# Developing a Simulation Framework for Modeling Biomolecules in Ionic Liquids

Vance W. Jaeger

A dissertation

submitted in partial fulfillment of the

requirements of the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Walter James Pfaendtner, Chair

Lutz Maibaum

François Baneyx

Cole DeForest

Program Authorized to Offer Degree:

Chemical Engineering

University of Washington

**Abstract**

Developing a Simulation Framework for Modeling Biomolecules in Ionic Liquids

Vance W. Jaeger

Chair of the Supervisory Committee:

Professor Walter James Pfaendtner, Ph.D.

Chemical Engineering

First, we describe the development of a protocol to quickly parameterize and validate ionic liquid force fields. Simulations that use these force fields are shown to accurately predict several thermopysical properties of ionic liquids. When compared to other solvent force fields, our force fields perform similarly well or better than many commonly used models for the prediction of solvent properties. Second, simulations containing enzymes solvated in ionic liquids and water are discussed. The mixture of ionic liquids and proteins is of interest because non-native solvents have been shown to affect the function of enzymes in both positive and negative ways. Yet, there is an incomplete understanding of the reasons why ionic liquids have such effects on proteins. The studies presented in this document focus on several different ionic liquid-enzyme systems and the ways in which the solvent affects the structure and dynamics of the solute. We also employ enhanced sampling techniques such as metadynamics to extend our understanding of the thermodynamics of ionic liquid-protein interactions. Third, we study systems that involve the self-assembly of biomolecules in ion-rich environments. One group of interesting peptides are the leucine-lysine (LK) peptides. LK peptides vary in secondary structure depending on the periodicity of their sequences, and often they have clearly separated hydrophobic and charged sides. LK peptides are known to precipitate silica bionanoparticles

from silicic acid solutions. Our simulations probe the structures of LK peptide aggregates and the formation of silicic acid structures at ordered LK peptide interfaces. In addition to LK peptides, we also study the formation of metal ion-surfactant complexes in organic solvents. The morphology of these complexes can be predicted using molecular dynamics, and our simulation results compare well with experimental data. Altogether, we demonstrate that molecular dynamics simulations of biomolecules in the presence of ionic liquids and other complex ions is a viable tool to understand solvent-solute interactions and that the results of such simulations compare well to available experimental data. We believe that the methods we have developed in this dissertation can be extended to study almost any commonly available solvent with almost any biomolecule.

**TABLE OF CONTENTS**

# Preface

In this document, I will present the rationale for combining ionic liquids (IL) and proteins and describe the research we have performed over the last several years toward understanding their interactions at the molecular level. I will demonstrate, through the contents of our publications and manuscripts, the feasibility of molecular simulations and experiments for studying IL-protein interactions and the reasons why IL-protein interactions are important.

Ionic liquids are a class of solvents that are composed of salts that have melting points below 100 °C. Within this dissertation, I will discuss room temperature ILs. Those are salts that are liquid at ambient conditions. Typically, ILs have a large organic positively charged cation. The negatively charged anions, on the other hand, can range in size from fluoride ions to complex molecules with molecular weights of hundreds of daltons. Because of the large size of at least one of the constituent ions, ILs do not pack well into crystal lattices in the same way as inorganic salts such as sodium chloride. This is the underlying reason why ILs can remain liquid at room temperature. *Within this presentation, I will occasionally refer to binary mixtures of IL and water as ILs. This is an area of some confusion within the literature, but the practice simplifies the discussion, and I will attempt to clarify this point at all necessary instances.*

Ionic liquids are considered to be some combination of organic solvent and salt. Their unique electronic structures impart upon them properties that can be important for several applications. As solvents, ILs tend to be highly viscous, nonvolatile, stable at high temperatures, and less flammable than many organic solvents. Because of these properties, ILs are considered more environmentally friendly than many common industrial solvents. They have the ability to solvate certain polymers and gases at ambient temperatures that other solvents cannot. For

example, they have been used in gas handling for the dissolution of carbon dioxide and the scrubbing of volatile organic chemicals[1, 2]. They can dissolve plastics and help in their separation[3]. They can also help in algal and cellulosic biomass preprocessing because of their rare ability to dissolve cellulose and detangle the complex matrix of hemicellulose, lignin, and cellulose that comprise the majority of biomass. More specifically for algal biomass, ILs can disrupt lipid membranes and segregate the resulting lipids into a non-aqueous phase[4]. As a charged medium, ILs can have electrochemical windows much larger than water[5]. This becomes useful for applications in batteries and other electronic devices. ILs have even been considered for heat transfer fluids[6] because they are liquid at room temperatures whereas other liquid salts used for the same applications liquefy only high temperatures.

Our research in this field began as an effort to understand the effects of ILs on the enzymatic activity of glycoside hydrolases (GHs). Glycoside hydrolases are a class of enzymes that catalyze the cleavage of bonds within cellulosic biomass. Since ILs can be used to preprocess algal and cellulosic biomass, and since enzymatic processing of this biomass produces biofuels and other important value-added chemicals, it is beneficial to understand the interactions of the solvent with functioning enzymes. If it is found that the enzymes can tolerate ILs, processes that separate dissolved cellulose from pretreatment ILs can be eliminated or severely reduced. Previously published literature has shown that certain GHs can maintain activity in low concentrations of IL, and we have been interested in describing the deleterious interactions of ILs and these enzymes at the molecular level in order to inform researchers who might want to engineer new IL-enzyme systems.

In order to facilitate molecular dynamics simulations of biomolecules in ILs, we established a protocol for creating IL force fields that can be applied to a wide variety of IL

molecules. Additionally, I automated the procedure by which we set up simulations. With these protocols, we can quickly begin simulations of nearly any biomolecule solvated in mixtures of water and organic solvent or ionic liquid. We know of no other research group that has this ability and expertise. Members of our research group have applied these protocols to many systems, and the methods have been found to be robust. We have since applied our knowledge of IL-biomolecule simulations to new scientific studies that involve the self-assembly of biomolecules in organic solutions and at silica interfaces.

## Overview

My dissertation consists of seven related studies. Of these seven studies, six have been published or are under review with publishers. The seventh study presently has a manuscript prepared for submission.

First, I will demonstrate a method by which molecular dynamics force fields for ILs can be developed quickly and easily using widely available software. Force fields contain atomic-level descriptions of molecular interactions — bonds, angles, dihedrals, van der Waals, and electrostatic potentials. By understanding the forces between the atoms in the system, we can propagate the system through time with numerical integration of Newton's equations of motion. This is the essence of molecular dynamics. In collaboration with Kayla Sprenger, I have extensively automated the process of creating such force fields by combining the capabilities several existing programs. Specifically, we must determine the electronic point charges that best represent the ions in their solvent environment. We have demonstrated that, for most ILs containing biologically common chemical elements, our force fields match the experimental observables of thermodynamic and transport properties quite well. By verifying these force fields, we are more confident in the results of the studies described after. This work has been published in *The Journal of Physical Chemistry B*.

Second, the initial project utilizing force fields parameterized using these techniques will be presented. We conducted experiments on a common family 11 xylanase from Trichoderma longibrachiatum to determine the extent to which the enzyme tolerated the presence of ILs. We determined that the enzyme tolerates 1-ethyl-3-methylimidazolium acetate and ethylsulfate ([EMIM][OAc] and [EMIM][$EtSO_4$]) in concentrations of up to 20 wt% in water. That is to say, while there is some loss of product yield, appreciable amounts of product are formed. Our

simulations sampled the structure and dynamics of the enzyme in the presence of varying concentrations of binary mixtures of IL and water. We observed affinity of the cations and anions for oppositely charged surface residues and of the cations for the binding pocket. Our experiments, our simulations, and a survey of the available literature on protein-IL interactions led us to three hypotheses that have shaped our thinking with regard to future projects.

*(A)* Ionic liquids can denature proteins by affecting the number and location of salt bridges thus exposing the protein's hydrophobic core.

*(B)* Ionic liquids can cause the aggregation of proteins in solution by screening the surface charges that normally separate them.

*(C)* Anions or cations can bind to the active site or other important sites on the protein and inhibit the regular function of the enzyme.

Ultimately, this research was published in *ACS Chemical Biology*. Shortly after the publication of our research, a separate group demonstrated inhibition of a closely related family 11 xylanase by [EMIM][7]. This corroborated one of the main hypotheses from the analysis of our simulations.

Third, in collaboration with Dr. Patrick Burney, I was inspired by an experimental study of the stability of three family 5 cellulases where each cellulase displayed a different tolerance for [EMIM][OAc] in mixtures with water. In order to address concerns about kinetically trapped states, as raised by reviewers of our previous manuscripts, we extended this set of simulations to explore longer timescales, and we performed each simulation in triplicate. The tolerance for the IL corresponded to the thermophilicity of the enzymes in both literature experiments and our simulations. We analyzed the different effects of the IL on the structure and dynamics of the enzymes. The least stable enzyme was found to have important salt bridges broken by the IL.

This led to a reshaping of the binding pocket. The second most stable enzyme was found to lose much of its secondary structure in the presence of ILs. This was likely due to the positive surface charge of the enzyme encouraging anions to penetrate the enzyme. The structure and dynamics of the most stable enzyme was found to be similar to water. We hypothesize that water-like behavior is a good predictor of the stability of enzymes in low concentrations of IL. This work has been published in *Biophysical Journal*.

Fourth, we studied an experimentally ubiquitous protein, human serum albumin (HSA). There has been important literature exploring the effects of the ILs 1-butyl-3-methylimidazolium tetrafluoroborate ([BMIM][$BF_4$]) and choline dihydrogen phosphate ([chol][dhp]) on the structure of both human and bovine serum albumin. One such study explores the unfolding of a specific region (loop 1 of domain I) of HSA in the presence of [BMIM][$BF_4$] by frequency-domain fluorescence spectroscopy. Another looks at the effects of [chol][dhp] on the distribution of bound fatty acids in the protein using double electron-electron resonance. Our simulations attempt to answer two questions: (a) Do the coupled motions of a well-defined multi-domain protein change in ILs? (b) Can we determine the free energy of unfolding for specific regions and thus discover metastable states? We answer the first question by utilizing principle component analysis to discover the slow modes of motion in the protein and to coarse grain the structure by predicting domains of motion. The second question is answered by employing well-tempered metadynamics. This method elucidates the shape of the free energy landscape across some collective variable of choice. In this case, we analyze transition from the folded to the unfolded state in water versus IL. This manuscript has been prepared for submission to *The Journal of Physical Chemistry B*.

Fifth, we wanted to answer questions related to the aggregation or salting out of proteins in the presence of ILs. We recognized that the time scales accessible by atomistic molecular dynamics were not long enough to observe aggregation events. Instead, we turned to coarse grained (CG) molecular dynamics. With coarse graining, simulations can be performed approximately forty times more quickly. As a model system, we chose short peptides consisting of leucine and lysine (LK peptides) that have been studied by other members of our group and other groups at our university. We collaborated with Prof. Joe Baio, Prof. Tobias Weidner and other experimentalists. The goal of our simulations was to add theoretical insight into the formation of biosilica nano and microparticles that are templated by LK peptides. We combined well-tempered metadynamics with the MARTINI CG force field to determine the relative free energy of different aggregates of LK peptides. The structure of the aggregates was shown to depend on the sequence and structure of the component peptides. While this study does not contain ILs, it gives us an understanding of the difficulties of simulating protein-protein interactions even with CG and metadynamics. This research has been published in *The Journal of the American Chemical Society*.

Sixth, in collaboration with Helmut Lutz, Prof. Tobias Weidner and others, we have studied the LK peptide-induced formation of silica at an air-water interface. Lutz and Weidner used sum frequency generation spectroscopy, x-ray photoelectron spectroscopy, and microscopy to analyze the ordering of the peptide and silica formations. With this information they hypothesized a mechanism of silica layer formation. We conducted molecular dynamics simulations of a peptide monolayer at an air-water interface with silica precursors solvated in the liquid layer. The results of the simulations indicated that the silica precursors interact favorably with the peptide monolayer and that the ordering of bonds at the interface is consistent with the

results obtained by our collaborators. The silica precursors and salts in the solution were parameterized using methods similar to those we developed for ILs. This indicates that our methods can be expanded to explore systems beyond proteins in a simple solution. The results of this study are in press with *Advanced Materials Interfaces*.

Seventh, we have demonstrated that molecular dynamics simulations can be helpful in determining the size, shape, and form of metal-surfactant complexes solvated in organic solvents. Dr. Pablo de la Iglesia and Prof. Lilo Pozzo found that surfactants cluster around trivalent metal cores in organic solvents to form interesting structures. These complexes impart increased conductivity and unique electrochemical properties to the organic solvents. The size and shape of the metal-surfactant complexes were observed by small angle neutron and x-ray scattering. Fitting the data required a hypothetical structure. Simulations were conducted to gain molecular insight into the preferred structures formed in the solution. These simulations once again utilized the tools that we have developed for IL-protein simulations. In general, our computational results agreed well with the hypothetical structures that Pozzo and de la Iglesia suggested. One of the simulated complexes had a different shape than was first hypothesized without the knowledge gained from molecular dynamics. Our simulations helped explain some unexpected results from the experiments. This study is one further illustration that our simulation techniques can provide testable insight into the molecular behavior of complex biomolecules in non-aqueous solvents. A manuscript containing this research is now in press in *Langmuir*.

# I. Developing Force Fields for Ionic Liquids



*Acknowledgement of coauthors:* Kayla Sprenger and I worked together closely on this project. My contribution to this work is the development of the set of scripts that automatically makes these force fields and sets up the molecular dynamics systems, and I assisted in the analysis of the data. We collaborated in the writing of the content.

## Abstract

We have applied molecular dynamics simulations to calculate thermodynamic and transport properties of a set of nineteen room temperature ionic liquids. Since accurately simulating the thermophysical properties of solvents strongly depends upon the force field of choice, we tested the accuracy of the general AMBER force field, without refinement, for the case of ionic liquids. Electrostatic point charges were developed using *ab initio* calculations and a charge scaling factor of 0.8 to more accurately predict dynamic properties. The density, heat capacity, molar enthalpy of vaporization, self-diffusivity, and shear viscosity of the ionic liquids were computed and compared to experimentally available data, and good agreement across a wide range of cation and anion types was observed. Results show that for a wide range of ionic liquids, the general AMBER force field, with no tuning of parameters, can reproduce a variety of thermodynamic and transport properties with similar accuracy to that of other published, often IL-specific, force fields.

## Introduction

Ionic liquids (ILs) are a class of solvents consisting of organic salts that are liquid below 100°C. Many ILs possess thermodynamic and electrochemical properties that make them optimal candidates for a large range of applications.[8, 9] For example, ILs tend to be thermally stable, noncombustible, and nonvolatile, making them more environmentally favorable compared to common organic solvents and thus increasingly used in green processes such as the separation of organic solvents from ILs by supercritical $CO_2$.[8] Cellulosic and algal biomass can be solvated and preprocessed by ILs for biofuel production.[4, 9] ILs have also been used to selectively absorb gases and later release them in temperature dependent processes.[2] Because of the favorable electrochemical window, ILs have been suggested as a replacement for water in certain batteries.[5] Understanding these interactions at the molecular level can help researchers optimize the species of IL used for a given device or process, and it can also guide the discovery and synthesis of new ionic liquids.

Molecular simulation, in particular Monte Carlo (MC) or molecular dynamics (MD) simulations provide a convenient route to predict thermophysical properties of ILs. The accuracy of these predictions relies on the accuracy of the force field used to describe the intra and intermolecular (bonded and non-bonded) interactions of the solvent molecules. Several force fields have been optimized to reproduce experimental data for small organic molecules, both synthetic and biological. Among these force fields are OPLS[10], CHARMM[11], UFF[12], AMBER and its companion the general AMBER force field[13] (GAFF). Since many ionic liquids are similar in structure to some of the molecules used to parameterize these popular force fields, they have commonly been used as templates for the creation of IL force fields, both in functional form and actual parameter values. Inspired by prior successes, we sought to perform a

comprehensive evaluation of a common force field (GAFF) in order to determine how well standard molecular simulation force fields perform for a new class of solvents (i.e. ILs) for which the parameters were not optimized. We note that while only MD simulations were performed in this study, the force fields generated using this method are not specific to MD, but can also be used in MC simulations.

Classical force fields consist of two sets of terms – bonded and non-bonded. Bonded parameters consist of bonds, angles, and dihedrals and are used to calculate energy potentials that include force or spring constants; non-bonded parameters include Lennard-Jones parameters and electrostatic point charges and are used to calculate 12-6 and Coulomb potentials, respectively. The sum of the bonded and non-bonded energies gives the overall potential energy of the system. For the ILs we will be testing in this paper, GAFF contains all of the needed parameters for calculation of the bonded and Lennard-Jones terms. This leaves the electrostatic point charges to be determined. Since the many interesting properties of ILs arise as a consequence of strong electrostatic interactions, which drive molecular scale fluctuations, accurately modeling the electrostatic landscape of each cation and anion is of primary importance.

The transport properties of ILs are not described well by classical force fields when full $+e/-e$ charges are assigned to the cations and anions.[14] The electronic point charges are commonly derived from quantum calculations in vacuum. It has been shown, both computationally and experimentally, that this description of a cation or anion in a vacuum with a full charge does not accurately describe the liquid system.[14-16] In a liquid, neighboring molecules can screen charge, and the electron density of a molecule can be affected by nearby molecules of the opposite charge. Because of this, previous researchers have suggested that classical force fields should assume that the charge of the cation and anion is actually some fraction of $+e/-e$.[17]

Maginn and coworkers have suggested and compared several methods to scale the ions' charges, and they have found that scaling is necessary to predict the shear viscosity or self-diffusivity of an IL using MD.[14] To overcome the shortcomings of calculating the electronic structure in vacuo, *ab initio* MD can be used with crystal structures of solidified ILs in order to scale the charges and very accurately reproduce the properties of liquid phase ILs. A more naïve approach is to simply scale the charges by multiplying by a factor of 0.8, as this is near what is typically found in the *ab initio* methods. While this approach sacrifices some accuracy, it is far more amendable to large-scale deployment of molecular simulation for predicting IL properties.

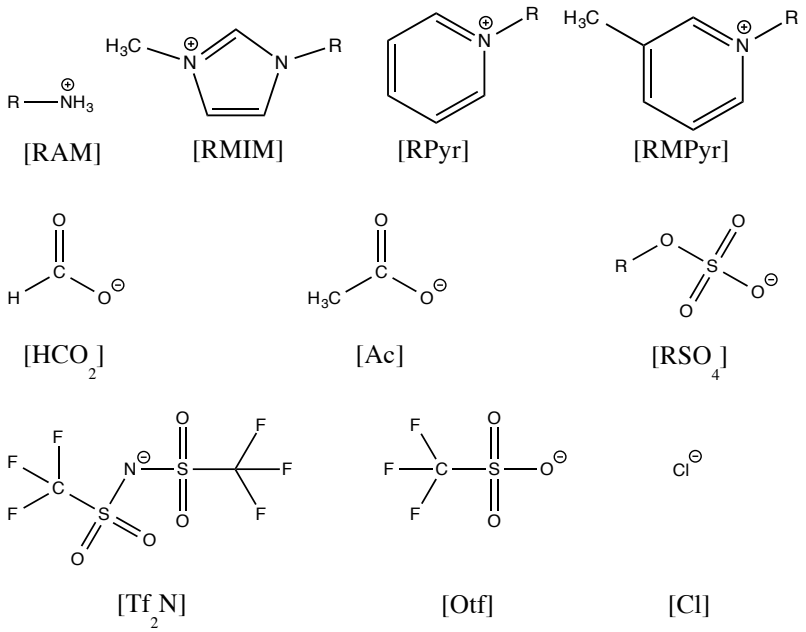



**Figure 1.1:** Ionic liquid ions. R = ethyl (Et or E), butyl (B), pentyl (P), hexyl (H), and octyl (O).

Ultimately, if IL force fields can be constructed quickly and easily without refinement, screening for the optimal ILs for certain applications could be accomplished with greater efficiency, and complex simulations involving ILs could be performed with greater ease. For example, simulating sugars, proteins, and enzymes in ILs and IL/water mixtures is becoming an

increasingly relevant area of research, supported by many recently published papers by us and others.[18-24] An even more interesting possibility is that we will be able to predict the thermophysical properties of yet undiscovered ILs or of complex interactions between solutes and ILs. These latter simulations could help to screen a large library of ILs and direct chemists in their efforts to discover useful cation-anion pairs. Other protocols have been proposed for predicting electrostatic point charges and properties of ILs.[14] These protocols require more rigorous calculations to develop the force fields. The approach we adopt herein, even though it could results in a modest loss of accuracy, can be implemented by a researcher with a basic understanding of standard MD using widely available and free programs. Specifically in this paper we propose a simple and fast protocol for accurately predicting the thermophysical properties of ILs using GAFF.

