow for low density to high density).

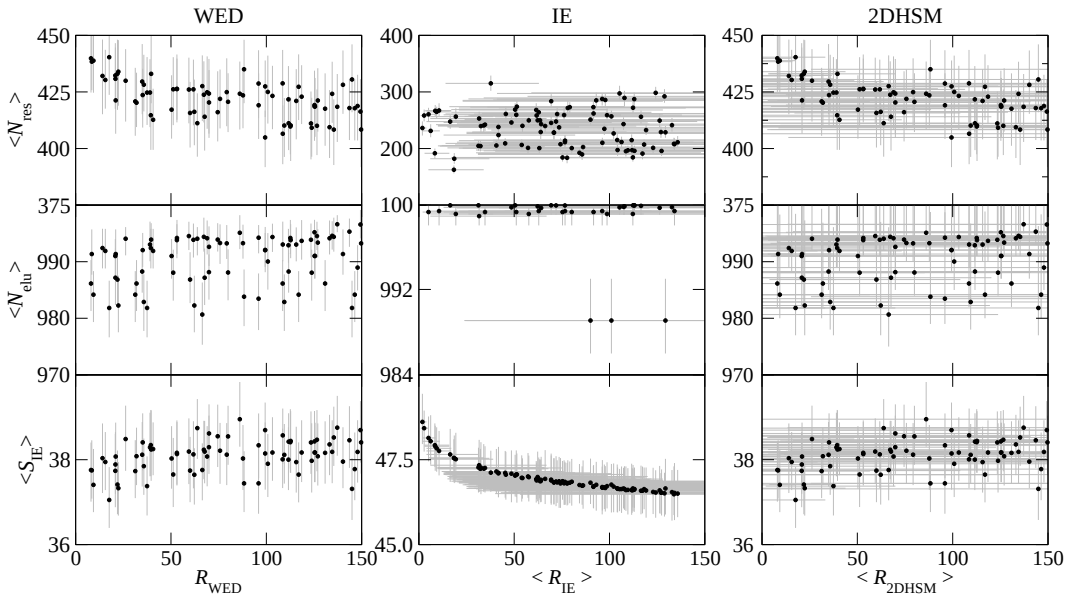

Figure B.4: Performance of top-100 columns selected by the WED, IE, and 2DHSM methods for parameter set $k'_{\max} = 1.375$ and $n_{\text{sample}} = 200$. Rows from top to bottom provide the average number of resolved peaks, average number of eluted analytes, and average orthogonality score, respectively, as a function of the average rank (following the 1^{st} place convention) assigned to a column pair by each metric. Error bars indicate the 10–90 intervals.

Table B.3: Rank and performance scores of the top column pairs obtained with the WED, IE, and 2DHSM methods, for parameter set $k'_{\max} = 1.375$ and $n_{\text{sample}} = 200$. The abbreviations EP, C18, and PBD refer to embedded polar, octadecyl, and polybutadiene phases, respectively. The super- and subscripts for each property denote the 10-90 interval.

R_{WED}	$\langle R_{\text{IE}} \rangle$	$\langle R_{\text{2DHSM}} \rangle$	Column 1	Column 2	Chem 1	Chem 2	$\langle N_{\text{res}} \rangle$	$\langle N_{\text{elu}} \rangle$	$10\langle S_{\text{IE}} \rangle$
1	21164^{28954}_{14796}	6617^{9421}_{4792}	ZirChrom-PBD	EC Nucleosil 100-5 Protect 1	PDB	EP	283^{296}_{269}	958^{965}_{950}	571^{580}_{561}
1	39281^{51432}_{29950}	1024^{1580}_{539}	EC Nucleosil 100-5 Protect 1	ZirChrom-PBD	EP	PBD	293^{307}_{280}	955^{964}_{946}	593^{603}_{583}
3	28358^{37510}_{19917}	5831^{8194}_{3581}	ZirChrom-EZ	EC Nucleosil 100-5 Protect 1	Other	EP	280^{293}_{269}	958^{965}_{950}	567^{575}_{557}
3	43217^{57451}_{31735}	815^{1337}_{354}	EC Nucleosil 100-5 Protect 1	ZirChrom-EZ	EP	Other	292^{306}_{280}	949^{958}_{939}	584^{593}_{574}
5	46737^{59705}_{35676}	314^{543}_{124}	EC Nucleosil 100-5 Protect 1	Apex II C18	EP	C18	300^{312}_{287}	942^{951}_{932}	573^{583}_{560}
5	37541^{49780}_{28071}	5323^{8407}_{2937}	Apex II C18	EC Nucleosil 100-5 Protect 1	C18	EP	286^{299}_{273}	958^{965}_{950}	564^{574}_{552}
7	52095^{63329}_{39515}	322^{511}_{130}	Zorbax Bonus RP	ZirChrom-PBD	EP	PBD	291^{305}_{277}	957^{964}_{948}	559^{567}_{548}
7	25347^{33462}_{18632}	6030^{8276}_{4230}	ZirChrom-PBD	Zorbax Bonus RP	PBD	EP	304^{319}_{288}	955^{964}_{946}	587^{596}_{576}
9	25488^{34158}_{19041}	3207^{5324}_{987}	Resolve C18	EC Nucleosil 100-5 Protect 1	C18	EP	310^{324}_{298}	948^{957}_{938}	587^{595}_{577}
9	35792^{45716}_{26128}	151^{249}_{60}	EC Nucleosil 100-5 Protect 1	Resolve C18	EP	C18	294^{308}_{281}	958^{965}_{950}	575^{583}_{566}
35291	$\mathbf{5.25}^{12}_{1}$	54988^{8319}_{8112}	Discovery CN	Vydac 218MS	EP	C18	189^{198}_{181}	1000^{1000}_{1000}	706^{714}_{697}
20415	$\mathbf{7.80}^{17}_{1}$	21256^{4121}_{3082}	ProntoSIL CN	Purospher RP-18	EP	C18	228^{238}_{219}	1000^{1000}_{1000}	704^{712}_{697}
31429	$\mathbf{7.80}^{14}_{3}$	47433^{7978}_{7885}	ACE 5 CN	Vydac 218MS	EP	C18	195^{203}_{184}	1000^{1000}_{1000}	704^{711}_{695}
44777	$\mathbf{7.87}^{21}_{1}$	139538^{14075}_{1298}	Cosmicsil Adore 100 CN	Vydac 218MS	EP	C18	162^{169}_{154}	1000^{1000}_{1000}	705^{714}_{696}
22443	$\mathbf{7.99}^{17}_{1}$	56896^{9041}_{7414}	Cosmicsil Adore 100 CN	Purospher RP-18	EP	C18	207^{217}_{197}	1000^{1000}_{1000}	704^{711}_{696}
42193	$\mathbf{10.80}^{21}_{3}$	113555^{10004}_{8936}	Fortis Cyano	Vydac 218MS	EP	C18	165^{175}_{157}	1000^{1000}_{1000}	702^{710}_{692}
34185	$\mathbf{10.95}^{23}_{3}$	49263^{7969}_{7724}	Thermo CN	Vydac 218MS	EP	C18	192^{201}_{183}	1000^{1000}_{1000}	702^{709}_{695}
37279	$\mathbf{11.33}^{24}_{3}$	64671^{9422}_{9138}	Genesis CN 300A	Vydac 218MS	EP	C18	184^{195}_{175}	1000^{1000}_{1000}	702^{710}_{693}
57931	$\mathbf{11.61}^{28}_{1}$	47464^{10442}_{9430}	Discovery HS PEG	Microsorb-MV 100 CN	EP	EP	202^{213}_{192}	1000^{1000}_{999}	702^{712}_{695}
29527	$\mathbf{16.09}^{33}_{4}$	40248^{6158}_{6931}	Discovery CN	Cogent UDC Cholesterol	EP	Other	203^{212}_{194}	1000^{1000}_{1000}	699^{707}_{692}