An extensive series of molecular dynamics simulations has been conducted featuring 19 combinations of ionic liquids at room conditions (298.15 K and 1 bar), including alkylammonium [RAM], 1-alkyl-3-methylimidazolium [RMIM], 1-alkylpyridinium [RPyr], and 1-alkyl-3-methylpyridinium [RMPyr] cations, and formate [$HCO_2$], acetate [Ac], alkylsulfate [$RSO_4$], bis[(trifluoromethyl)sulfonyl]imide [$Tf_2N$], trifluoromethanesulfonate [Otf], and chloride [Cl] anions. Calculations of IL density, heat capacity, molar enthalpy of vaporization, self-diffusivity, and shear viscosity have been computed for the ILs and compared to experimental data. Representations of the IL cations and anions are shown in Figure 1.1. As was noted earlier, many recent studies have focused on combining water and ILs to look at protein or enzyme structure and/or stability, thus additional simulations were performed on pure boxes of water using the TIP3P and SPC/E water models. This was done for two reasons: first, we surmised that the accuracy of TIP3P and SPC/E, given their widespread use, would be a useful

standard with which to compare the accuracy of GAFF in predicting IL properties and second, we wished to provide this comparison as a starting point for future work that could assess the ability of GAFF to predict IL/water mixture properties.

## Methods

*GAFF*. To represent the ILs, GAFF[13] was the chosen force field, the functional form of which is shown in Equation 1.1. Despite the reputation and widespread use of AMBER as a force field to study proteins and nucleic acids, AMBER lacks the parameters necessary to study many organic molecules. Thus, GAFF was designed to improve upon this aspect of AMBER and permit the detailed study of organic molecules such as those present in ILs in conjunction with studies of biomolecules. In a similar fashion to AMBER, bond stretching and angle bending terms in GAFF are evaluated with a simple harmonic functional form and the dihedral term is evaluated with a cosine function. Coulombic and 12-6 Lennard-Jones terms are used to calculate the electrostatic and van der Waals non-bonded interactions, respectively, considering all pairs of atoms (*i* and *j*) in different molecules or in the same molecule but separated by a minimum of 3 bonds. One-four interactions, or non-bonded interactions separated by 3 bonds, are reduced by a scale factor, in order to achieve identical parameters for intra- and intermolecular interactions. Parameters include the force constants $k_r$, $k_\theta$, and $V_n$, the equilibrium structural parameters $r_{eq}$ and $\theta_{eq}$, the partial atomic charges $q_i$ and $q_j$, and the Lennard-Jones well depth, $\varepsilon$, and radii, $A_{ij}$ and $B_{ij}$; $n$ is multiplicity and $\lambda$ is a phase angle for torsional angle parameters.

$$V_{tot} = \sum_{bonds} k_r(r - r_{eq})^2 + \sum_{angles} k_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\phi - \lambda\right] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \quad (1.1)$$

*Quantum Mechanics Calculations*. Individual quantum mechanics (QM) calculations were performed on all ions, i.e. [RAM] (R = E, B, and P), [RMIM] (R = E, B, H, and O), [RPyr]

(R = P and H), and [RMPyr] (R = E) cations, and [$HCO_2$], [Ac], [$RSO_4$] (R = Me and Et), [$Tf_2N$], [Otf], and [Cl] anions. Geometry optimization and energy calculations were performed at the Hartree-Fock (HF) level of theory using the 6-31G(d)//6-31G(d) basis set and the electronic structure program Gaussian.[25] All force constants and equilibrium and Lennard-Jones parameters were taken directly from the GAFF force field. Antechamber[26] was used to assign electrostatic point charges by the RESP method.[27] After the determination of partial charges, tleap[28] and ACPYPE[29] were used to generate input files for each ion. Atomic charges were scaled by multiplying each by a factor of 0.8. These methods are very similar to some of those employed by Maginn and coworkers.[30] This work aims to determine how universal this approach is for ILs.

*Ionic Liquid Simulations*. Cubic boxes of 125 $nm^3$ in size and containing 200 to 1000 ion pairs were generated using Packmol[31] at close to experimental liquid densities. IL boxes were simulated with periodic boundary conditions and used the GROMACS 4.6[32, 33] package and GAFF force field. Under 1.4 nm, electrostatics were calculated explicitly; above 1.4 nm, particle-mesh Ewald (PME) summations were used to calculate long-range electrostatic interactions. Van der Waals interactions were shifted to 0 at 1.2 nm, in order to eliminate artifacts. All bonds between hydrogen and heavy atoms were frozen at equilibrium values.

IL ions were placed randomly in the box to carry out the liquid-phase simulations. Following energy minimization, each of the 19 ILs were simulated in the NVT ensemble for 10 ns, at 500 K, to achieve thorough mixing of the ions and to eliminate any artifacts in the initial configurations, thus establishing coordinates independent of the Packmol input. Constant temperature was maintained with a global stochastic thermostat[34].

From these 10 ns NVT simulations, unique starting structures for subsequent simulations were taken from points along the simulation trajectories at 2, 4, 6, 8, and 10 ns. For each IL, the 5 unique starting structures then underwent 5 ns of pressure equilibration in the NPT ensemble, at room conditions (298.15 K and 1 bar), using the Berendsen barostat.[35] The boxes then underwent a further 10 ns of pressure equilibration in the NPT ensemble at the same conditions, using a timestep of 1 fs and the Parrinello-Rahman barostat[36] to accurately capture the correct ensemble fluctuations for calculations of the densities and heat capacities. The reported density and heat capacity values represent the values averaged across the 5 simulations.

One of the 5 simulation boxes for each IL was used to initiate triplicate non-equilibrium MD NPT simulations, 1 ns in length, wherein a cosine acceleration factor, or shear rate, of 0.03 $nm/ps^2$ was applied to each particle. This periodic perturbation method was used to calculate the shear viscosity of the ILs.[37] NPT simulations of length 10 or 20 ns depending on the IL (see Results and Discussion) were run to calculate the self-diffusivity of 4 ILs.

To calculate molar enthalpies of vaporization, gas-phase simulations were conducted in the NVT ensemble on a single cation/anion pair in a small box without periodic boundary conditions, at 298.15 K. The gas-phase simulations were run for 100 ns and the resulting trajectories divided into thirds to obtain three "trials" of gas-phase enthalpies. Three trials of liquid-phase enthalpies were calculated from the NPT production runs described earlier.

## Results and Discussion

An overview of the results for IL density, heat capacity, molar enthalpy of vaporization, and shear viscosity is shown via parity plots in Figure 1.2; results will later be discussed in detail.

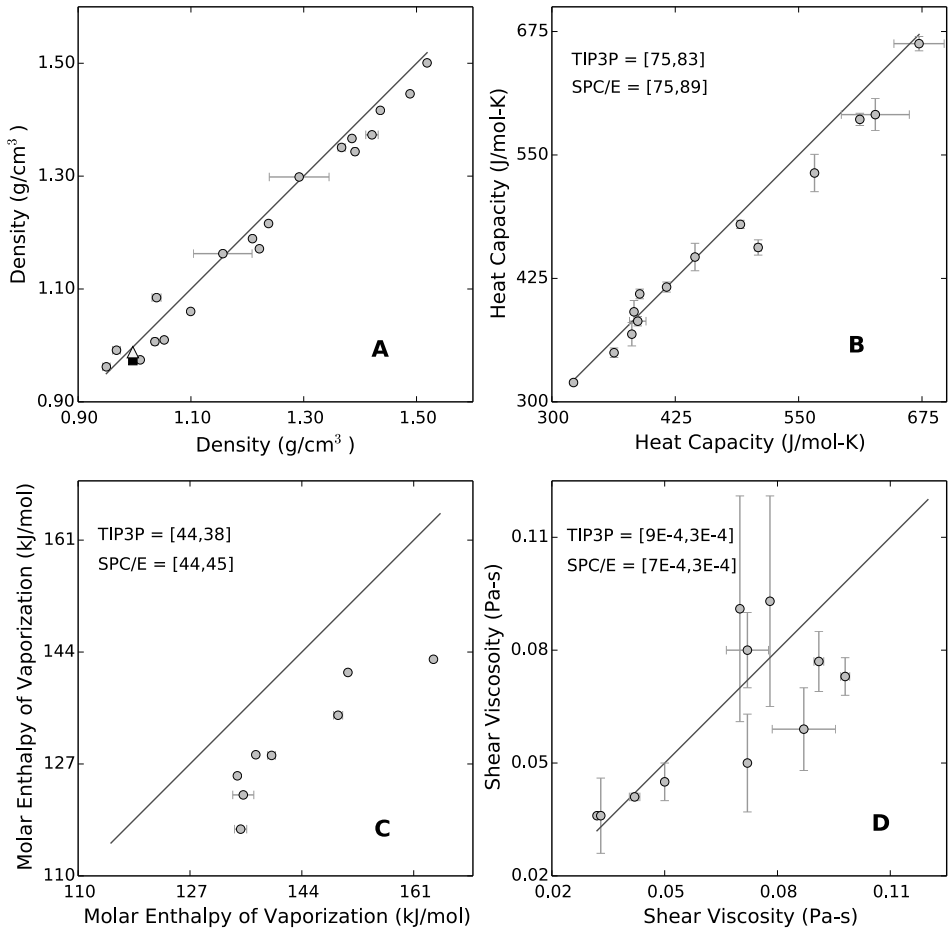



**Figure 1.2:** Simulated vs. experimental IL properties: (A) density (g/cm$^3$), square = TIP3P water, triangle = SPC/E water; (B) heat capacity (J/mol-K); (C) molar enthalpy of vaporization (kJ/mol); and (D) shear viscosity (Pa•s). The x and y axes show experimental and simulated results, respectively. Brackets show results for TIP3P and SPC/E water with experimental results listed first.

*Density*. Using the equilibrium volume of the simulation cell, we computed the density of the ILs using the standard definition. Computed and experimental densities of the 19 ILs and TIP3P and SPC/E water are reported in Table 1.1. When multiple experimental data points were available, an average of the values was taken to get the single reported experimental value in Table 1.1.

**Table 1.1:** Experimental and simulated IL densities ($g/cm^3$). Values in parentheses represent simulated densities from Vega and Abascal.[38]

| IL | Experiment | Simulation | Error (%) |
|---|---|---|---|
| [BAM][$HCO_2$] | 0.968[39] | 0.992 | 2.4 |
| [BMIM][Ac] | 1.053[40-42] | 1.010 | -4.1 |
| [BMIM][$MeSO_4$] | 1.209[43-47] | 1.189 | -1.6 |
| [BMIM][$Tf_2N$] | 1.436[48-52] | 1.416 | -1.3 |
| [EAM][$HCO_2$] | 1.039[39] | 1.085 | 4.4 |
| [EMIM][Ac] | 1.099[53-55] | 1.060 | -3.6 |
| [EMIM][$EtSO_4$] | 1.238[56-59] | 1.216 | -1.7 |
| [EMIM][Otf] | 1.386[45, 53, 60, 61] | 1.367 | -1.4 |
| [EMIM][$Tf_2N$] | 1.519[50, 58, 62] | 1.501 | -1.2 |
| [EMPyr][$EtSO_4$] | 1.221[63-65] | 1.171 | -4.1 |
| [EMPyr][$Tf_2N$] | 1.489[66] | 1.446 | -2.9 |
| [HMIM][Cl] | 1.036[67-70] | 1.007 | -2.8 |
| [HMIM][$Tf_2N$] | 1.367[54, 71-74] | 1.351 | -1.2 |
| [HPyr][$Tf_2N$] | 1.391[75, 76] | 1.343 | -3.4 |
| [OMIM][Cl] | 1.010[44, 67-69, 77, 78] | 0.974 | -3.5 |
| [OMIM][Otf] | 1.157[79, 80] | 1.163 | 0.5 |
| [OMIM][$Tf_2N$] | 1.292[72, 81, 82] | 1.298 | 0.5 |
| [PAM][$HCO_2$] | 0.950[39] | 0.962 | 1.3 |
| [PPyr][$Tf_2N$] | 1.421[83] | 1.373 | -3.4 |
| Mean Absolute Error | | | 2.4 |
| Max Absolute Error | | | 4.4 |
| **Water Model** | **Experimental** | **Simulation** | **Error (%)** |
| TIP3P | 0.997[38] | 0.973 (0.980[38]) | -2.4 |
| SPC/E | 0.997[38] | 0.988 (0.994[38]) | -0.9 |

A parity plot showing trends in the data across the full spectrum of simulated densities is shown in Figure 1.2A. Uncertainty in the experimental data, when possible to assess, is drawn in the plot with error bars representing the standard deviation of all experimental or simulation data. We note that some data points are based upon a single experimental source and thus no error bars are available. Simulated error bars are too small to be seen due to the good reproducibility of the simulations. In general, GAFF appears to slightly underestimate the densities of most of the 19

ILs; however, the same is true of TIP3P water with excellent agreement seen between SPC/E and experiment.

For the five points that lie above the parity line in Figure 1.2A, i.e. overestimate the density, three of the points correspond to the three ILs with [RAM] cations. In contrast to most other experimental densities, we note that only a single source could be found in the literature for the experimental densities of the [RAM]-based ILs. Interestingly, the simulated density error of these specific ILs appears to decrease with increasing chain length, from 4.4% (C2) to 2.4% (C4) to 1.3% (C5), and the same trend is also mildly seen for the [$Tf_2N$]-based ILs, from 1.2% (C2) to 0.5% (C8). This is not surprising given the extensive use of hydrocarbon-containing compounds in the development of GAFF. The other two points that lie just above the parity line represent two of the three [OMIM]-based ILs. Notably, these two ILs have the largest experimental standard deviations of the 19 simulated ILs (see Figure 1.2A), however they also represent the ILs with the smallest deviation in simulated density from the reported experimental values (0.5% for both). The largest deviation from experiment values occurs for [EAM][$HCO_2$] (4.4%), followed closely by [EMPyr][$EtSO_4$] and [BMIM][Ac] (4.1% for both ILs).

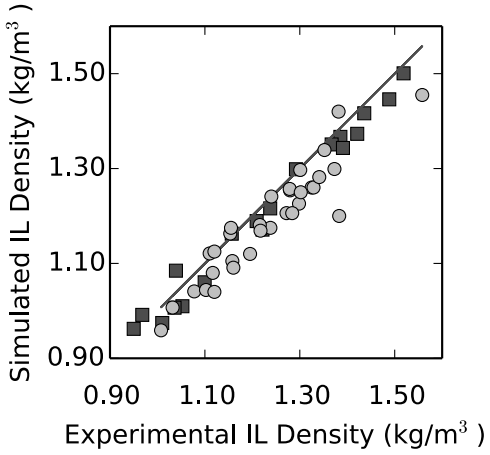



**Figure 1.3:** Simulated vs. experimental IL densities for this work (black squares) and the work of Sambasivarao and Acevedo[84] using a refined IL-specific OPLS force field (gray circles).

The trend of underestimating IL density is not unique to the GAFF force field. Simulated densities for 33 [RMIM]-based ILs, computed using a refined OPLS-AA force field[84], are compared against results of this work and to experimental results in the following parity plot. The majority of the points, as with GAFF, lie below the parity line. Contrary to GAFF, however, increasing the chain length of the cation generally leads to larger deviations from experiment; since the OPLS-AA force field was parameterized for types of molecules such as carbohydrates, rather than hydrocarbons, this makes sense.[85] For estimating densities, these results clearly show that GAFF, with no tuned parameters, provides results of comparable accuracy to an IL-specific force field.

*Heat Capacity*. Constant pressure heat capacities for the ILs were calculated using the classical statistical mechanical definition,

$$C_p = \left( \frac{\left\langle (\partial H)^2 \right\rangle}{nk_B T^2} \right)_P = \left( \frac{\left\langle H^2 \right\rangle - \left\langle H \right\rangle^2}{nk_B T^2} \right)_P \tag{1.2}$$

where $H$ and $T$ are the enthalpy and temperature of the system, respectively, $k_B$ is Boltzmann's constant, and $n$ is the total number of particles.

Simulated heat capacities for the 19 ILs and corresponding experimental values (averaged across the specified experimental sources) for 14 of the 19 ILs are reported in Table 1.2; results for TIP3P and SPC/E water are also reported in Table 1.2. These results are also graphically displayed in Figure 1.2B, which shows the simulated and experimental error of each data point (note that some points have no experimental error bar due to a single source, and that for some points, simulation errors are too small to be seen). Error bars represent the standard deviation of all experimental or simulation data.

**Table 1.2:** Experimental and simulated IL heat capacities (J/mol-K). Values in parentheses represent simulated heat capacities from Vega and Abascal[38].

| IL | Experiment | Simulation | Error (%) |
|---|---|---|---|
| [BAM][$HCO_2$] | N/A | 258 | N/A |
| [BMIM][Ac] | 383[86] | 391 | 2.1 |
| [BMIM][$MeSO_4$] | 416[45] | 416 | 0.0 |
| [BMIM][$Tf_2N$] | 566[52, 87, 88] | 532 | -6.1 |
| [EAM][$HCO_2$] | N/A | 185 | N/A |
| [EMIM][Ac] | 322[53] | 320 | -0.7 |
| [EMIM][$EtSO_4$] | 387[45, 89, 90] | 382 | -1.4 |
| [EMIM][Otf] | 363[53, 91] | 350 | -3.6 |
| [EMIM][$Tf_2N$] | 509[62, 92, 93] | 456 | -10.3 |
| [EMPyr][$EtSO_4$] | 389[94] | 409 | 5.2 |
| [EMPyr][$Tf_2N$] | 491[95] | 480 | -2.3 |
| [HMIM][Cl] | 381[96] | 369 | -3.3 |
| [HMIM][$Tf_2N$] | 628[90, 94, 97, 98] | 591 | -5.9 |
| [HPyr][$Tf_2N$] | 612[94] | 586 | -4.2 |
| [OMIM][Cl] | 445[96] | 447 | 0.4 |
| [OMIM][Otf] | N/A | 572 | N/A |
| [OMIM][$Tf_2N$] | 672[87, 94] | 663 | -1.4 |
| [PAM][$HCO_2$] | N/A | 289 | N/A |
| [PPyr][$Tf_2N$] | N/A | 561 | N/A |
| Mean Absolute Error | | | 3.6 |
| Max Absolute Error | | | 10.3 |
| **Water Model** | **Experiment** | **Simulation** | **Error (%)** |
| TIP3P | 75[38] | 83 (78[38]) | 9.9 |
| SPC/E | 75[38] | 89 (87[38]) | 18.7 |

Similar to density, GAFF slightly underestimates heat capacities of the majority of the ILs. It appears, however, that GAFF more accurately estimates heat capacities that lie within the lower half of the experimental range of heat capacities of the 14 ILs. Incidentally, it can be pointed out that the 4 most underestimated heat capacities, all in the upper range of Figure 1.2B, represent ILs with fluorinated anions, specifically [$Tf_2N$]. This underestimation appears to lessen quite drastically with increasing chain length, from 10.3% (C2) to 1.4% (C8). As was noted earlier in regards to density, this result is not unusual in light of the fact that GAFF was parameterized with a significant fraction of hydrocarbon-containing compounds and increasing

the alkyl chain length has the net result of reducing the impact of inaccuracies introduced from the fluorinated components.

In reality, the quantity that we calculated in this work represents only the excess, or residual portion of the heat capacity, which accounts for the intermolecular interactions in the condensed phase.[99] Conversely, the other portion of the heat capacity is an ideal gas contribution that takes into account intramolecular interactions.[99] These two contributions to the heat capacity are typically represented as shown in Equation 1.3.

$$\langle H \rangle = \langle H^{ig} \rangle + \langle H^{res} \rangle = \langle \Phi^{\text{int}} + K + Nk_B T \rangle + \langle \Phi^{nb} + PV - Nk_B T \rangle \tag{1.3}$$

Typically, ideal gas portions of heat capacities can be found through experiments, but as this type of experimental data is currently not available for ILs, the values are calculated from a frequency analysis of optimized cation and anion structures from *ab initio* MD simulations.[100] Since GAFF appears to give good agreement with experimental heat capacities across a broad range of ILs using the quick screening method we are proposing here, we conclude that the residual heat capacity is the dominant term and thus deemed it unnecessary to compute the ideal gas portions of the heat capacities.

*Molar Enthalpy of Vaporization*. Experiments taking place over the last decade or so have served to refute the idea of ILs as nonvolatile liquids, and on the contrary have proved that to some extent ILs can be distilled.[101] Because of this revelation, there has been a push to understand the nature of ILs in the vapor phase, and thus to be able to accurately model and even predict molar enthalpies of vaporization of ILs is of growing importance.

Previous studies have shown that ILs most likely transition into the gas phase via single, neutral ion pairs (i.e. one cation and one anion).[99] Thus, as was described in the Methods section, the IL molar enthalpies of vaporization were calculated by simulating a single

cation/anion pair in the gas phase. The equation for calculating the molar enthalpy of vaporization is shown below,

$$\begin{aligned}\Delta h_{vap}(T,P) &= \left\langle h_{gas}(T)\right\rangle - \left\langle h_{liquid}(T,P)\right\rangle \\ &= \left[\left\langle u_{gas}(T)\right\rangle + RT\right] - \left[\left\langle u_{liquid}(T,P)\right\rangle + Pv\right] \\ &\approx \left[\left\langle u_{gas}(T)\right\rangle + RT\right] - \left[\left\langle u_{liquid}(T,P)\right\rangle\right] \\ &= \Delta u_{vap}(T,P) + RT\end{aligned} \tag{1.4}$$

where Chevron brackets represent ensemble averaged molar quantities. The gas is assumed to be ideal, and $P_v$ contributions (i.e. pressure-volume work, where $P$ is the pressure and $v$ is the molar volume) in the liquid can be considered negligible when compared to the internal energies and thus can be neglected. Additionally, the liquid internal energies ($u_{liquid}(T,P)$) were divided by the total number of ion pairs in the system, since the desired final quantity was the heat of vaporization ($\Delta h_{vap}$) for a single ion pair.

Due to limited experimental data, experimental molar enthalpies of vaporization and results from this work for only 8 of the 19 ILs are reported in Table 1.3, as well as in a parity plot in Figure 1.2C; simulation error bars are too small to be seen, and multiple experimental sources could only be found for the four [$TF_2N$]-based ILs, thus some points have no experimental error bars. Error bars represent the standard deviation of all experimental or simulation data. For the [$Tf_2N$]-based ILs in Table 1.3, the reported experimental value is an average of those from the specified sources. Also reported in Table 1.3 is data for TIP3P and SPC/E water, as well as molar enthalpies of vaporization of the [RMIM][$TF_2N$] ILs from 2 other simulation groups[102, 103] for comparison to the results of this work.

**Table 1.3:** Experimental and simulated IL molar enthalpies of vaporization (kJ/mol). Values in parentheses represent simulated molar enthalpies of vaporization from Santos et al.[102], Kelkar and Maginn[103], and Vega and Abascal[38].

| IL | Experiment | Simulation | Error (%) |
|---|---|---|---|
| [EMIM][$Tf_2N$] | 135[104, 105] | 117 (159[102], 146[103]) | -13.0 |
| [BMIM][$Tf_2N$] | 135[104, 105] | 122 (174[102], 151[103]) | -9.5 |
| [HMIM][$Tf_2N$] | 139[104, 105] | 128 (184[102], 157[103]) | -8.0 |
| [OMIM][$Tf_2N$] | 150[104, 105] | 134 (201[102], 162[103]) | -10.1 |
| [PPyr][$Tf_2N$] | 134[106] | 125 | -6.7 |
| [HPyr][$Tf_2N$] | 137[106] | 128 | -6.3 |
| [EMIM][$EtSO_4$] | 164[104] | 143 | -12.9 |
| [OMIM][Otf] | 151[104] | 141 | -6.7 |
| Mean Absolute Error | | | 9.1 |
| Max Absolute Error | | | 13.0 |
| **Water Model** | **Experiment** | **Simulation** | **Error (%)** |
| TIP3P | 44[38] | 38 (42[38]) | -12.9 |
| SPC/E | 44[38] | 45 (49[38]) | 2.4 |

Our results show that GAFF is able to calculate molar enthalpies of vaporization of the 8 ILs with fair agreement to experiment, with the largest percent error occurring for [EMIM][$Tf_2N$] (-13.0%; note this does not exceed the error of the TIP3P water model). Additionally, at least in the case of the 4 [RMIM][$Tf_2N$] ILs, GAFF predicts the enthalpies of vaporization with similar or better accuracy than the computed values of other published studies[102, 103], illustrated in Figure 1.4.

For the 6 [$TF_2N$]-based ILs in Table 1.3, GAFF captures the well-documented trend, both experimentally and computationally, of increasing vaporization molar enthalpy with increasing cation carbon chain length. For the four [$Tf_2N$]-based ILs, our results, in agreement with those of Kelkar and Maginn[103], show a 5-6 kJ/mol increase in the enthalpy of vaporization for each 2-carbon increase in cation chain length. This trend also generally agrees with the experimental results, though virtually no increase in the molar enthalpy of vaporization is seen with an increase in carbon chain length from C2 to C4, experimentally, and a larger increase (10 kJ/mol)

accompanies a jump from a length of C6 to C8. Both past simulations and experiments have proposed that these increases are due to additional van der Waals interactions brought about by the increasing chain lengths.[102, 103]

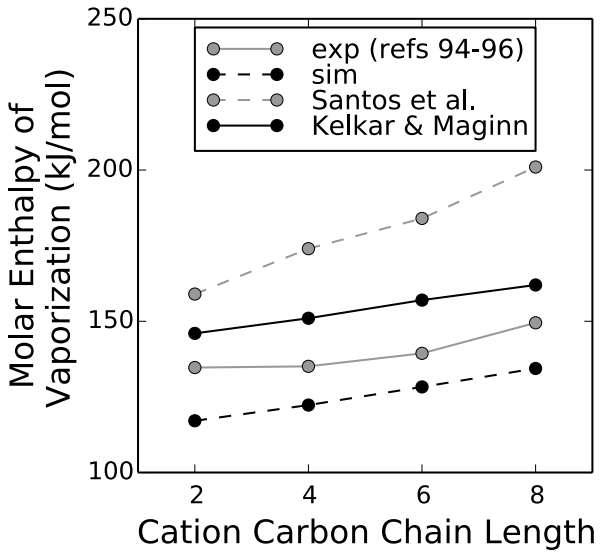



**Figure 1.4:** Simulated vs. experimental IL molar enthalpies of vaporization for [RMIM][$Tf_2N$] ILs (R=E, B, H, and O). Simulation results are also shown from the work of Santos et al.[102], and Kelkar and Maginn[103].

*Self-Diffusivity*. Though the ability to predict thermodynamic properties with accuracy is a good first assessment of a force field, it does not necessarily ensure that transport properties will also be well predicted. Thus, in assessing the accuracy of a force field, it is often very important to calculate common dynamic properties, such as self-diffusivity and shear viscosity, to complement calculations of static and equilibrium properties.

The Einstein relation shown in Equation 1.5 was used to compute the self-diffusivity of the cations and anions through calculation of the mean square displacement, or MSD, of the ion centers-of-mass,

$$D = \frac{1}{6}\lim_{t\to\infty}\frac{d}{dt}\left\langle \sum_{i=1}^{N}\left[\vec{r}(t)-\vec{r}(0)\right]^2 \right\rangle \tag{1.5}$$

where the bracketed term is the MSD over time, and the factor of 1/6 arises for a three-dimensional system.

As with the molar enthalpy of vaporization, limited experimental data led to our decision to calculate the self-diffusivity of only the 4 [RMIM][$TF_2N$] ILs. Computed and experimental self-diffusivities for the cation and anion of each of the ILs are reported in Table 1.4, along with TIP3P and SPC/E water data. All production simulations for were run in triplicate for 10 ns except where later noted. For ease of visualizing the trends, the data is also graphically displayed in Figure 1.5; error bars represent the standard deviation of all experimental or simulation data, and simulation error bars are generally too small to be seen, whereas points lack experimental error bars due to the availability of only a single experimental source.

**Table 1.4:** Experimental and simulated IL self-diffusivities ($10^{-11}$ $m^2/s$). Values in parentheses represent simulated self-diffusivities from Vega and Abascal[38].

| **IL** | **Cation Exp** | **Cation Sim** | **Error (%)** | **Anion Exp** | **Anion Sim** | **Error (%)** |
|---|---|---|---|---|---|---|
| [EMIM][$Tf_2N$] | 4.98[107] | 2.29 | -53.9 | 3.10[107] | 1.11 | -64.1 |
| [BMIM][$Tf_2N$] | 2.76[107] | 1.13 | -51.8 | 2.20[107] | 0.83 | -62.2 |
| [HMIM][$Tf_2N$] | 1.69[107] | 0.87 | -48.7 | 1.54[107] | 0.61 | -60.6 |
| [OMIM][$Tf_2N$] | 1.17[107] | 0.41 | -65.1 | 1.17[107] | 0.31 | -73.4 |
| Mean Absolute Error | | | 54.9 | | | 65.1 |
| Max Absolute Error | | | 65.1 | | | 73.4 |
| **Water Model** | **Exp** | **Sim** | **Error (%)** | | | |
| TIP3P | 230[38] | 586 (549[38]) | 154.8 | | | |
| SPC/E | 230[38] | 264 (254[38]) | 14.7 | | | |

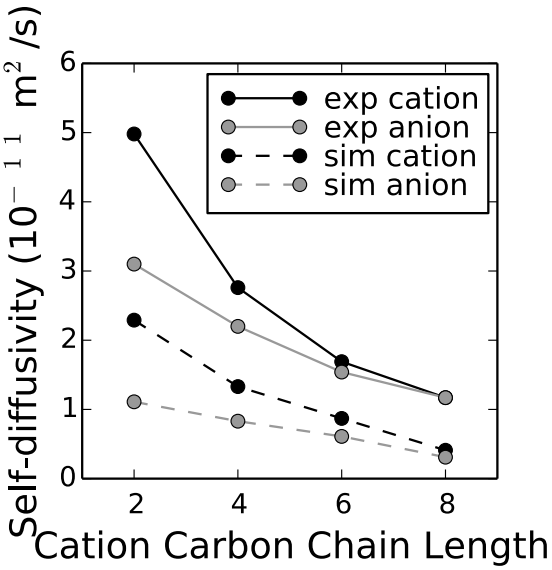



**Figure 1.5:** Simulated vs. experimental IL cation and anion self-diffusivities for [RMIM][$TF_2N$] ILs (R=E, B, H, and O).

The results show that GAFF is able to accurately capture the trend of decreasing cation/anion self-diffusivity with increasing cation chain length, and even to some extent, the relatively large nonlinear trend in self-diffusivity of the cation in moving from a cation chain length of C2 to C4. The trend of decreasing cation/anion self-diffusivity with increasing cation chain length has been observed by others, and can be explained by an increased steric hindrance brought on by longer cation chain lengths that serve to slow down the diffusion of the cations through the IL.[100] GAFF is also able to capture the interesting trend of the anion self-diffusivity tracking that of the cation, a trend that has been observed by Cadena et al.[100] In other words, our results show that the self-diffusivity of the [$TF_2N$] anion is highest when paired with the fastest cation, [EMIM], and that it is lowest when paired with the slowest cation, [OMIM]; as others have noted, this suggests that cations and anions diffuse through the liquid in pairs or clusters, rather than on their own.[100] Our results also show that for the 4 IL cation chain lengths tested, the self-diffusivity of the cation at 298.15 K is always higher than that of the anion, though

studies show this is not the case for all ILs.[100] From a quantitative perspective, deviations in cation/anion self-diffusivity from experiment range from 49-73%. There is consistently a greater deviation from experiment of the simulated anion self-diffusivities than of the simulated cation self-diffusivities, with the largest percent error occurring for [$Tf_2N$] in [OMIM][$TF_2N$].

When considering the quantitative and qualitative accuracy of the results and trends of IL self-diffusivity, respectively, it is important to consider the ability of each IL to reach diffusive behavior over the timescale of the simulations. In a review article on computing IL properties with simulation, Maginn suggests computing the time-dependent quantity beta, or $\beta(t)$, as a way to measure whether a system is in the diffusive regime.[99] The equation to calculate $\beta(t)$ is shown below, where t is the simulation time and $\Delta r^2$ is the MSD of the ion centers-of-mass, similar to Equation 1.5.

$$\beta(t) = \frac{d\log(\Delta r^2)}{d\log(t)} \tag{1.6}$$

The system can be said to have reached diffusive behavior when β(t) has reached a value of 1; values below 1 thus indicate sub-diffusive behavior.[99] The triplicate TIP3P water simulations all reached $\beta(t)$=1 after only 1 ns, while the triplicate simulations of [RMIM][$TF_2N$] (R = Ethyl, Butyl, and Hexyl alkyl chains) had all reached diffusive behavior after 5 ns. Neither simulation of [OMIM][$TF_2N$], however, reached diffusive behavior over the 20 ns; after leveling off at around 17 ns, the maximum $\beta(t)$ value reached for either simulation was 0.6. Thus, for ILs with long-chained ions, such as [OMIM][$TF_2N$], it would be advisable to extend the simulation time, or increase the simulation temperature, to have more confidence in the calculated ion self-diffusivities. Indeed, of the 4 [RMIM] cations, the greatest percent error relative to experiment is seen for [OMIM] (65%), and as was noted earlier, of the 4 [$Tf_2N$] anions, [$Tf_2N$] in [OMIM][$TF_2N$] has the greatest deviation (73%).

*Shear Viscosity*. Past simulations of IL shear viscosity have roughly fallen into two categories: 1) equilibrium MD simulations that focus on the intrinsic pressure and momentum fluctuations of the system, and 2) non-equilibrium MD simulations that monitor the linear response of the system after it is driven away from equilibrium. Nearly all previous attempts using equilibrium methods have led to overestimations of IL shear viscosities by at least an order of magnitude, while calculations using non-equilibrium methods have tended to agree better with experiment and generally require significantly shorter simulation times for calculations.[99] For example, in a study performed by Hess, the shear viscosity of a Lennard-Jones fluid using the SPC and SPC/E water models was calculated using two equilibrium and two non-equilibrium methods; of the four methods, the non-equilibrium, periodic perturbation method (PPM) provided the best results.[108] Accordingly, we adopted the PPM method, first developed by Gosling et al.[37], to assess GAFF's ability to predict IL shear viscosities.

In using PPM to calculate shear-viscosity, an external force is applied to the system, which induces a velocity field in the liquid that is described by the transverse portion of the Navier-Stokes equation, shown below,

$$\rho \frac{\partial u_x(z,t)}{\partial t} = \rho a_x(z,t) + \eta \frac{\partial^2 u_x(z,t)}{\partial z^2} \tag{1.7}$$

where $a_x(z,t)$ and $u_x(z,t)$ are the external acceleration and velocity field, respectively. An acceleration can be added to each particle at each MD step, according to the acceleration profile shown below,

$$a_x(z,t) = a_0 \cos(kz) \qquad k = \frac{2\pi}{l_z} \tag{1.8}$$

where $k$ is the wavenumber and $l_z$ is the box height in the z-dimension. The acceleration profile is assumed to take this shape because it satisfies the conditions that the acceleration and velocity

profiles be both periodic (since the simulations feature periodic systems) and smooth (so local shear rates are small).[108] This assumption, together with the assumption that the initial velocity $u_x(z,t=0)$ is zero, leads to the following solution to the velocity profile,

$$u_x(z,t) = a_0\tau(1-e^{-\frac{t}{\tau}})\cos(kz) = u_0\cos(kz) \tag{1.9}$$

where $\tau$, the system relaxation time, is given by Equation 1.10.

$$\tau = \frac{\rho}{\eta k^2} \tag{1.10}$$

At sufficiently long times, $e^{-t/\tau}$ goes to zero, and the amplitude of the steady-state velocity profile, $u_0$, is equal to $a_0\tau$. By fitting this steady-state velocity profile to a $cos(kz)$ form, one can then measure $u_0$ and use Equation 1.11 to calculate the IL shear viscosity.

$$\eta = \frac{a_0}{u_0}\frac{\rho}{k^2} \tag{1.11}$$

As was described earlier in the Methods section, 1 ns NPT simulations were run in triplicate to calculate the shear viscosity. Simulations were run at shear rates of 0.01, 0.03, 0.05, and 1.0 nm/ps$^2$, and though the shear viscosity calculated using the smallest perturbation should theoretically be nearest to the true shear viscosity at the hydrodynamic limit (i.e. $k=0$)[109], the best overall agreement with experiments was achieved with a shear rate of 0.03 nm/ps$^2$. Thus, Table 1.5 shows the computed and experimental shear viscosities for 17 of the 19 ILs (experimental data at 298.15 K was unavailable for [EMPyr][$Tf_2N$] and [OMIM][Otf]), as well as for TIP3P and SPC/E water. Reported experimental values represent averages of the values from the specified experimental sources. The data is also displayed in a parity plot in Figure 1.2D, though the range is restricted to highlight good fits (note that some points have no experimental error bar due to a single source, and that for some points, simulation errors are too

small to be seen). Error bars shown in Figure 1.2D represent the standard deviation of all experimental or simulation data.

**Table 1.5:** Experimental and simulated IL shear viscosities (Pa•s) using a shear rate of 0.03 nm/ps$^2$. Values in parentheses represent simulated shear viscosities from Vega and Abascal[38]. Bolded ILs are not included in Figure 1.2D, which is restricted to a range that highlights good fits.

| **IL** | **Experimental** | **Simulation** | **Error (%)** |
|---|---|---|---|
| [BAM][$HCO_2$] | 0.070[39] | 0.091 | 29.7 |
| **[BMIM][Ac]** | 0.349[40, 41, 74, 94, 110] | 0.055 | -84.2 |
| **[BMIM][$MeSO_4$]** | 0.201[43, 74] | 0.082 | -59.1 |
| [BMIM][Tf2N] | 0.050[51, 82, 111] | 0.045 | -10.6 |
| [EAM][$HCO_2$] | 0.032[39] | 0.036 | 11.3 |
| **[EMIM][Ac]** | 0.135[41, 53, 112] | 0.027 | -80.4 |
| [EMIM][$EtSO_4$] | 0.098[59, 74, 113-115] | 0.073 | -26.1 |
| [EMIM][Otf] | 0.042[53, 116] | 0.041 | -2.4 |
| [EMIM][$Tf_2N$] | 0.033[82, 117, 118] | 0.036 | 9.3 |
| **[EMPyr][$EtSO_4$]** | 0.156[63, 94] | 0.053 | -66.2 |
| [EMPyr][$Tf_2N$] | N/A | N/A | N/A |
| **[HMIM][Cl]** | 9.408[67, 70] | 0.380 | -96.0 |
| [HMIM][$Tf_2N$] | 0.072[71, 74, 82, 119, 120] | 0.080 | 10.2 |
| [HPyr][$Tf_2N$] | 0.087[75, 76, 94] | 0.059 | -31.7 |
| **[OMIM][Cl]** | 20.933[67, 114, 121] | 0.419 | -98.0 |
| [OMIM][Otf] | N/A | N/A | N/A |
| [OMIM][$Tf_2N$] | 0.091[82, 122, 123] | 0.077 | -15.3 |
| [PAM][$HCO_2$] | 0.078[39] | 0.093 | 19.1 |
| [PPyr][$Tf_2N$] | 0.072[76] | 0.050 | -30.2 |
| Mean Absolute Error | | | 17.8[a], 40.0[b] |
| Max Absolute Error | | | 31.7[a], 98.0[b] |
| **Water Model** | **Experimental** | **Simulation** | **Error (%)** |
| TIP3P | 9.0E-4[38] | 3.1E-4 (3.2E-4[38]) | -65.1 |
| SPC/E | 9.0E-4[38] | 7.0E-4 (7.3E-4[38]) | -22.2 |

[a]Excludes bolded ILs.

[b]Includes bolded ILs.

As Table 1.5 shows for the sets of [$Tf_2N$] and [$HCO_2$]-based ILs, GAFF is able to reproduce the experimental trend of increasing shear viscosity with increasing cation chain length, even despite the significantly large errors that were seen among the triplicate simulations

of [BAM][$HCO_2$] and [PAM][$HCO_2$] (see Figure 1.2D). Additionally, as Figure 1.2D shows, below an experimental shear viscosity of approximately 0.1 Pa•s, GAFF is able to predict IL shear viscosity with surprising quantitative accuracy; over this range of experimental values, which accounts for just over 60% of the total data set, errors range from as little as 2% to a maximum of 34%. Above an experimental shear viscosity of 0.1 Pa•s (bolded ILs shown in Table 1.5), GAFF predicts shear viscosities with deviations of 60-98%; it is important to note that these deviations, however, are no larger than the 65% deviation in the simulated shear viscosity of TIP3P water compared to experiment. Nonetheless, a few similarities can be found among these 6 most deviant IL values. Recurring anions include [Cl], [$RSO_4$], and [Ac]. In the case of the [Cl]-based ILs, we do not recommend using GAFF to calculate transport properties. ILs are generally considered to be highly viscous fluids, and [Cl]-based ILs have significantly higher shear viscosities than most ILs, as Table 1.5 shows. While GAFF does correctly predict that the 2 [Cl]-based ILs have the highest viscosities of the 17 ILs at the temperature used in this study (298.15 K), the simulation dynamics of the [RMIM][Cl] ILs were visibly so slow as to imply that a glass-like state may have been reached. This may reflect that these particular systems did not come close to reaching the hydrodynamic limit, where the shear viscosity is a number rather than a function, due to too small of a box length (and thus too large of a wavenumber). Indeed, a study by Hu and Margulis on the shear viscosity of [HMIM][Cl] showed that with extremely large box sizes and higher temperatures, much better agreement with experiment could be achieved.[104]

*Opportunities for Improvement*. Unfortunately, GAFF does not have force field parameters for some of the most common constituent ions in ILs, such as $AlCl_4$, $BF_4$, and $PF_6$. In such cases, the free tool paramfit can be used to derive unknown force field parameters from *ab initio* calculations.