Table B.4: Continued from Table B.3: Rank and performance scores of the top column pairs obtained with the WED, IE, and 2DHSM methods, for parameter set $k'_{\max} = 1.375$ and $n_{\text{sample}} = 200$. The abbreviations EP, C18, and F refer to embedded polar, octadecyl, and fluoro phases, respectively. The super- and subscripts for each property denote the 10-90 interval.

R_{WED}	$\langle R_{\text{IE}} \rangle$	$\langle R_{\text{2DHSM}} \rangle$	Column 1	Column 2	Chem 1	Chem 2	$\langle N_{\text{res}} \rangle$	$\langle N_{\text{elu}} \rangle$	$10\langle S_{\text{IE}} \rangle$
129	58976^{71092}_{46338}	$\mathbf{1.24}^{2}_{1}$	Zorbax Bonus RP	Allure PFP Propyl	EP	F	329^{341}_{316}	957^{964}_{948}	552^{563}_{543}
57	40043^{50946}_{30284}	$\mathbf{3.44}^{5}_{2}$	Zorbax Bonus RP	ProntoSIL SpheriBOND 80 ODS1	EP	C18	314^{326}_{301}	957^{964}_{948}	570^{580}_{560}
219	30736^{38237}_{22592}	$\mathbf{8.32}^{17}_{3}$	Zorbax Bonus RP	Spherisorb ODS-1	EP	C18	311^{325}_{298}	957^{964}_{948}	580^{590}_{571}
219	16259^{20967}_{11520}	$\mathbf{9.13}^{19}_{3}$	Spherisorb ODS-1	Zorbax Bonus RP	C18	EP	330^{343}_{317}	990^{994}_{986}	601^{611}_{590}
205	16838^{21936}_{12289}	$\mathbf{10.56}^{23}_{2}$	Allsphere ODS1	Zorbax Bonus RP	C18	EP	331^{345}_{319}	991^{995}_{987}	600^{610}_{590}
129	57638^{69249}_{46509}	$\mathbf{10.95}^{32}_{1}$	Allure PFP Propyl	Zorbax Bonus RP	F	EP	338^{351}_{324}	958^{966}_{950}	554^{563}_{544}
205	31683^{40124}_{24626}	$\mathbf{11.04}^{23}_{5}$	Zorbax Bonus RP	Allsphere ODS1	EP	C18	311^{324}_{299}	957^{964}_{948}	579^{588}_{569}
17	45753^{57253}_{35958}	$\mathbf{13.35}^{27}_{4}$	Zorbax Bonus RP	Resolve C18	EP	C18	312^{324}_{300}	957^{964}_{948}	565^{573}_{555}
31	42164^{54679}_{30364}	$\mathbf{14.90}^{33}_{4}$	Zorbax Bonus RP	ZirChrom-PS	EP	Other	295^{305}_{285}	957^{964}_{948}	568^{578}_{558}
451	39457^{49243}_{30039}	$\mathbf{20.85}^{37}_{6}$	BetaMax Acid	ProntoSIL SpheriBOND 80 ODS1	EP	C18	309^{321}_{297}	966^{974}_{959}	571^{581}_{560}