## Conclusions

Densities, heat capacities, molar enthalpies of vaporization, self-diffusivities, and shear viscosities of 19 ionic liquids have been calculated using the general AMBER force field and compared to experiment. Results show that GAFF is able to reproduce these properties with good accuracy compared to experiment and with similar accuracy compared to other published force fields. No refinement to GAFF was required for such accuracy. Specifically, deviations in simulated IL densities from experimental values were ca. 1-4%, heat capacity deviations ranged from 0-10%, and deviations in molar enthalpies of vaporization from experiment were ca. 6-13%. In general, across the range of ILs tested, GAFF tended to underestimate the densities, heat capacities, and molar enthalpies of vaporization. Significantly larger errors occurred for calculations of IL transport properties; cation and anion self-diffusivity values were all underestimated by GAFF and deviated from experiment in the range of 50-70%. GAFF was, however, able to capture important trends in the data, such as decreasing self-diffusivity with increasing cation chain length, as well as anion self-diffusivities tracking those of the cations. Deviations in IL shear viscosity from experimental values varied widely, from as little as 2% to as much as 98%. The largest errors occurred for ILs with [Cl], [$RSO_4$], and [Ac] anions, and may indicate that those simulations were unable to reach the hydrodynamic limit and thus that larger box sizes and higher temperatures could be used to achieve better agreement with experiment. Excluding these most deviant ILs from the data set, the range of viscosity error is much more

reasonable, with a max absolute error of ~30%. Considering the difficulty of calculating dynamic properties of ionic liquids with simulation[99], GAFF does surprisingly well at predicting these properties and capturing important trends.

After determining that GAFF, with no refinement, is suitable for predicting thermodynamic and transport properties of ILs with reasonable accuracy, we suggest that the protocol we describe in this paper might be of interest to other researchers in the following situations:

1. When IL properties are not known experimentally.
2. To get a good starting point for force field refinement.
3. To screen known ILs for unknown properties, quickly, easily, and efficiently.
4. To guide in the discovery of new ILs.

Because there are many additional applications that feature systems of water/IL mixtures, future research will focus on GAFF's ability to predict thermodynamic and transport properties of water/IL mixtures, as well as mixtures of multiple different ILs.

# II. Structure, Dynamics, and Activity of a Family 11 Xylanase



## Abstract

We have discovered that a family 11 xylanase from *Trichoderma longibrachiatum* maintains significant activity in low concentrations of the ionic liquids (IL) 1-ethyl-3-methyl-imidazolium acetate ([EMIM][OAc]) or 1-ethyl-3-methyl-imidazolium ethyl sulfate ([EMIM][$EtSO_4$]) in water. In order to understand the mechanisms by which the ionic liquids affect the activity of xylanase, we conducted molecular dynamics simulations of the enzyme in various concentrations of the cosolvent. The simulations show that higher concentrations of ionic liquid correlate with less deviation from the starting crystallographic structure. Dynamic motion of the protein is severely dampened by even the lowest tested concentrations of ionic liquid as measured by root mean square fluctuation. Principal component analysis shows that the characteristics of the main modes of enzyme motion are greatly affected by the choice of solvent. Cations become kinetically trapped in the binding pocket, allowing them to act as a competitive inhibitor to the natural substrate. Dynamic light scattering and kinetic studies evaluated the stability of the enzyme in the new solvents. These studies indicate that likely factors in the loss of enzyme activity for this xylanase are the dampening of dynamic motion and kinetic trapping of cations in the binding pocket as opposed to the denaturing of the protein.

## Introduction

Hemicellulose, a major component of biomass, can be hydrolyzed into monosaccharides for fermentation or catalytic upgrading. Xylan, a common hemicellulose, is typically hydrolyzed into constituent xylose subunits by one of two methods that decompose the crucial β-1,4-glycosidic bond by either high temperature acid-catalyzed decomposition[124] or by enzymatic decomposition using a β-1,4-xylanase. The latter requires more costly enzymes but uses less energy and produces less caustic waste.

In both approaches the use of ionic liquids for pretreatment of cellulosic feedstock, which includes hemicelluloses and lignins, has been thoroughly explored. Ionic liquids are organic salts with melting points below 100°C. They can solvate the complex mixtures of cellulose, hemicellulose, and lignin contained in the biomass of interest.[125, 126] By solvating biomass using ionic liquids, more area is exposed for attack by the enzyme, and the reaction proceeds much more quickly.[127] However, a key challenge remains in that the IL and expanded biomass are typically separated prior to enzymatic processing to break down the polysaccharide.

Certain biomass degrading enzymes have been shown to remain active in binary mixtures of water and organic solvents,[128] while some other classes of enzymes have displayed unique characteristics in pure organic solvents.[129] Hyperthermophilic cellulases, for example, have been found to maintain complete activity in levels of up to 20% (v/v) ionic liquid in water.[130] This is counterintuitive, since one might expect a charged solvent of high ionic strength to disrupt the hydrogen bonding patterns of a well-formed enzyme and lead to loss of activity by loss of higher order structure. Moreover, a more viscous medium might be expected to reduce product yield, but this is not always true.[131] Still, the aqueous-IL solvated enzyme performs relatively well even during experiments on the time scale of hours. Other proteins have been shown to maintain

higher order structuring in concentrations of ionic liquid in the range of 20–25% (v/v) ionic liquid in water,[132, 133] indicating they have not denatured. Molecular simulation of the solvated enzyme can provide molecular scale details complementary to modern experimental techniques and indicate the properties of the enzyme and solvent that lead to activity loss. The tools of molecular simulation have been used to study interactions of ILs and cellulose,[134-136] but have seen limited application to IL/enzyme interactions.[137]

Molecular simulation could be used to provide new insight and complement experiments with atomic scale understanding of the enzyme/solvent interactions. Mechanistically, explanations for the IL-induced reduction in apparent reaction rates include solvent/substrate effects, competitive inhibition, enzyme solvation, and disruption of enzyme structure. All of these could contribute to changes in the dynamic motions of the enzyme, interfere with the active site, or affect conformational changes related to the enzymatic mechanism. Evidence suggests that some dynamic motions are related to ligand binding, catalytic activity, and thus apparent reaction rates in general.[138, 139] However, this is a vigorous topic of discussion in the literature and the role of dynamics in enzyme activity continues to be debated.[140] Herein we refer to dynamic motions as those studied on timescales of 50–500 ns using molecular simulations. For biomolecules of the size studied in this work this timescale is not sufficient to probe large-scale structural change, in essence limiting discussion to local conformational changes.

Family 11 glycoside hydrolases (GH11), such as the one we have tested, have been extensively studied both experimentally and computationally in water. Therefore there exists a wealth of quantitative and qualitative data for comparison to new studies in ionic liquids. Previous work has revealed connections between the structure of the protein and its enzymatic activity. The shape of the enzyme (Figure 2.1) is analogous to a left hand with the characteristics

of a thumb, fingers, palm, cord, and a single alpha-helical region. The thumb has been shown to have a profound effect on the binding and cleaving of the substrate.[141, 142] The fingers constitute the bulk of the binding pocket. The palm is the location of the catalytic glutamate and glutamic acid.[143] The cord regulates the entrance of substrate into the binding pocket, and stabilizing mutations of the alpha helix have been shown to increase the thermal stability of the enzyme.[144] Additionally, one study compared several similar xylanases to the specific one studied herein to find the properties that indicate higher thermophilicity.[145] Other computational studies have shown that even at short time scales molecular dynamics (MD) simulations perform well in predicting the relative thermal stability of family 11 xylanases.[146]

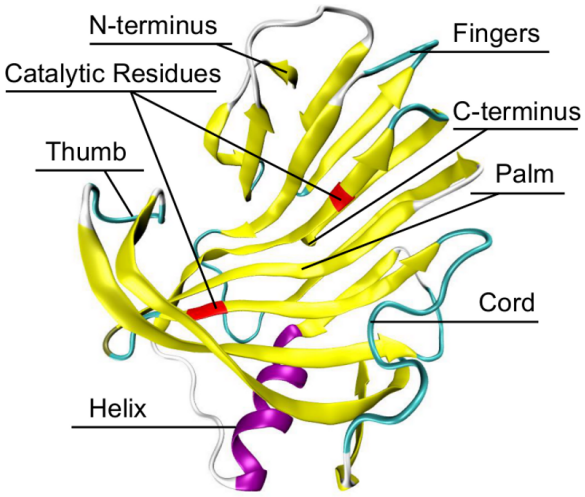



**Figure 2.1:** Family 11 xylanase from *Trichoderma longibrachiatum*. Colored according to secondary structure. Catalytic residues are in red.

To address our significant gap in understanding of the mechanisms by which ILs change the activity of glycoside hydrolases, we have undertaken an extensive computational study. Our representative GH11 enzyme, xylanase II from *Trichoderma longibrachiatum* was studied by systematically changing the IL type, the concentration of the IL-water solution, and the temperature of our simulations. Approximately 4.6 μs of atomic resolution MD simulation was performed at temperatures ranging from 310-363 K. We complement the simulations with

experimental findings that demonstrate partial retention of activity and retention of structure in multiple IL-water solutions.

## Methods

*Enzyme Simulations*. All simulations were performed using widely accepted techniques with AMBER-compatible force fields implemented in NAMD2.7b3[147] with TIP3P[148] water. Protein forces were calculated using AMBER99sb,[149] and ionic liquid force fields were developed with the generalized amber force field (GAFF)[150] using point charges calculated by RESP[151] (Figure S2.1, Table S2.1). The van der Waals cutoff was set to 12 Å with a switching function implemented at 10 Å. Electrostatic interactions were treated with particle mesh Ewald summation using a mesh size of 1 Å. Constraining intramolecular hydrogen bond lengths with the SHAKE algorithm[152] permitted a time step of 2 fs. Temperature and pressure were regulated with a Langevin thermostat and a Nosé-Hoover Langevin piston barostat with a dampening coefficient of 0.5 $ps^{-1}$ and a piston decay period of 2 ps. Replicate simulations were performed by reinitializing velocities from minimized structures. More details about the simulations and verification of the force field are included in the Supporting Information.

We selected xylanase from *Trichoderma longibrachiatum* (PDB ID: 3AKP)[153, 154] as the model system. The first of the two chain sequences provided in the PDB file was used for all simulations. Crystallographic waters were maintained. We selected the ionic liquids (a) because of the ability of established force fields to reproduce experimental data for similar structures, (b) so that we could compare the results to other experimental studies of glycoside hydrolases in IL,[128, 130] and (c) because of the promising abilities of [EMIM][OAc] as a solvent for cellulosic biomass.[155]

*Xylose Yield*. Xylanase from *Trichoderma longibrachiatum*, xylan from birchwood, 1-ethyl-3-methylimidazolium acetate [EMIM][OAc] (90% purity), and 1-ethyl-3-methylimidazolium ethyl sulfate [EMIM][$EtSO_4$] (95% purity) were purchased from Sigma Aldrich. Premixed concentrated PBS buffer was obtained from Calbiochem and diluted to a concentration 150 mM. Enzymatic hydrolysis of xylan was carried out in 5 g of solution with varying concentrations of ionic liquid in buffer. For all experiments, 1 mg of enzyme was rehydrated in the aqueous buffer portion of the solution at its working temperature for 30 minutes. This is meant to help the system overcome any kinetic barriers associated with being placed in a new solvent. Next, 100 mg of xylan was added, and the solution was incubated in a thermostatic shaker table at the working temperature with a rotation speed of 150 rpm. Two samples of 50 μL were extracted from the reaction solution after 24 hours of reaction time. The samples were diluted fifteen fold and centrifuged before they were analyzed for xylose content using HPLC or a Multiparameter Bioanalytical System. Specific details are included in the supporting information. The procedure was repeated for an incubated system, where the enzyme was incubated in the IL-water mixture for 24 hours prior to the addition of substrate. To facilitate easy comparison across all experiments, we defined the yield of this reaction relative to the baseline case of an aqueous solution without incubation.

*Dynamic Light Scattering*. A Malvern Zeta Sizer ZS was used to determine the presence and size of proteins and aggregates in solution. Xylanase from T. longibrachiatum was obtained from Hampton Research. We chose to use an enzyme from a different supplier for the DLS studies for two reasons. First, the lyophilized enzyme used for the kinetic study displayed some aggregation even in simple PBS buffer, which was likely caused by damage due to lyophilization and rehydration. This confounded the ability of DLS to detect natively folded enzymes. Second,

crystal structures of the enzyme used for DLS studies showed well-bound glycerol near the binding pocket. This had the potential to affect any kinetic studies by acting as a competitive inhibitor. The amino acid sequences of the two batches are identical, but differences in behavior can arise from the differences in preparation techniques (lyophilization vs. non-lyophilization) and the mixture of other chemical additives present with the enzyme. For the DLS studies, the stock solution of xylanase was diluted one hundred times to achieve a concentration of 1 mg ml–1 of protein in each concentration of ionic liquid. The first set of samples was allowed to sit for 30 minutes before sizing. A second set of samples was incubated at 310 K for 48 hours. This set of samples was passed through a filter of pore size 0.64 μm, and each was allowed to sit for 30 minutes before sampling. All light scattering experiments were conducted five times on an individual sample at 310 K using Brand disposable UV microcuvettes, single back scattering at 173°, and viscosity and refractive index parameters obtained by linear interpolation from literature or manufacturer data.[59, 156]

## Results and Discussion

*Determining Xylanase Activity in IL.* The enzyme maintained significant activity in both [EMIM][$EtSO_4$] and [EMIM][OAc] (Figure 2.2). Incubation in low concentrations of ionic liquid does not adversely affect the yield when compared to the loss in yield for the analogous aqueous system. In fact, for the lowest tested concentration of IL, the [EMIM][$EtSO_4$] incubated system yields more xylose as product than the water incubated system. The relationship between product yield and IL concentration is not simply linear. Additionally, we observed increased turbidity over the timespan of the experiment for some IL-containing systems, which could be caused by the aggregation of the enzyme. This hypothesis is tested by DLS in the following section.

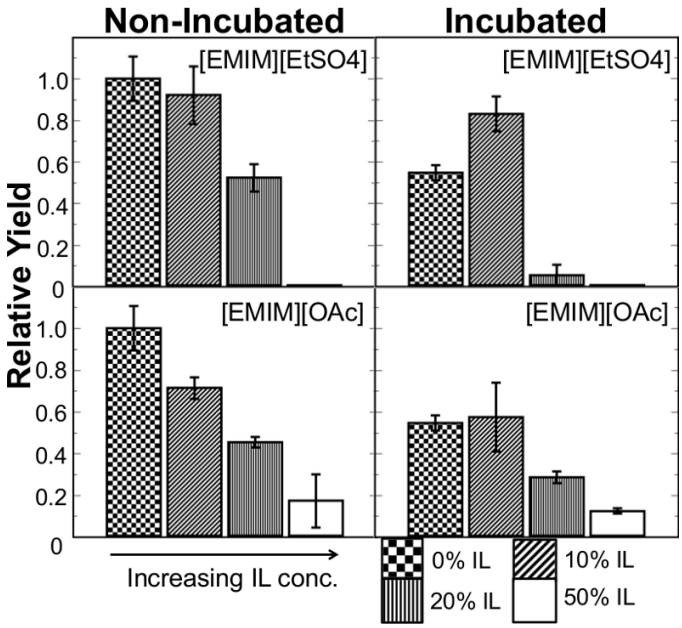



**Figure 2.2:** Experimental yield of xylose. Xylanase in ionic liquid containing systems with and without substrate-free incubation. Relative yield measured versus a baseline of the non-incubated water system. Error bars are the standard deviation six measurements (three independent samples measured in duplicate).

*Effect of IL on Xylanase Structure*. An analysis of the root mean square deviation (RMSD) of the Cα atoms indicates that the systems with higher concentrations of ionic liquid generally maintain closer agreement to the experimentally determined crystal structure of the enzyme over the time scale studied (Figure 2.3A,B). Kinetic trapping has been a problem reported by other researchers simulating enzymes in IL on the timescale of up to 20 ns.[137] Yet these trends hold over the full range of variables tested and over replicate simulations. Loops in particular begin to transition between similar but different folded states. Even at elevated temperatures of up to 363 K, highly ionic systems show few signs of denaturing or disruption of higher order structuring. Two of the most interesting features in the simulations occur at elevated temperatures. First, as the temperature increases for the aqueous system, the fluctuations near the helix (THR152 to GLN162) increase in magnitude. This behavior may explain the loss of activity at high temperatures for similar GH11 proteins and the increased stability afforded by

stabilizing mutations in the helix as shown by Turunen and coworkers[157] who added a disulfide bridge near the N-terminus of the helix by mutating SER110 and ASN154 to cysteine. Second, in the highest temperature system, 20 wt% [EMIM][$EtSO_4$], the fluctuations around the thumb (GLN121 to THR133) become very large compared to the rest of the system and compared to its counterpart in 20 wt% [EMIM][OAc]. In addition, the distance between the fingers and the thumb increases. This behavior may indicate an approaching limit of stability for the system and predict a severe loss in enzyme activity near this concentration.

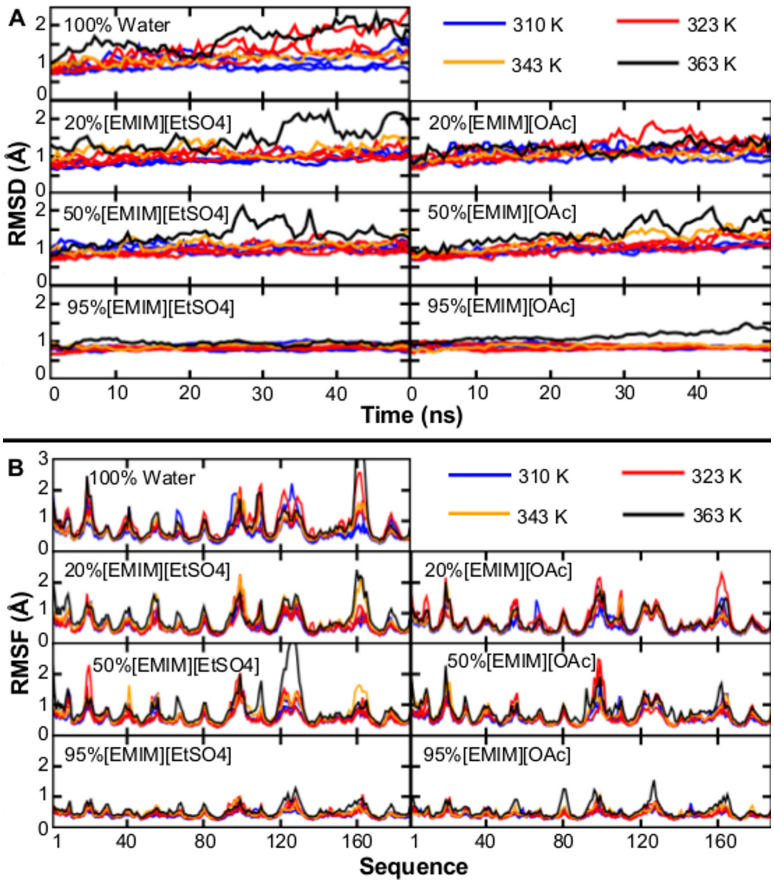



**Figure 2.3:** Root mean square deviation (RMSD) [A] and fluctuation (RMSF) [B] of xylanase alpha carbons across a range of cosolvent concentrations and temperatures. Triplicate simulations at 310 and 323 K are shown.

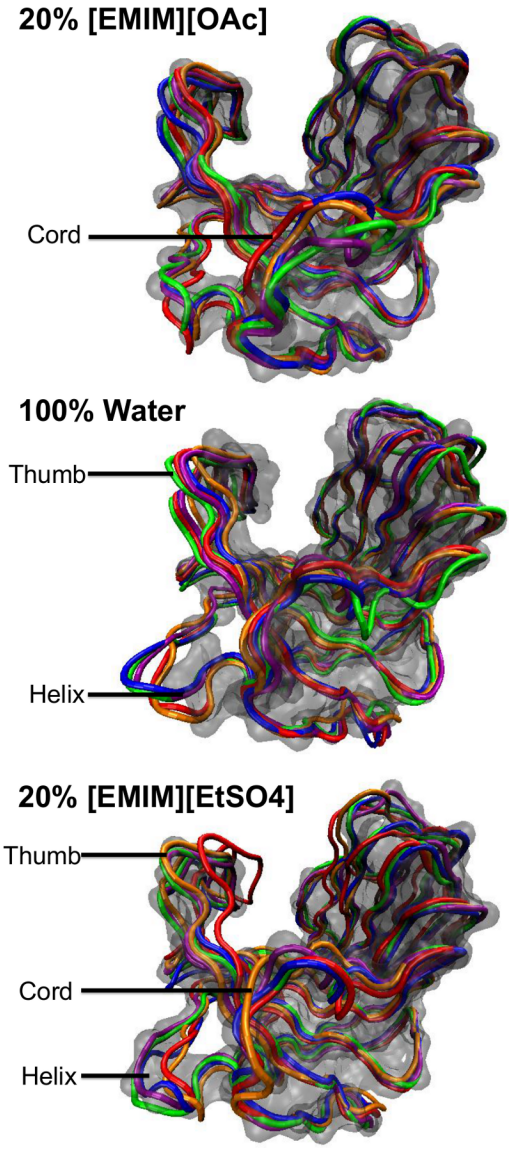



**Figure 2.4:** Snapshots of the molecular dynamics trajectories (310 K). Areas of interest are indicated. Grey cloud (minimized structure). Red (100 ns). Orange (200 ns). Green (300 ns). Blue (400 ns). Violet (500 ns).

Several simulations were extended to 500 ns of simulation time in order to obtain more sampling of conformational changes in the enzyme (Table S2.2). The results from 50 ns are similar at a timescale 10 times longer. The enzyme remained stable in IL, as measured by RMSD, which never exceeded 2 Å on a Bezier smoothed plot (Figure S2.2). On this timescale,

systems of 20 wt% IL and the water system show similar displacement from crystallographic structure. Figure 2.4 contains snapshots of the systems at 100 ns intervals and highlights the features of each system that deviate the most from the experimentally determined structure. The cord was the most active of the main features. This was likely due to its interactions with the cation. Moreover, in water at the elevated temperature of 363K, the enzyme did not have an RMSD greater than 3 Å. The residues following the N-terminus of the helix showed particularly high thermal activity compared with the rest of the protein. More specific differences between these systems are discussed in the following sections.

In addition to the studying the atomic scale details of GH11 at the sub-microsecond timescale, we probed the solvent-induced features using light scattering, which can be used to roughly estimate changes in the size of the enzyme and its aggregates. Use of more direct measurements of the protein structure (e.g., circular dichroism) in ILs is confounded by the absorbance spectrum of the ions.[158] We employed DLS to gain semi-quantitative insights into whether the enzyme is undergoing aggregation or large-scale denaturation in ionic solvents. Other scattering techniques, such as SAXS and SANS, have shown that other proteins maintain much of their secondary and tertiary structure in ionic liquid concentrations of up to 20% (v/v).[159] The measurement of particle size with this approach is very sensitive to the parameters of refractive index and viscosity of the solvent. We interpolated these data from previously reported values across both temperature and composition and therefore use the DLS results primarily to gain mostly qualitative insights.

The enzyme's hydrodynamic radius was measured before and after incubation across the range of concentrations studied by simulation and in 8M urea (Table 2.1). We interpreted the data to determine the presence or absence of proteins near their aqueous hydrodynamic radius (of

about 2 nm) and the presence of much larger aggregates. Additionally, the samples were inspected for turbidity that would also indicate large aggregates. DLS indicates that the buffered aqueous system maintains most of its protein near its native state even after incubation, and no turbidity is apparent in either the DLS or the kinetic studies. At the same time, incubation of the enzyme led to a more than 40% drop in the yield of product for the kinetic studies (c.f., Figure 2.2). Two possibilities to explain the seemingly apparent contradiction are that (a) either the structural changes leading to loss in activity are too small to be detected by DLS or that (b) the native protonation state of the lyophilized enzyme used in the kinetic studies was changed from a more optimal state to a less optimal state after incubation. In the first case, a more sensitive scattering method could be used. In the second case, we observe that 10 wt% [EMIM][$EtSO_4$] affords some protection from this loss in activity even though its pH of 8.1 is farther from the optimum of about 5.5 suggested by the manufacturer. The presence of small amounts of sulfuric acid, hydrogen sulfide and ethanol, which are the products of the slow degradation of [$EtSO_4$] in water,[160, 161] could lead to a more primed state for the protonated catalytic residue.

Interestingly, DLS indicates that there is a critical range of concentrations of IL where aggregation occurs. Using [EMIM][$EtSO_4$] as an example, at a concentration of 10 wt%, aggregation is not detected by DLS or by turbidity both before and after incubation. Conversely, at 20 wt% aggregation is not detected immediately. At 50 wt%, aggregation is very rapid, and the aggregates or the incubated sample are large enough to be completely filtered, yielding no measurement of protein in the sample. We were unable to determine whether these aggregates were multimers of natively folded enzyme or if they are constituted of unfolded enzyme. At the highest tested concentration of 95 wt% IL, the enzyme displays a normal hydrodynamic radius and no aggregation is detected by DLS or by turbidity. The formation of large protein aggregates

is not generally accessible with atomistic MD simulations due to scale limitations in both size and time, but insights into the propensity for aggregation can be gained through the investigation of solvent-protein interactions displayed in the simulation.

**Table 2.1:** Dynamic light scattering (DLS) analysis of xylanase. Enzyme solvated in various wt% concentrations of two ionic liquids, water, and urea. Hydrodynamic radius ($R_h$) and peak width are taken from an average of measured peaks below 250 nm.

| | No Incubation | | Turbid | 48 hr Incubation | | Turbid |
|---|---|---|---|---|---|---|
| | $R_h$ (nm) | width (nm) | yes/no | $R_h$ (nm) | width (nm) | yes/no |
| 100% Water | 1.91 | 0.70 | no | 1.72 | 0.59 | no |
| 10% [EMIM][OAC] | 1.75 | 0.58 | no | 1.44 | 0.32 | no |
| 10% [EMIM][$EtSO_4$] | 1.61 | 0.36 | no | 1.60 | 0.43 | no |
| 20% [EMIM][OAC] | 1.93 | 0.53 | no | 1.65 | 0.30 | no |
| 20% [EMIM][$EtSO_4$] | 1.64 | 0.51 | no | 33.59 | 6.64 | yes |
| 50% [EMIM][OAC] | 56.63 | 18.48 | yes | 51.68 | 12.89 | yes |
| 50% [EMIM][$EtSO_4$] | 26.43 | 7.58 | yes | - | - | yes |
| 95% [EMIM][OAC] | 1.10 | 0.20 | no | 3.65 | 0.59 | no |
| 95% [EMIM][$EtSO_4$] | 2.11 | 1.33 | no | 1.76 | 0.47 | no |
| 8M Urea | 2.11 | 0.46 | no | 3.36 | 0.74 | no |

In the simulations we observed unique interactions between the anion and the positively charged surface residues of the enzyme. Specifically, lysine and arginine tend to attract the negatively charged oxygen atoms of acetate and ethyl sulfate. White and coworkers have suggested that charged surface residues, specifically lysine and glutamic acid, help deter nonspecific interactions including aggregation.[162] Therefore, interruption of the environment around these charged residues may lead to higher levels of aggregation, as in the case of [EMIM][$EtSO_4$] but not in the case of [EMIM][OAc]. The average number of anion-oxygen atoms within 4 Å of lysine, arginine, or histidine was tracked over the length of the 500 ns simulations for 20 wt% solutions of both [EMIM][OAc] and [EMIM][$EtSO_4$]. For acetate, 15 ± 5 oxygen atoms occupied the space, while for ethyl sulfate, 27 ± 7 occupied the space. High concentrations of these anions around the positively charged residues may also facilitate the salt

bridging of separate proteins and encourage aggregation. On the other hand, the simulations do not indicate an affinity of the cations for the negatively charged residues, which we expect is caused by much greater charge delocalization on the cation compared to the anions.

*Changes in Structure and Dynamics*. In order to quantify the magnitude of deviation from the average position for specific sites along the backbone of the enzyme the root mean square fluctuation (RMSF) of the Cα atoms in the enzyme was computed. First, the RMSF was calculated over an equilibrated sample of each trajectory in order to identify regions of interest. We observed that the regions of highest fluctuation for the enzyme in water match the features of the enzyme that have been previously identified as important to enzymatic activity. The main features of the enzyme's Ca RMSF (Figure 2.3C,D) are centered about the following residues: the thumb, ARG128; the alpha helix, ASN157; the fingers HIS22, SER42, GLY70, and SER181; the cord, PRO98. The palm constitutes most of the remainder of the structure. Four of the five main features (the thumb, the alpha helix, the fingers, and the cord) show the highest levels of fluctuation in 100% water. These regions have specifically been implicated in the proper binding of the substrate and the release of the product.[163-165] A larger dynamic range of these features may help the protein move the substrate to its proper binding site and eject the product from the binding pocket. The fifth region is the palm, which is the active site of GH11 xylanase. Since biocatalytic activity requires careful arrangement of the catalytic sidechains, large scale fluctuations within this region could conceivably affect reaction rates.

In addition to the average fluctuations in an equilibrated portion of each simulation, we used the simulations that were extended to 500 ns to study how the fluctuations changed over time. The RMSF was measured around the arithmetic average structure over one-nanosecond windows (Figure S2.3). In the aqueous systems, the temporal RMSF plots show activity in the

region between the thumb and the cord that is notably less pronounced in the two ionic systems. As detailed below, we observed that some simulations with IL show the presence of a cation trapped near the chord. This might disrupt the natural motion of the chord region observed in aqueous simulations. The same windowed RMSF analysis was performed for the entire set of data, including the 50 ns runs. Even at the highest tested temperatures of 363 K, the ionic systems do not display the same characteristic fluctuations as the aqueous system. At this temperature, the aqueous protein fluctuates with a greater magnitude, and the ionic systems are still much more rigid.

Multiple researchers have implicated the dynamic motions of enzymes as being a key to their activity.[166] By moving in a semi-coordinated fashion and specifying the distribution of states that the sidechains occupy, the enzyme can (a) efficiently bind substrate, (b) direct it toward the catalytic site, (c) possibly aid in the chemical reaction by inducing new geometries or electronic structures that stabilize transition states, and (d) release the product. The introduction of a new solvent, when it does not denature the protein or cause kinetic trapping, can interrupt this order and disrupt any of the steps along this process, thereby leading to changes in activity and yield of product. In the following section, solvent-induced changes in the dynamics of the enzyme and in the coordination of its residues' motions will be discussed. This assumes that the system is at a new equilibrium state in the IL systems and that it is not kinetically trapped.

*Principal Component Analysis*. Using principal component analysis (PCA), the MD trajectories at 310 K were projected onto representative sets of eigenvectors in order to gain insight into the slow modes of motion. We used PCA study effect of solvents on correlated and anti-correlated motions in the enzyme. The covariance matrices were calculated for systems containing the lowest simulated concentrations of ionic liquid versus one of a purely aqueous

system (Figure 2.5). As can be seen in the ionic systems, correlated motions are not completely preserved. Differences can be seen in the location of the regions of correlation as well as the size of the spans containing correlated motions. We would not expect the modes to completely disappear as the enzyme does retain activity in the IL solvent. Moreover, regions displaying little correlation in the aqueous system, such as region spanning the thumb and the helix, show more pronounced correlation and anti-correlation in the ionic systems. This potentially affects the ability of the enzyme to carry out its function as discussed above.

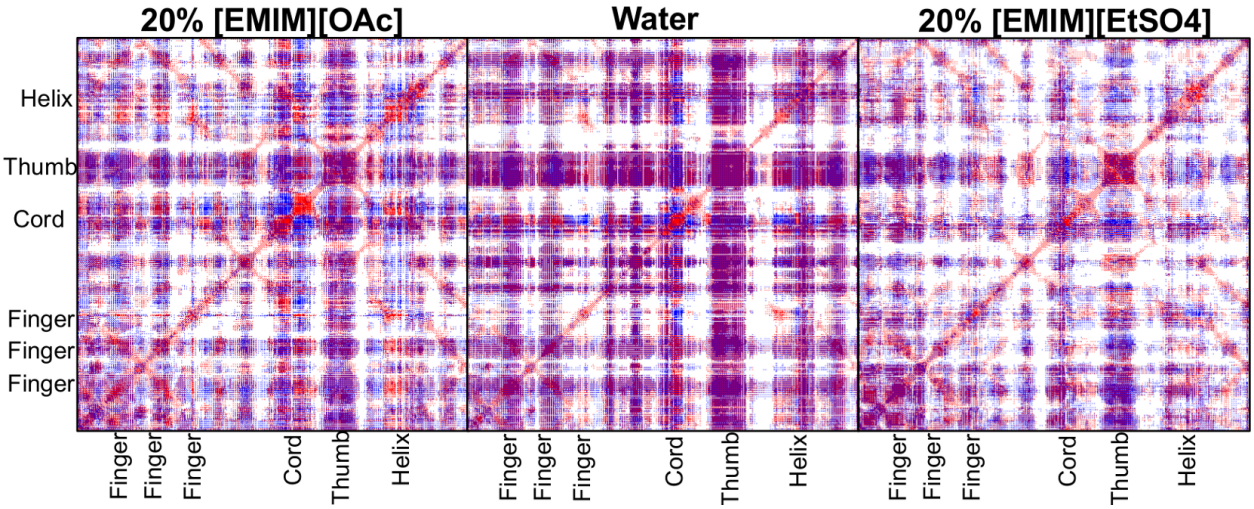



**Figure 2.5:** Covariance matrices of xylanase. Backbone motion in ionic liquids in xyz resolution. Red pixels are correlated while blue are anti-correlated.

Next the PCA modes were visually represented via porcupine plots by projecting the first and second eigenvectors onto the alpha carbons of the protein backbone (Figure 2.6). In order to confirm that these modes are significant, the contribution of each mode was plotted (Figure S2.4). For each of the samples taken in equilibrated regions of the trajectory, as determined by inspection of RMSD, systems with higher concentrations of ionic liquid display disrupted motion. The vectors are different in orientation, and the ordering of primary, secondary, and higher modes of motion from the aqueous systems is not maintained. Inspection of the trajectory of the aqueous systems reveals a shuffling motion of the hand-like enzyme as the thumb and

fingers move in concert, as depicted by the dominant PCA mode in Figure 2.6. Conversely, the addition of IL changes this, and a new dominant mode takes over. The dominant motion is a slow opening and closing of the hand-like structure, which accounts for a large portion of the displacement from the crystal structure. The dominant mode in the aqueous system is preserved in the IL, but it is consistently present with a noticeably higher eigenvalue in the aqueous system. This is consistent with our expectations, as any motions related to enzymatic function would need to be preserved in the IL since it shows activity experimentally. We observed these effects over all 100 ns windows of analysis we investigated from the long 500 ns trajectories.

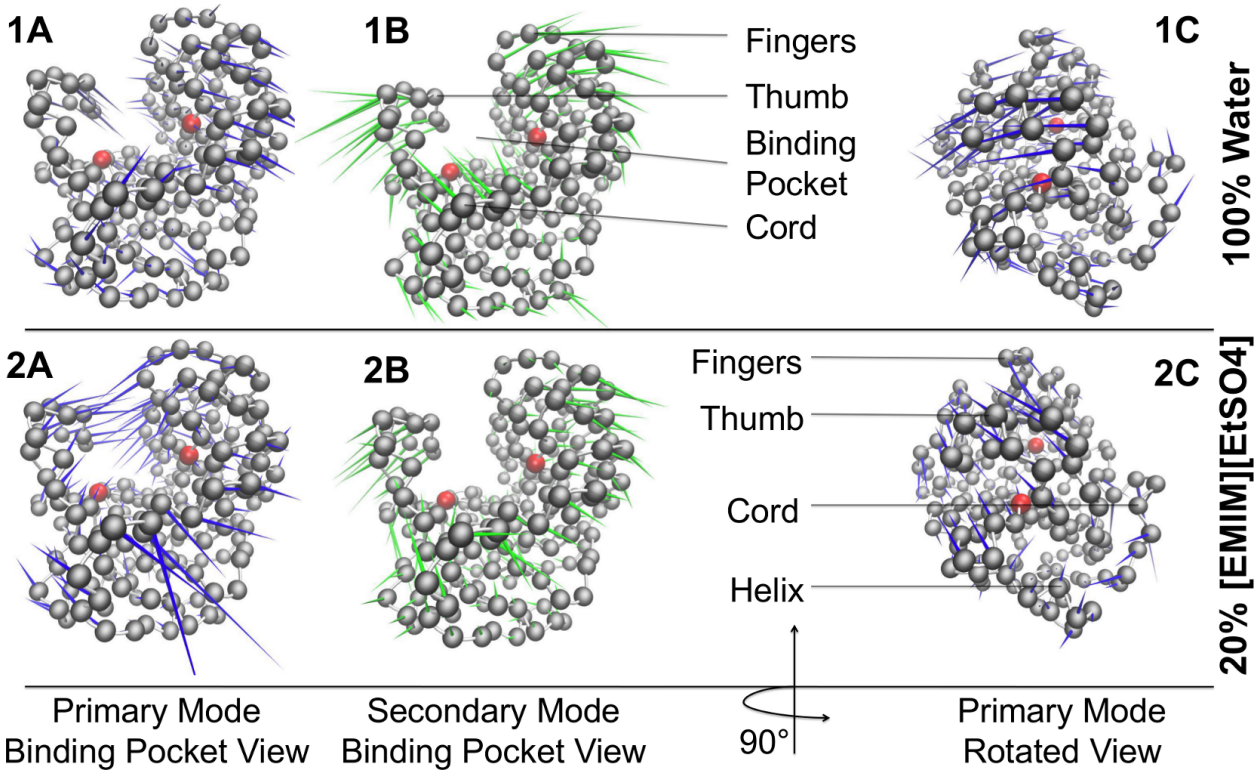



**Figure 2.6:** Principle components of xylanase motion. The aqueous enzyme (1A,1B,1C) is compared to its counterpart in 20 wt% [EMIM][$EtSO_4$] (2A,2B,2C). Columns A and B displays a view through the binding pocket, while column C displays a rotated view. Primary modes are blue, while secondary are green. Catalytic residues are red.

*Competitive Inhibition from Cations*. Since the active site is charged, we next investigated whether the strong interactions between solvent and catalytic residue could disrupt the catalytic cycle. The large number of possible instigators for the loss of enzymatic activity in these systems makes the task of experimentally verifying and measuring inhibition difficult by conventional means such as double reciprocal plots. The simulation trajectories were analyzed for the presence of organic ions in the binding pocket. For simulations containing either of the ILs, two or three cations preferentially occupy space very near the binding pocket (Figure 2.7). We observed this for all concentrations and temperature ranges investigated in this study. The cation is roughly similar in shape and size to the enzyme's natural product of xylose, but the composition of the imidazolium ring differs very much from that of the xylose ring.

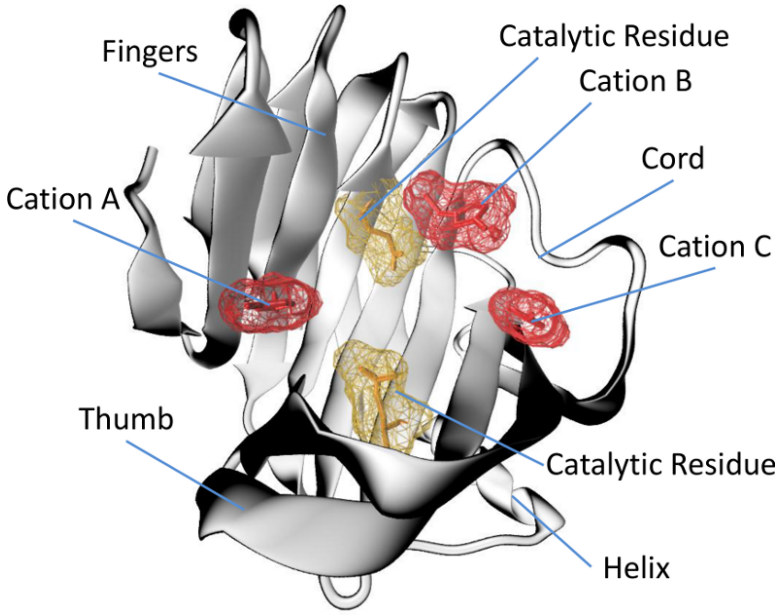



**Figure 2.7:** Kinetically trapped cations in the xylanase binding pocket. Catalytic residues are yellow. Cations are red.

It is therefore interesting to note that, during the inspection of the trajectories, it was incredibly rare for either of the first two cations to exit the binding region once they have become bound; they become kinetically trapped, at least on the 50–500 ns time scale. On the rare

occasion that a cation leaves the binding pocket, it is quickly replaced by another cation. Because of how quickly the pocket became inhabited with cations at the beginning of the simulation, and because of the observed low likelihood of cations leaving the pocket over the timescales studied, competitive inhibition could play a role in the loss of activity. No kinetically trapped anions were observed. As for the crystallographic waters, they leave the surface quickly. Yet, if the number of water molecules near the surface is tracked as a function of time, most IL-containing simulations converge to a similar number of molecules and then slightly fluctuate from there (Figure S2.5). Exclusion of water from the catalytic site is also not observed.

## Conclusions

The activity, stability, and dynamics of a family 11 glycoside hydrolase solvated in mixtures of ionic liquid and water were characterized by molecular simulation and experiment. Several hypotheses for the partial deactivation the enzyme were tested. Analysis of simulations and experimental data lead to several observations about the interactions of IL with this enzyme.

1. GH11 from T. longibrachiatum solvated in higher concentrations of ILs generally remains more stable with respect to its crystal structure than when solvated in water on the time scales studied.

2. The solvent environment of IL-containing systems causes fluctuations of the enzyme to be muted when compared to a purely aqueous system.

3. Natural correlated motions of the enzyme are not preserved as concentration of IL increased.

4. [EMIM] binds strongly in the binding pocket, potentially reducing activity by inhibition.

5. Aggregation must be considered when mixing enzymes and ILs.

Future work in this area should decouple the role of the ionic strength and viscosity of the solutions, both simulation and experiment both offer possible routes to do so. Our work also suggests that more insights are needed into the role of cations in the substrate binding pocket.

# III. Comparison of Three Ionic Liquid Tolerant Cellulases



*Acknowledgement of coauthors:* Dr. Patrick Burney's contribution was largely in the analysis of the data and the production of figures. All other work was conducted by me. What follows is the full text of our paper.

## Abstract

We have employed molecular dynamics to investigate the differences in ionic liquid tolerance among three distinct family five cellulases from *Trichoderma viride*, *Thermotoga maritima*, and *Pyrococcus horikoshii*. Simulations of the three cellulases were conducted at a range of temperature in various binary mixtures of the ionic liquid 1-ethyl-3-methyl-imidazolium acetate with water. Our analysis demonstrates that the effects of ionic liquids on the enzymes vary in each individual case from local structural disturbances to loss of much of one of the enzyme's secondary structure. Enzymes with more negatively charged surfaces tend to resist destabilization by ionic liquids. Specific and unique structural changes in the enzymes are induced by the presence of ionic liquids. Disruption of the secondary structure, changes in dynamical motion, and local changes in the binding pocket are observed in less tolerant enzymes. Ionic liquid induced denaturation of one of the enzymes is indicated over the 500 ns time scale. In contrast, the most tolerant cellulase behaves similarly in water and in ionic liquid containing mixtures. Unlike the heuristic approaches that attempt to predict enzyme stability using macroscopic properties, molecular dynamics allows us to predict specific atomic level structural

and dynamical changes in an enzyme's behavior induced by ionic liquids and other mixed solvents. Using these insights, we propose specific experimentally testable hypotheses regarding the origin of activity loss for each of the systems investigated in this paper.

## Introduction

Lignocellulosic biomass is an important potential feedstock for the sustainable production of liquid fuels and other chemicals. Development of technologies toward effective production of biofuels from these substrates has led to the requirement of pretreatment processes to improve viability [167]. Pretreatment is a general term for the processing steps used to convert raw cellulosic biomass to a form that is suitable for enzymatic hydrolysis. This includes but is not limited to the solubilization of cellulose and hemicellulose and the separation of lignin, which inhibits enzymatic activity [168, 169]. Several methods of pretreatment have been developed (e.g., steam explosion, ammonium fiber explosion, dilute acid solutions, lime solutions), each with advantages and disadvantages. A detailed review of pretreatment methods has been provided by Abor et al. [170]. In summary, the high temperatures, high pressures, or caustic chemicals utilized in these processes are undesirable from the standpoint of capital and energy costs, safety, equipment lifetime, and environmental impact. To varying degrees, these processes also produce byproducts that interfere with enzymatic hydrolysis such as phenolic compounds and furan derivatives, degrade potentially useful secondary substrates, or causes loss of saccharides [171, 172]. It would be preferential to develop new pretreatment methods that run at lower temperatures, at atmospheric pressure, with fewer hazardous chemicals, and most importantly that allow hydrolase activity without neutralizing or changing the solvent from step to step.

Many researchers have demonstrated that ionic liquids (ILs), organic salts that have melting temperatures near 100 °C, can completely solubilize cellulosic biomass – separating the

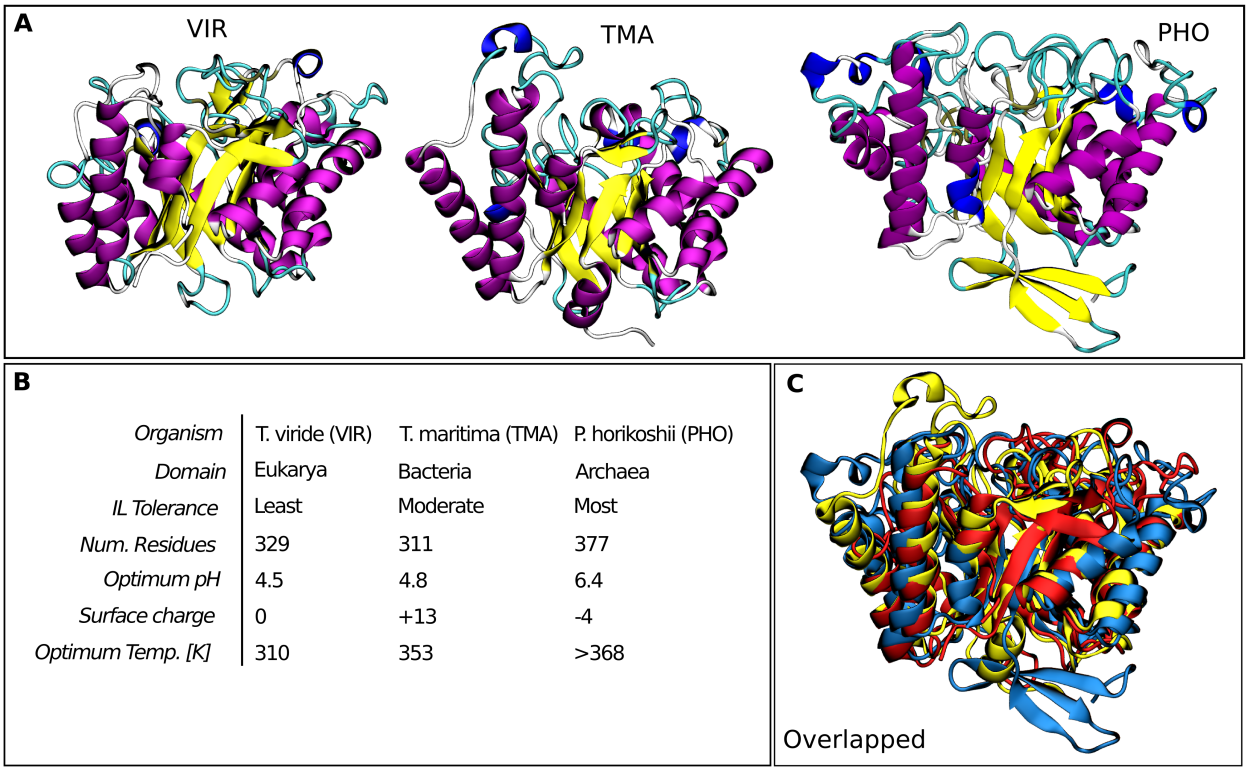
hemicellulose and cellulose from lignin while overcoming the recalcitrance of crystalline cellulose [173-175]. In addition to the rare ability to dissolve crystalline cellulose, ionic liquids are relatively benign compared to the chemicals used in other pretreatment processes such as sulfuric acid, lime, and ammonia. The chemical and thermal stability of ILs coupled with their low vapor pressure gives them the ability to be recycled through the process many times without loss, and effective separation of ILs from biomass has been demonstrated. IL pretreatment can operate at lower temperatures and atmospheric pressures while posing less of an environmental risk than other pretreatment processes. This facilitates the use of less robust equipment and fewer safety precautions. Therefore ILs have been proposed as an alternative and "green" pretreatment solvent [176-180].

One aspect of an IL based pretreatment scheme is that many imidazolium based ILs can support enzymatic catalysis at a variety of temperatures both as pure solvents and in binary mixtures with water. When used as solvents for biochemical reactions, the right choice of IL can improve our ability to control many properties of the reaction, especially substrate and product solubility. Some hydrolases such as chymotrypsin retain activity in pure ILs [181], and others such as certain lipases exhibit enhanced thermal stability [182, 183], selectivity [184] and enantioselectivity [185, 186] compared to traditional solvents. Yet, there are no known cellulases that can tolerate ILs at high concentrations. Instead, these enzymes tend to deactivate at much lower concentrations of IL in water. Specific examples will be discussed later. Still, the potential to dissolve cellulosic biomass while maintaining hydrolase activity in low concentrations of IL affords a unique opportunity to avoid a secondary separation step during biomass pretreatment. However, the enzyme cocktails used to hydrolyze solubilized biomass show insufficient or even no activity in many potentially useful ILs. Without discovery of a method for improving enzyme activity in the

current ILs or of novel ILs that permit sufficient enzyme activity, improving the viability of biocatalytic conversion of biomass using ILs will be greatly hindered.

Glycoside hydrolase (GH) enzymes catalyze the degradation of long chain carbohydrates such as cellulose and xylan, whose subunits are key substrates for the production of biofuels and other value added chemicals. GHs are an important target for engineering ILs or the enzyme itself toward improved activity in the pretreatment of biomass. Recent publications from several research groups have provided important results toward discovering both ILs capable of supporting GH enzymes and enzymes that exhibit improved activity in ILs. In particular, Wang et al. have demonstrated up to 65% retention of activity for a cocktail of GH enzymes in 20 wt% 1-ethyl-3-methyl-imidazolium acetate ([EMIM][OAc]) in water [128]. Zhang et al. in a separate publication have reported full or even enhanced activity of a cellulase (family 5 GH) from a halophilic microbe in 20 wt% [EMIM][OAc], [EMIM][Cl], [BMIM][Cl], and [AMIM][Cl] relative to its activity in water [187]. This group later demonstrated activities for three other GH family 5 cellulases that vary considerably in their ability to function within increasing concentrations of [EMIM][OAc] in water [188].

These preliminary studies offer a wealth of research opportunities toward further understanding the interaction between ILs and biomass hydrolyzing enzymes. Of particular interest are the results showing that structurally similar cellulases from three separate species have drastically different activity in increasing concentrations of [EMIM][OAc] [188]. The three enzymes are: a cellulase from *Trichoderma viride*, cellulase 5A from *Thermotoga maritima*, and endoglucanase from *Pyrococcus horikoshii*.




| | T. viride (VIR) | T. maritima (TMA) | P. horikoshii (PHO) |
|---|---|---|---|
| *Organism* | | | |
| *Domain* | Eukarya | Bacteria | Archaea |
| *IL Tolerance* | Least | Moderate | Most |
| *Num. Residues* | 329 | 311 | 377 |
| *Optimum pH* | 4.5 | 4.8 | 6.4 |
| *Surface charge* | 0 | +13 | -4 |
| *Optimum Temp. [K]* | 310 | 353 | >368 |

**Figure 3.1:** (A) Renderings of the initial configurations for each enzyme. (B) Details highlighting their differences. (C) Aligned structures overlapped. Blue - PHO. Yellow – TMA. Red – VIR.

Hereafter referred to as VIR, TMA and PHO, respectively, structures and relevant details of these enzymes are provided in Figure 3.1, A and B. Optimum temperature and pH are taken from Datta et al [188]. Surface charge is calculated as described in the Methods section. Even though the enzymes are all classified as the same family, the organisms they are extracted from belong to three different domains of life: Eukarya, Bacteria, and Archaea respectively. VIR, an industrial benchmark for cellulase activity in water, is mesophilic; both TMA and PHO are identified as hyperthermophylic enzymes. Figure 3.1C superimposes the structures over each other. MultiSeq [189] was used to analyze the differences in the sequences and structures among the three enzymes. The root-mean-square displacement (RMSD) of the coordinates of mutual residues and the percentage of identical residues for VIR and TMA are 2.3 Å and 13.1%, for VIR

and PHO are 2.9 Å and 9.2%, and for TMA and PHO are 2.6 Å and 14.8%. Twenty-five residues are conserved across the three structures, all of which are near the beta-barrel that contains the catalytic residues.

Datta and coworkers hypothesized that thermophilic enzymes will tolerate ILs more readily than their mesophilic relatives. The presence of a higher number of ion pairs (that is charged amino acids) in the structures of hyperthermophilic enzymes has been observed [190]. These ion pairs form salt bridges that stabilize the protein against unfolding at high temperatures. Datta's results provide details on the macroscopic properties of the enzymes such as activity and melting temperatures. They identified the two thermophilic enzymes, TMA and PHO, as IL tolerant to some degree, but their methods do not allow them to gain molecular level insight into the structure and dynamics of the enzymes themselves. The results provided by Datta and coworkers provide us with the ability to connect the fine detail provided by molecular simulation with relevant macroscopic properties such as thermophilicity and IL tolerance.

Molecular simulation, in particular molecular dynamics (MD) can be used to study the effects of ILs on GHs at atomistic resolution and help to resolve several of the hypotheses previously put forth in the literature. A few MD studies have been conducted on the role that ILs play in the activity and stability of enzymes and proteins [137, 191-194]. One study compares several GH5 enzymes solvated in water using metrics taken from MD trajectories. The researchers were able to demonstrate the feasibility of MD-guided protein engineering by suggesting specific mutations that stabilized a GH5 enzyme at high temperatures [195]. Based on Datta's hypothesis about thermal stability being connected to IL tolerance, mutations that stabilize the protein at high temperatures are relevant to our goals. Our previous simulations in the area of IL-enzyme compatibility analyzed the stability and fluctuations of two different hydrolases', *Trichoderma*

*longibrachiatum* GH11 and *Candida rugosa* lipase, structures in IL containing solvents [18, 196]. We suggested that the presence of IL in concentrations of up to 20 wt% does not appreciably denature the enzyme in simulations on time scales of up to 500 ns. We also suggested that the dynamics of the structure as measured by root-mean-square fluctuation (RMSF) and principal component analysis could be used as a predictor for the maintenance of activity of the enzyme in IL containing solvents. In our MD study of GH11 we observed that the tested cation molecules penetrate the binding pocket of the enzyme, and we predicted that this would competitively inhibit the enzyme's function. This hypothesis was supported by experiments that observed the inhibition of a closely related GH11 by the same cation [197]. Herein, we will test some of these previous hypotheses (structural stability, fluctuations, cation-catalytic residue interactions) and develop new hypotheses with different enzymes of significantly different secondary structure and a larger set of data. This will allow us to more thoroughly analyze enzyme structure and dynamics over larger time scales as well as observe less common events. In this study we focus on the observations that ILs destabilize (or even denature) the enzyme or disrupt the flexibility of the structure that is critical to the enzyme activity.

Previous simulation studies have given some insight into the relationships between IL containing solvents and some specific enzymes' structures and dynamics, but a comparative study to investigate IL effects across a family of enzymes has not yet, to the best of our knowledge, been conducted. To address the disparate understanding of IL-enzyme relationships at the molecular level, we performed a series of MD simulations on a set of family 5 glycoside hydrolases (GH5). Our simulations focus on the work of Datta and coworkers and seek to reproduce several of their experimental conditions at the molecular scale. Specifically, we vary the enzyme species, temperature, and IL concentration to see how each affects the structure and

dynamics of the protein. The three enzymes share a structural motif being as they are all from the GH5 family, but they have very different activity in IL-water mixtures. These simulations will help us understand which differences in the proteins' structure lead to the observed loss in activity.

## Methods

All simulations were performed using AMBER compatible force fields implemented in GROMACS 4.6 [198] using the TIP3P water model [199]. The enzyme forces were calculated with AMBER99SB-ILDN force field [200], and the ionic liquid force field was adopted from our previous studies [18] and utilizes the general AMBER force field [150] and restrained electrostatic potential fitting for point charges [151] from quantum calculations using Gaussian09 [25] at the HF/6-31g(d) level of theory. We scaled the charges by 0.8 to more accurately represent dynamic properties of the solvent, as has been demonstrated in the literature [201]. Cubic boxes of the solvent and protein were generated using Packmol [202]. Systems were minimized using a steepest descent method for 10,000 steps or until machine precision was reached and then heated over 100 ps.

For our MD simulations, a time step of 2 fs was attained by constraining the bonds between hydrogens and heavy atoms using the LINCS algorithm [203]. Periodic boundary conditions were applied to all dimensions. Lennard-Jones interactions were calculated over 1.0 nm and shifted to ensure no artifacts were created from discontinuities in energy. Particle mesh Ewald summation is used to account for the long-range electrostatic interactions. Temperature was maintained using the Parrinello stochastic velocity rescaling algorithm [204], and pressure was held at 1 bar using a Berendsen barostat [205].

We have conducted 54 total simulations: 3 enzymes at 2 temperatures in 3 different concentrations of IL containing solvent in triplicate, for a total of 27 microseconds of simulation time. A list of all of the simulations performed is provided in Table S3.1 in the supplemental information. The IL chosen for these simulations was [EMIM][OAc]. Other GHs have been shown to tolerate this IL [128, 187, 206]. We chose three different concentrations of IL that corresponded with those found by Datta et al. [188] to cause large changes in enzyme activity (0%, 15%, and 50%). Weight percent was used in our simulations whereas volume percent was used in the experiments. This is not a perfect comparison, but the approximation can be made because the densities of the two solvents are similar and the excess volume of mixing is small [207]. In this case weight percent will approximately equal volume percent. These three concentrations are important to illustrate the differences in the enzymes. First, 0% was chosen as a standard. This allows us to see if the experimental crystal structures are stable using our simulation techniques. Without this standard, we might erroneously explain destabilization of the enzyme observed in MD as being caused by the presence of IL when it might actually be caused by a less stable experimental starting structure. Second, 15% was chosen because it differentiates among the three enzymes; in 15% [EMIM][OAc] VIR loses all activity in experiments, TMA loses about half of its activity, and PHO maintains nearly all of its activity. Finally 50% was chosen because all three enzymes lose nearly all of their activity at this concentration. The two temperatures (310 K, 353 K) were chosen because the former is the optimum temperature for VIR and the latter is the optimum for TMA and near the optimum for PHO. Triplicate simulations were launched from a single minimized structure for each enzyme.

The experimental structures were taken from the protein databank [208]. VIR was modeled using an x-ray structure from *Trichoderma reesei* (PDB: 3QR3) [209]. GLU54 was mutated to

ASP to match a known sequence for *T. viride* endoglucanase [210]. TMA and PHO were modeled using x-ray structures (PDB: 3AMD, 3AXX) [211, 212] without mutation. PROPKA [213, 214] was used to assign protonation states to the enzyme residues at the pH that optimizes cellulase activity: 4.5, 4.8, and 6.4 for VIR, TMA, and PHO respectively. Table S3.2 contains a list of the charged residues for each system. The surface charge of each enzyme noted in Figure 3.1B was calculated by first enumerating all residues with solvent-accessible surface area (SASA) greater than at least 30% of the maximum possible SASA for that residue type [162, 215], then counting the number of positive and negatively charged surface residues at the pHs mentioned above. RMSD, RMSF, solvent accessible surface area (SASA), secondary structure analysis, energetic and side chain data were all obtained from the toolkit included with GROMACS 4.6.

## Results and Discussion

*General Effects of IL on Enzyme Structure and Dynamics*. In this section we first analyze general trends in the structure and dynamics of the three cellulases. This serves as a baseline for comparison among the three enzymes. Afterwards, we discuss each enzyme individually and identify important observations unique to each system.

*Structural Evolution*. The structural evolution of the systems over the 500 ns trajectories is an indicator of the stability of the enzymes in mixtures of IL and water. To provide some insights into the extent of the structural changes we observed, snapshots for the three enzymes after 500 ns of simulation in the three solvent states (0%, 15%, 50% IL) are compared to its crystallographic structures in Figure 3.2.

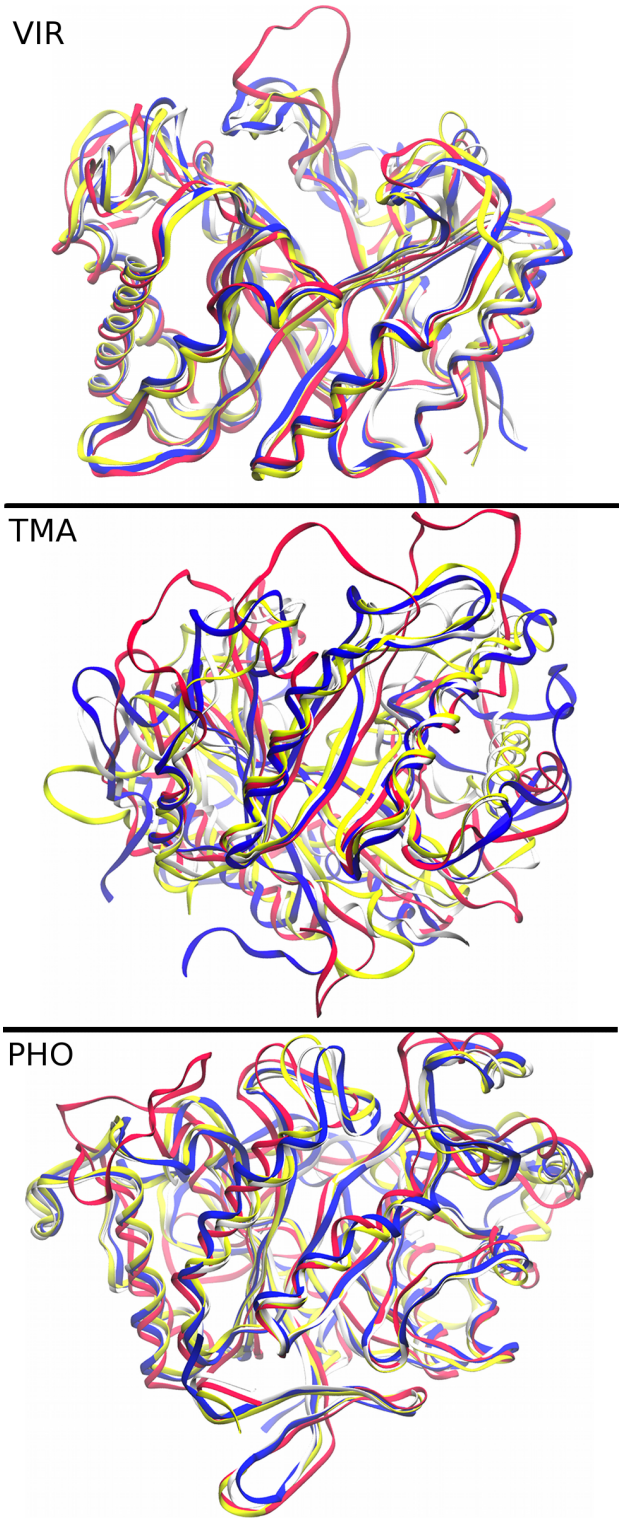



**Figure 3.2:** Structure of each enzyme after 500 ns of simulation compared to the crystallographic structure after aligning all Cα atoms. Crystallographic – white. Water – yellow. 15% IL – Red. 50% IL – Blue.

The ends of the simulations rather than average structures are used because the highest RMSD tends to occur near the end of the simulation. This accentuates the differences in structures to help distinguish among them. The structures are taken from simulations at 310 K for VIR and 353 K for TMA and PHO. The structures of VIR and PHO remain close to their respective crystallographic structures. Only a few largely unstructured loops (PHE 15 to ASP43 for VIR and LYS111 to ASP131 for PHO) are greatly displaced in the presence of ionic liquid. TMA, on the other hand, deviates noticeably from its crystallographic structure in several regions. Several alpha helices are less structured in the IL containing systems, indicating that at least a part of the TMA structure is unfolding in the presence of ionic liquids. Much of the deviation from crystal structure is due to the unfolding of the alpha helix at the N-terminal end. This event precedes the destabilization and unfolding of nearby structures later in the trajectory.

Deviations of the enzymes' structures from their initial configurations were analyzed over their complete trajectories using Cα RMSD. In contrast to our earlier studies in ILs, several of the simulations, specifically of the enzymes VIR and TMA, display major deviations from the starting structure over long time scales, possibly indicating the start of denaturing. The enzymes that we have previously studied with molecular simulation have very different secondary structures than these family 5 cellulases[18, 20]. We find that alpha helices tend to be less stable than beta sheets in the presence of ionic liquids. Figure 3.3 contains plots of alpha carbon RMSD for each protein in different concentrations of IL and key statistics summarizing the RMSD data. The data presented in this figure are for the optimum temperature of 310 K for VIR and 353 K for TMA and PHO. The N-terminal (first 15 residues) of TMA were not included in the RMSD calculations (Figures S3.1 and S3.2 contain all of the data for all carbons at both temperatures). This is because this region has a very high RMSD, and it might lead to the incorrect conclusion

that TMA is unfolding in even the water simulations. For example, the large changes we observe in RMSD are sometimes characterized by large structural changes over relatively short time periods, occurring after significant elapsed simulation time (hundreds of nanoseconds) for multiple trajectories. That is to say the RMSD remains low for an extended period at the beginning of the simulation, suggesting that the protein is stable. At some point later in the trajectory, a portion of the protein adopts a different structure relatively quickly. Specific examples of this behavior can be seen in Figure 3.3 for TMA in water at 100 and 250 ns, for TMA in 50% IL at 200 and 400 ns, and for PHO in 15% IL at 250 ns. It should be noted that these RMSD data should be compared back to the water standard in each case to determine if there is unfolding of the proteins' structures. Different crystallization techniques will lead to different RMSD values. Based on the observations of the MD trajectories it is clear that these events are indicative of least partial protein unfolding for TMA. Without dramatically longer simulation times it is impossible to determine we are observing a change in the preferred conformation of the enzyme in the ILs (i.e., solvent-induced unfolding) or high amplitude, long-wavelength structural fluctuations (i.e., solvent-induced changes in the conformational dynamics).

The RMSD statistics for PHO, the most IL tolerant of these enzymes, show similar behavior among all three of the concentrations except for those systems in 15% IL at 353 K. This difference is caused by a transition of a surface loop (LYS111 to APS131) that rearranges to make contact with a different portion of the surface in one of the three replicate simulations late in the trajectory as (seen in Figure S3.2). This replica was chosen for Figure 3.2 to show that the maximum extent of structural disruption is limited mostly to one unstructured loop. Since the transition is late in the trajectory, the average remains near that of the other replicas. Taken

together, these data show that the enzyme PHO maintains behavior similar what it exhibits in water for both of the tested IL concentrations in terms of its departure from the native structure.

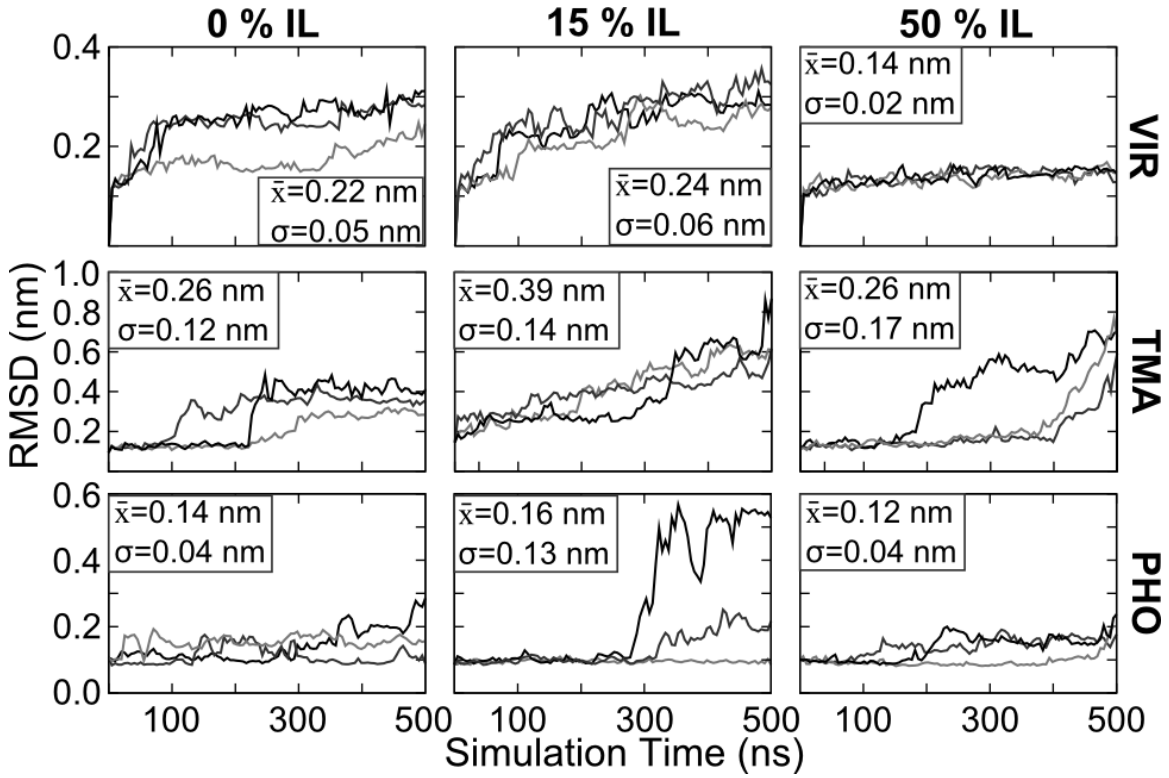



**Figure 3.3:** RMSD of alpha carbons. Three replicas. VIR at 310 K. TMA and PHO at 353 K. First 15 alpha carbons excluded for TMA. Average and standard deviation of all RMSD values shown.

The two less IL tolerant enzymes, TMA and VIR, exhibit increased deviation from crystal structure and more varied behavior in IL mixtures. Generally, moderate deviations are seen at 310 K whereas the largest departure from the native states are observed at 353 K. This is surprising since TMA was expected to be quite stable (melting temperature of 380 K) in water. One explanation for the deviation from the crystal structure for TMA simulations in water is the large difference in pH between the simulation conditions and the experimental crystallization conditions (4.8 versus 8.5). VIR comes from a mesophilic organism, and it is expected to deactivate at the higher temperature – possibly suggesting that the native structure is unstable at

elevated temperatures. On the other hand, TMA comes from a thermophilic organism and its highest activity in water was reported by Datta and coworkers at 353 K. Viewing the trajectories reveals that the high RMSD of TMA can be explained by fluctuations in the N-terminal region of the protein and by the loss of secondary structure in a few key regions, details for which are provided below. For PHO and VIR, systems solvated in 50% IL exhibit more stability, even compared to the simulations of the enzymes in pure water. From our previous experience with IL-solvated proteins, this is expected. On time scales of up to 500 ns, proteins in 50%+ IL tend to have less or equal deviation from crystal structure when compared to water solvated systems. Furthermore, in near-pure ILs we have observed that enzymes exhibit severely diminished structural fluctuations and dynamics on the 150 ns time scale [20]. This may be a consequence of the higher viscosity and lower sampling times, but experimental evidence shows that some proteins are actually stabilized over long incubation times in ionic liquids [216, 217]. TMA is the exception to these previous observations. TMA in 50% IL appears to be unfolding on these timescales, but that unfolding is slower than in the 15% IL system. The IL-containing TMA systems appear to still be in the process of unfolding at the end of the trajectories as will be shown in later discussion.

The character of the RMSD plots differs from VIR to TMA. The RMSD of VIR deviates linearly with time whereas the RMSD of TMA increases more often with fast transitions (cf. Figures S3.1 and S3.2). Visualizing the trajectories shows that these events are preceded by the breaking of key salt bridges within the protein, formation of non-native salt bridges, and exposure of hydrophobic residues to the solvent. We hypothesize that the IL mediates this unfolding by stabilizing the exposure of charged sidechains that begin internal to the protein and later rearrange to form salt bridges that are nearer the exterior of the protein. Figure S3.3 shows

that higher concentrations of IL correlate with higher numbers of non-native salt bridges. The statistical entropy of TMA's sidechains and the internal energy of the enzyme (discussed later) are consistent with this mechanistic interpretation.

*RMSF*. To analyze changes in the characteristic fluctuations of the protein backbone, Cα root-mean-square fluctuations (RMSF) of the enzymes are calculated for the last half of each trajectory and then averaged over the three replicate trajectories. The results for these calculations are shown in Figure 3.4. The observations that follow hold whether individual trajectories or the aggregated/averaged data is used. The RMSF of PHO is similar across all concentrations of IL. The presence of IL affects only one region of the enzyme in a substantial way. This is due to the previously mentioned event observed in one of the three high temperature-15% IL simulations in which an unstructured loop (LYS111 to APS131) changes contact from one section of the bulk of the protein to another. Since enzyme function can be related to dynamic motions during the binding and release of substrates, and since the catalytic function of the enzyme depends highly upon the precise arrangement of the catalytic region, we surmise that enzymes that maintain water-like character in IL-containing solvents will also maintain significant activity in those same solvents. Therefore, the retention of water-like character in increasing concentrations of IL for PHO may be indicative of its high tolerance to ILs as observed by Data et al [188].

TMA and VIR, in comparison to PHO, exhibit diverse features in their RMSF at key segments in the enzyme sequence. For TMA the presence of IL affects the magnitude of the fluctuations much more than it affects the location of the fluctuations. The N-terminus shows remarkably high fluctuations in all of the simulations. Upon inspection of the trajectories, we observe that the helix nearest the N-terminus comes undocked from the protein bulk and loses

secondary structure in 15% IL. This unfolding precedes unraveling of another longer helix (MET69 to ARG86) and subsequent destabilization of an unstructured loop (ANS20 to LYS35) near the entrance to the enzyme's binding site, both of which are connected closely linearly in primary sequence and in geometric proximity to the N-terminus. Similar behavior is also observed in the 50% IL simulations.

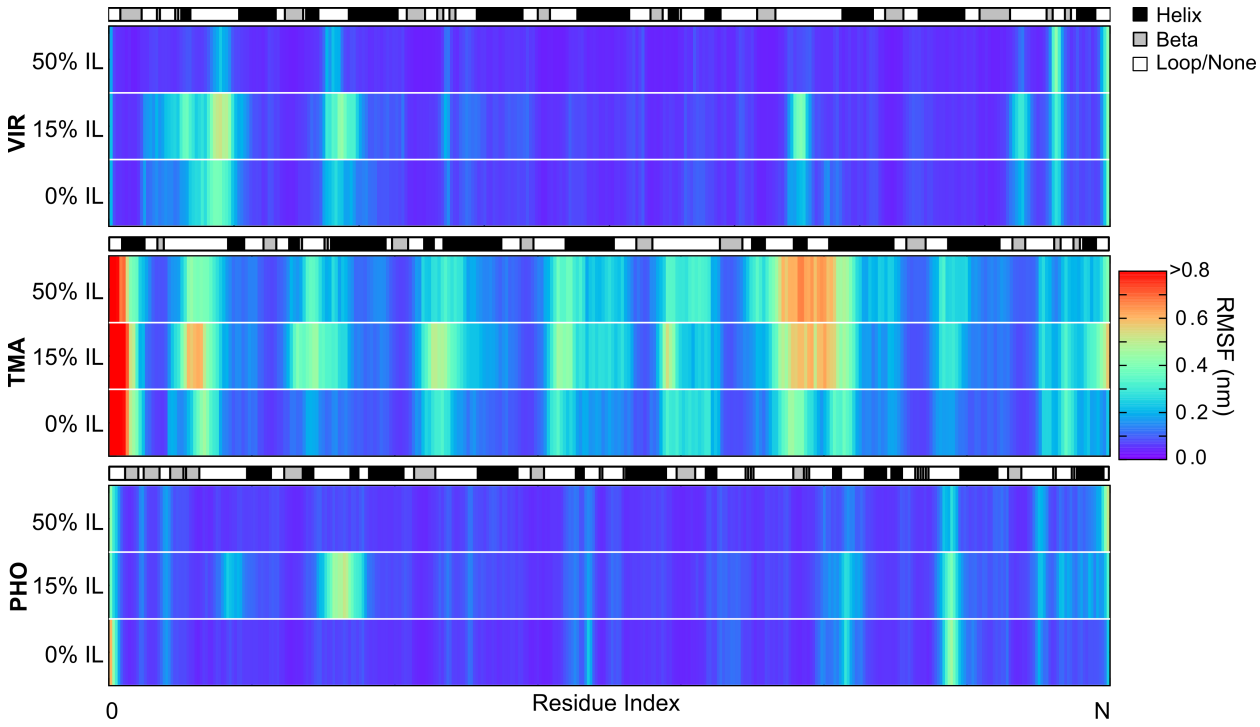



**Figure 3.4:** Root-mean-square fluctuation (RMSF) averaged over the triplicate simulations at 310 K for VIR and 353 K for TMA and PHO. N = number of amino acids in each enzyme. Secondary structure from DSSP in grayscale.

The magnitude of the fluctuations in VIR is lower than that of TMA. This is because the data presented for VIR are at 310 K whereas those for the other two enzymes are at 353 K. The data are presented in this way because a temperature of 353 K is unnatural for VIR, and the data at 310 K better represent experimental conditions. The width of the regions of high fluctuation tend to increase from 0% to 15% IL and then decrease to relatively thin slivers as the concentration increases to 50% IL. The decrease in fluctuations in 50% IL indicate that the

enzyme may be more structurally stable in these concentrations of IL at these temperatures and timescales. Inspection of the trajectories reveals that the highest fluctuations come from regions in which there are multiple salt bridges that intermittently break and reform contact. RMSD and RMSF are useful gauges of the global average behavior of the three enzymes, but these measures can obscure important details of the trajectories. Therefore, below we analyze each enzyme individually to discuss unique aspects of the interactions between each IL/enzyme combination.

*Surface Charges and Salt Bridges*. The arrangement and ratio of charged surface residues in chymotrypsin and lipase have been shown experimentally to affect the enzymes' activity in binary mixtures of IL and water [218, 219]. Interactions of the cations with the surface residues are hypothesized to play a role in IL tolerance by creating a shell of positive charge around the protein that cannot be penetrated by anions, which tend to disrupt protein structure more readily. Of the three enzymes studied, TMA is the only one with a positively charged surface as predicted by the previously described SASA method. If the hypothesis about the protective properties of a layer of cations at the enzyme's surface is true, TMA should be the most prone to attack by anions, and secondary structure should be lost more readily. We measured the average number of acetate carbonyl carbons within 0.4 nm of the protein for the natural-temperature trajectories. We then divided this count by the hydrophilic SASA of each enzyme, which averaged approximately 110 nm2 in all cases. The number of anions was 0.28, 0.32, and 0.23 anions/nm2 in for VIR, TMA, and PHO respectively. This supports the hypothesis of more positively charged protein surfaces attracting more anions. Figure 3.5 illustrates the coulombic potential at the surface of each protein for its experimental X-ray structure protonated according to predictions by PROPKA. These representations demonstrate the differences in the electrostatic character of the proteins, which greatly affects the structuring of ILs around the enzymes. The

number of salt bridges was enumerated to see the effect of the presence of cations and anions. Generally, higher concentrations of IL lead to higher numbers of salt bridges. This formation of additional, non-native salt bridges is especially prevalent for TMA in 50% IL, but it can be noted in all IL systems. The evolution of salt bridges over time is visualized in Figure S3.3.

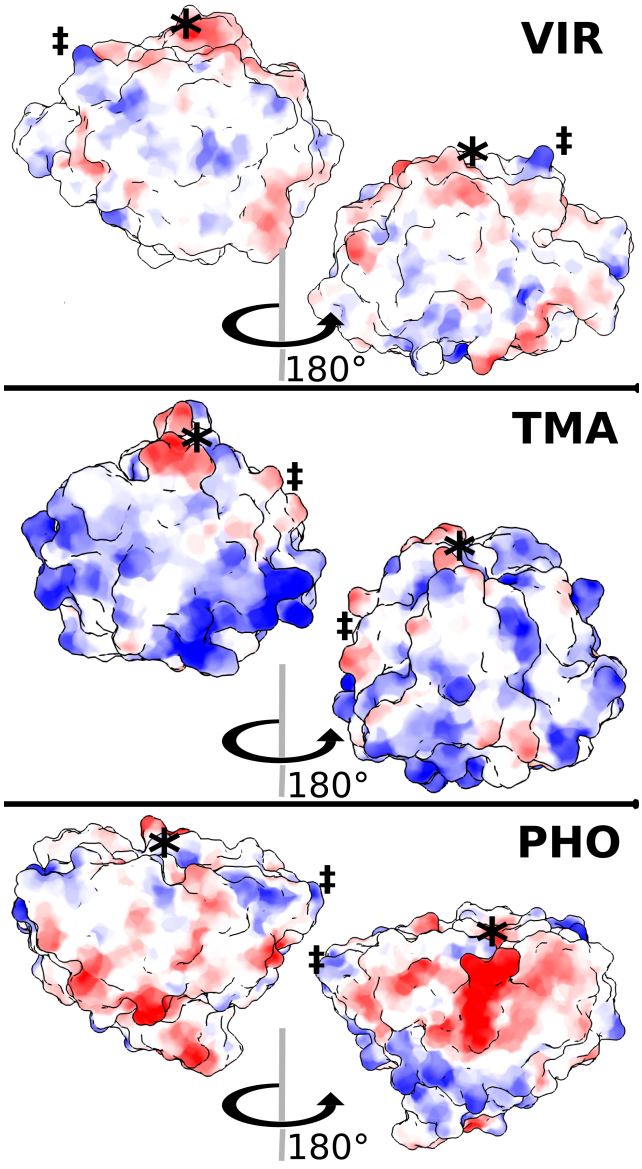



**Figure 3.5:** Coulombic potential surfaces for the crystallographic structures protonated by PROPKA. Red is positive. Blue is negative. Color intensity scales with potential. The * and ‡ are placed to guide the eye through rotation.

*Energetic and Entropic Contributions to Cellulase/IL Interactions*. To gain insight into the differences in the thermodynamics of the enzymes in water versus ILs, we calculated the internal energy of the protein and the statistical entropy of its sidechains in each solvent. Statistical entropy of the $\chi_1$ dihedral angle was calculated for each residue in each solvent using the definition $S = R \Sigma Pln(P)$ where $R$ is the ideal gas constant and $P$ is the probability of the dihedral angle [18, 196]. Internal energy was calculated as the potential energy of all molecular interactions within the protein. Results of the calculations at their most optimum temperatures are presented in Figures S3.4 and S3.5. In general, the greatest differences in $\chi_1$ histograms between IL and water systems are observed in histidine, cysteine, and protonated aspartic acid residues. Interestingly, the presence of ILs has little effect on the positively charged lysine and arginine histograms, which are often used as targets in stabilization studies [218, 219].

The effects of the solvent on the protein sidechain entropy and internal energy differ among the three enzymes. Entropy is generally reduced in the presence of ILs for VIR, indicating a significantly reduced conformational flexibility in the sidechains. In 50% IL the internal energy of the protein averaged over the last half of the trajectory is lower than that of the water system. This is consistent with structural changes that increase protein-protein interactions in order to reduce contact with the solvent; the RMSD of VIR in 50% IL and observation of the trajectories reveal that the protein does not exhibit any large structural changes. On the other hand, sidechain entropy is generally increased in the presence of ILs for TMA. The internal energy of the protein averaged over the last half of the trajectory is higher in IL than it is for water. These calculations indicate that the sidechains are exploring more phase space, and that there are fewer favorable interactions within the enzyme structure as breaking an intramolecular protein contact will increase the internal energy in accordance with the Lennard-Jones pairwise

functions used in this study. The entropy increased especially in regions near helices where the majority of unfolding takes place. PHO has similar sidechain entropy in water and IL. The internal energy of the protein over the last half of the trajectories is slightly higher in ILs than it is in water. In the case of 15% IL much of the change in internal energy can be accounted by the transition of the unstructured loop (LYS111 to APS131) described before. In the case of 50% IL, the increase in internal energy is much smaller than that of TMA, and the internal energy of the enzyme is close to that of water.

*Enzyme Specific IL Induced Changes*. VIR was deactivated significantly by even the lowest experimentally tested concentration of IL [188]. However, large-scale denaturation of the enzyme on the 500 ns timescale was not apparent in the RMSDs and RMSFs calculated from our simulations. Subsequently, we further investigated the trajectories for evidence of any persistent changes in the enzyme structure or other features when comparing the water simulations and the 15% IL simulations. The trajectories show specific changes in the binding pocket, which we propose may lead to deactivation (whereas the native structure is mostly maintained) even at low concentrations of IL. This analysis reveals that several events take place in the substrate-binding pocket of each of the 15% IL systems for VIR. Initially, ARG109 and ASP152 establish a salt bridge. ASP20 is nearby this salt bridge, but it does not participate. Within the first 200 ns of each 15% IL trajectory, ASP20 along with the nearby charged residue LYS27 move farther away from the bulk of the protein into more solvated states. This helps expose the loop that contains ASP20. Interestingly, the hydrophobic VAL24 follows its two neighbors and becomes exposed to the surrounding solvent. Late in each of these trajectories (after 250 ns), ARG109 breaks contact with ASP152 and forms a new salt bridge with ASP20. Altogether, these rearrangements significantly reshape the binding cleft of the enzyme and could lead to the observed deactivation.

The aforementioned changes are highlighted in Figure 3.4. Some salt bridges break and reform in the water solvated trajectories, but the reproducible and significant rearrangement of contacts observed in 15% IL is not observed in water.

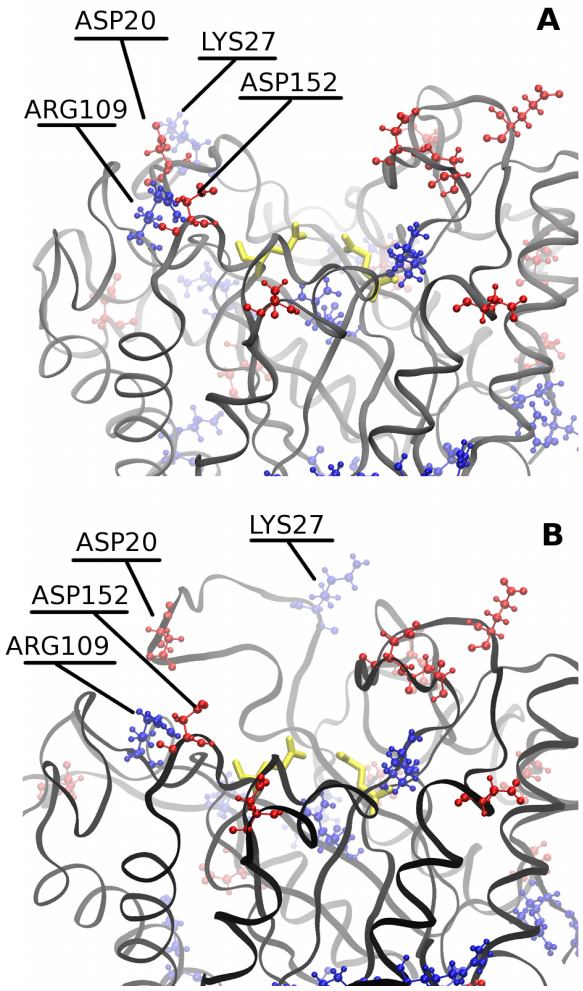



**Figure 3.6:** A view of VIR's binding cleft in 15% IL. (A) Structure after minimization. (B) Structure near the midpoint of the trajectory. Positive residues in blue. Negative residues in red. Catalytic residue heavy atoms in yellow.

Conservation of the binding pocket's structure is key to maintenance of activity in two ways. First, as a rule for biocatalysis, the alignment of the substrate near the catalytic residues can significantly affect the energy of the transition state, and thus the rate of catalysis. Second, the rate of binding and unbinding affects the turnover rate of the enzyme through its effect on the maximum rate of reaction ($V_{max}$). Figure 3.6 shows disruption of the shape of the binding pocket.

In its natural state, the substrate would be directed into the page in this image. Therefore the more occluded binding pocket seen in IL systems is less accessible to natural binding of the substrate. To be clear, there are cases where a reshaping of the binding pocket can lead to higher enzyme turnover, but because of the shape of the substrate (long oligomers of carbohydrates), we suspect that steric hindrances will cause binding to be less energetically favorable in IL-containing solvent than in water. It is also possible that the substrate would have a lower rate of unbinding in IL mixtures, but we observe nothing in our simulations that would lead us to this conclusion. The relative energy of binding and unbinding of the substrate in ILs versus water could be tested by isothermal titration calorimetry, but this will not differentiate between the possibility of inhibition by the IL and reshaping of the binding pocket. Kinetic studies in very low concentrations of IL to measure inhibition constants would give even more information.

*TMA*. Compared to VIR, TMA was deactivated less severely at low concentrations of IL [188]. Complete deactivation was observed by 20% v/v IL in water. TMA is different from the other two enzymes in that it has a net positive surface charge at both neutral pH and pH 4.8, the optimum pH for enzyme activity as observed by Datta et al. This is not characteristic of halophilic enzymes. It has been hypothesized that a negative surface charge attracts cations, thus increasing protein hydration while repelling anions that disrupt salt native bridges and hydrogen bonding [187, 218]. In the case of imidazolium based ILs, the cation tends to be larger than the anion, and the charge is more dispersed on the molecule. In our system specifically, the positive charge of the cation is dispersed over the several atoms of the imidazole ring, whereas the negative charge of the anion is concentrated on the two carboxyl oxygen atoms. Following the same hypothesis, a protein with a negative surface charge will attract these large cations and be shielded from attack by the relatively small anions with high charge densities [218].

Compared to the other enzymes, TMA experiences much higher loss of secondary structure. Figure 3.7 compares the alpha helical content, as determined by DSSP [220], over time for each enzyme across the range of IL concentrations at the temperature nearest each enzyme's optimum. It can be seen that the alpha helical content of VIR and PHO remain stable throughout the simulations. On the other hand, TMA loses a majority of its alpha helical character in IL containing solvents, indicating an underlying reason for the large RMSD and RMSF values discussed above. It is interesting to note the highly positive surface charge shown in Figure 3.5 for TMA. In water, a highly positive surface charge appears to not be particularly detrimental to the stability of the secondary structure of the enzyme, but when anions are introduced from the IL the enzyme denatures much more dramatically. A pH of 4.8 is optimal for activity in water, but this enzyme may need a higher pH to deprotonate the surface so that when ILs are introduced, anions do not attack penetrate the enzyme so readily. Similar DSSP calculations were also completed for the enzymes' beta sheets and showed that this loss of secondary structure is specific to helices. Beta sheets were highly stable across all enzymes and concentrations tested. As seen in Figure 3.1, the helices may shield the beta sheets from the solvent because they form the outer surface of the protein surrounding the inner "beta barrel". Although it is difficult to perform circular dichroism in imidazolium based ILs because of the absorption spectrum of the cation and because of the sensitivity of the technique to even small concentrations of contaminants, we predict that IL mediated loss in secondary structure could be observed for TMA even in concentrations as low as 15% v/v. FTIR might also be suited to detect loss in secondary structure.

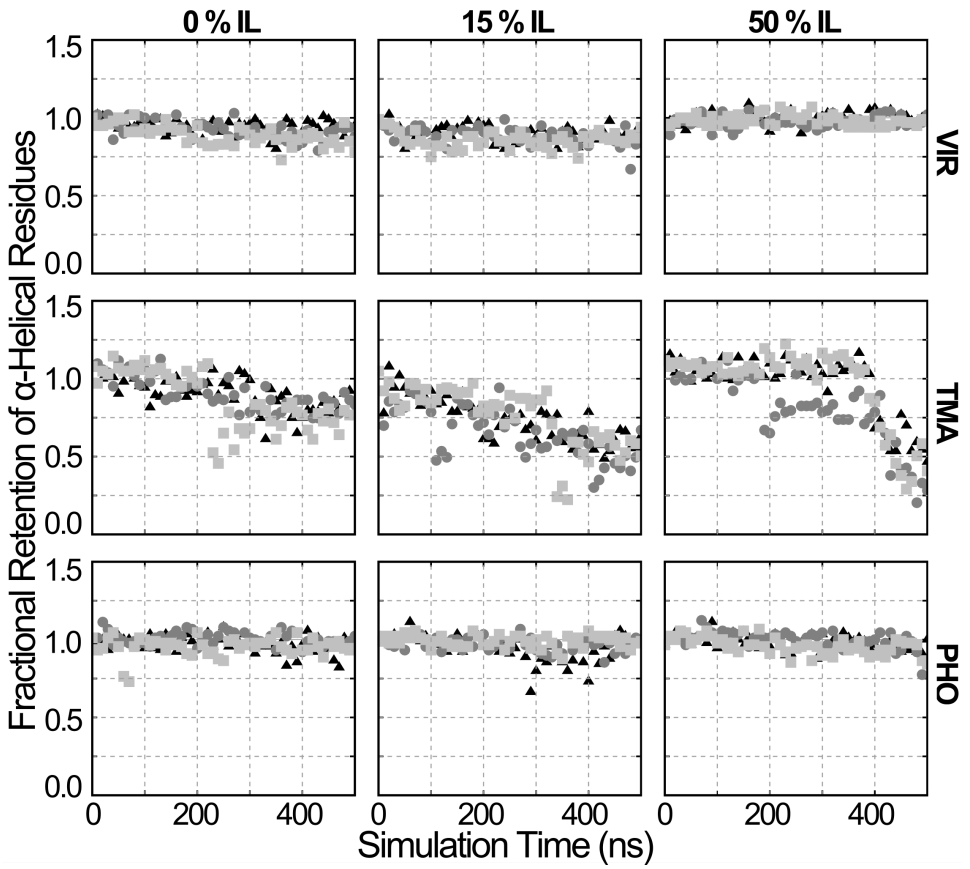



**Figure 3.7:** Retention of alpha helical structure. Three different replicate simulations are shown. Secondary structure contents calculated using DSSP and normalized to the crystallographic secondary structure.

*PHO*. PHO is surprisingly stable at all simulated concentrations of IL and both temperatures. This is consistent with the experimental observation that PHO loses relatively little activity at 20% v/v IL in water, but it is inconsistent with the experimental observation that PHO loses nearly all activity at 50% v/v IL. From the previously discussed figures, it can be seen that the water-like behavior, as measured by RMSD and RMSF and observation in the trajectories, of PHO is maintained in even in 50% w/w IL. Since the structure of the catalytic site affects catalytic activity and since dynamic motions of the enzyme affect catalytic turnover rates, we propose that the maintenance of enzyme in activity in ILs could be predicted using MD simulations to carefully screen for enzyme/IL combinations in which enzyme/water interactions

like structural fluctuations are preserved. However, note that there remains the untested possibility that our simulations of PHO in high concentrations of IL were simply not long enough to observe some of the changes we noted above in the TMA and VIR systems. Yet, there are several reasons why an enzyme might lose activity in IL even when the structure and dynamics are unchanged (as in this case). The structure of the substrate might change, or the more viscous medium might lead to lower turnover rates. Yet, changes in activity due to viscosity are expected to manifest themselves with a gradual decrease in activity with increasing IL concentration. This is not observed experimentally. The IL might act as a competitive inhibitor. Observation of our trajectories reveals that one of the catalytic residues (GLU342) preferentially interacts with single imidazolium cations, trapping them within a distance of 4 Å for over 100 ns. If the cation is an inhibitor, it is not a strong inhibitor in the experiments at concentrations of up to 20% v/v. Interestingly, there is little decrease in the activity between 0% and 20% v/v IL and a sharp decrease to no activity between 20% and 50% v/v IL. We know that this range of concentrations can be an important transition point for protein stability and activity. Previous experimental findings show that denaturation cytochrome c and human serum albumin begin to occur somewhere in the range between 25% to 50% v/v 1-butyl-3-methylimidazolium chloride [132]. For human serum albumin, the structural changes lead to aggregate formations. Our previous study of a GH11 xylanase shows that aggregation occurs at similar concentrations of [EMIM][OAc] and [EMIM] ethylsulfate [18]. In accordance with these previous studies, we speculate that aggregation may play a large role in the loss of activity for PHO in higher concentrations of [EMIM][OAc]. Another possibility is that the time-scales of our simulations are too short to see large structural changes for PHO. Unfortunately the computational methods employed in this study are inadequate to study aggregation as will be discussed below.

## Conclusions

*Other Considerations*. Datta and coworkers previously determined the unfolding temperature of each protein in different concentrations of IL. They observed a decrease in the unfolding temperature of the proteins as the concentration of IL increased [188]. However, unfolding of PHO in ILs may take place over a much longer timescale than is accessible by atomistic MD. Another possibility is that the lower melting temperature is due to new protein-protein interactions introduced with aggregation as discussed briefly before. Based on knowledge of how other enzymes behave in ILs, we propose that aggregation plays a role in the deactivation of PHO at high IL concentrations. Aggregation could occur as the consequence of the exposure of some portion of hydrophobic residues, or the protein could precipitate as a consequence ILs screening of the surface charges that normally keep proteins suspended as colloids [221]. Some studies mention that there are limits to the solubility of certain proteins in certain ILs [221, 222]. There are a number of papers on the positive effects of certain ILs on the long-term stability of proteins [223, 224]. These studies usually focus on low or high concentrations of IL in water, and they skip over the important intermediate region where denaturation and aggregation effects have been observed previously. Still, in certain neat ILs, aggregation of other enzymes like lipase are observed [159]. There are many factors that play a role in protein stability and aggregation in ILs including IL species, protein structure, and temperature. Unfortunately standard MD is not able to access timescales necessary to study aggregation. A coarser description of the protein and solvent and other methods of simulation beyond MD could help researchers access these events.

We believe that the simulations of lone proteins are still useful in determining the effects of high concentrations of IL on the protein structure and dynamics for the prediction of IL tolerance. We observe no unfolding in the simulations of PHO in 50% w/w IL at 353 K on the 500 ns

timescale. This indicates to us that our simulations can predict high halophilicity and IL tolerance for the enzyme, but the simulation would not in and of itself predict that this enzyme is not active at this temperature and concentration. Other simulation methods and experiments would need to be employed as predictors of aggregation, precipitation, and effects on substrate-protein interactions. For example, coarse-grained simulations that predict the osmotic second virial coefficient might help us determine if a specific protein will salt out of an IL solution[225]. The outlook for future IL-protein simulations is promising. With the development of more accurate coarse grained force fields such as the MARTINI polarizable force field and more efficient methods of enhanced sampling such as metadynamics, temperature accelerated molecular dynamics, and parallel tempering, we believe that future researchers will be able to use simulations to guide or compliment experiments in this field.

*Towards Improved Enzyme-IL Simulations*. We reiterate that in addition to the limitations noted above, there remain definite areas of future work in order to improve enzyme-IL simulations and their use in interpreting and inspiring new experimental work. While it is clear from the TMA simulations that large structural changes can occur after hundreds of nanoseconds, there is still essentially no way to predict *a priori* if and when these transitions may occur. Future work using enhanced sampling free-energy calculations could investigate folding free-energies for small model systems in an attempt to generalize some of these observations to larger systems. It is also clear that the structures are still evolving at the end of the trajectory, indicating significantly longer time scales are needed. Emerging hardware solutions such as accelerators or special purpose MD machines could help address this in the future. Ultimately, the observations and hypotheses presented herein are meant to compliment and guide experimental studies and add some molecular-level insight into IL-protein interactions.

*Causes of Enzyme Deactivation*. We analyzed the behavior of three distinct family 5 glycoside hydrolases in binary mixtures of IL and water using molecular dynamics. By drawing on a previously published study on these enzymes and comparing that study's results with our simulations, we developed the following suggested mechanisms could be the most important to explain the loss of activity of the enzymes in IL containing solvents, and we made predictions on how our hypotheses could be experimentally confirmed:

1. VIR is deactivated by deformation in the binding pocket that lead to the substrate to be unable to bind efficiently. We know from the published experimental data that the activity of this enzyme drops significantly even in very low concentrations of IL. In the absence of large structural changes in our simulations, smaller structural changes in key regions could therefore be a likely cause of the deactivation. This could be supported by isothermal titration calorimetry experiments to energetically describe binding/unbinding. Incorporating a kinetic study to determine if competitive inhibition plays a role in substrate binding would also be useful.

2. TMA is deactivated by extensive loss of secondary structure in IL. The large-scale denaturation of this enzyme's helices in the presence of ILs is obvious in our simulations. This concurs with the deactivation observed in experiments. X-ray scattering or small angle neutron scattering would be able to provide some detail about large changes in structure without being hindered by the unorthodox solvent. Fast Fourier infrared spectroscopy could also be used to monitor the evolution of the secondary structure. Increasing the pH of the solvent might protect the enzyme from attack by anions.

3. PHO shows no signs of deactivation in IL for our simulations. PHO exhibits water-like behavior in the presence of ILs, and there are no large-scale structural changes to structured portions of the enzyme or the binding pocket. We speculate that aggregation is a possible cause

of the deactivation observed at 50% IL, as we have observed experimentally for another glycoside hydrolase [18]. Unfortunately the methods employed in this paper do not cover the timescales necessary to sample protein-protein interactions, but we would be able to predict high thermal stability and halophilicity. Dynamic light scattering would be able to detect large aggregates of this protein and determine the IL concentration at which it salts out.

More generally, any of these suggested experimental techniques could be applied to any of these proteins. For example, observation of PHO trajectories shows an affinity of the cation for the catalytic residues, but there is almost no deactivation for low concentrations of IL experimentally indicating that inhibition is not a large problem. The experiments suggested for VIR could also be applied to PHO. At these timescales and temperatures we observe a few general trends that are consistent with our previous experience in IL-enzyme simulations. Higher concentrations of IL correlate with lower RMSD from experimental structures. High (>50%) concentrations of IL correlate with lower RMSF. It is surprising to see such a variety of character in enzymes that belong to the same family of glycoside hydrolases. The simulations described herein have the ability to predict the stability of enzymes in water and in IL, they can predict local changes in the structure that cause deactivation rather than relying on heuristics based on macroscopic properties, and they can be used as a tool to better select those enzymes or mutants that tolerate ILs.

# IV. Destabilization of Human Serum Albumin by Ionic Liquids

*Publication details:* This chapter contains the majority of a draft of the manuscript we plan to submit to *The Journal of Physical Chemistry B*. I have authored this manuscript under the supervision of Prof. Pfaendtner. This chapter contains a detailed explanation of metadynamics that will be truncated for publication.

## Abstract

Ionic liquid (IL) containing solvents can change the structure, dynamics, function, and stability of proteins. To determine the viability of molecular simulations as a tool for predicting IL-induced structural changes in a large multi-domain protein, we studied the interactions of human serum albumin (HSA) with two different ILs, 1-butyl-3-methylimidazolium tetrafluoroborate and choline dihydrogenphosphate. Root mean square deviation and fluctuation calculations indicated that high concentrations of ILs in mixtures with water led to protein structures that were kinetically trapped near the crystallographic starting structure. We hypothesize that this effect is due to the high viscosity of the solvent. To overcome these limitations, we employed enhanced sampling techniques that uncover the energy of experimentally determined important transitions within the protein structure. Metadynamics simulations show that the free energy landscape of the unfolding a loop 1 of domain I is different in the presence of ILs than it is in water, consistent with previously published experimental evidence. We then used essential dynamics coarse graining to systematically predict differences in the dynamics of proteins solvated in IL-water mixtures versus pure water systems. We demonstrate that this method can be used to predict other domains of the protein that might also change structure in the presence of ILs.

## Introduction

Ionic liquids (ILs) are a class of solvents that consist of salts that are liquid at temperatures below 100° C. Because of the unique ionic structure of these solvents, they tend to have a set of properties that make them useful in several applications. ILs tend to have low volatility, high thermal stability and low flammability compared to some of the most common organic solvents, leading some researchers to classify them as "greener" solvents.[131] ILs are useful for the solvation and separation of biomolecules related to biofuel synthesis. For example, lignocellulosic biomass processing can be enhanced by the addition of certain ILs such as [EMIM][OAc] which has the rare ability to solvate crystalline cellulose.[155] Algal biomass processing can be enhanced by using ILs to degrade the microbe's cell walls, releasing the important lipids within.[4] ILs have been shown to be useful in a number of applications related to chemical synthesis and catalysis. For example, IL-containing solvents can enhance reaction rates when combined with transition metal catalysts as well as change enantioselectivity or reaction equlibria when combined with enzymes such as lipases.[186, 217] All of the above properties make ILs an interesting solvent for enzymatic hydrolysis of biomass. In addition to characteristics that make ILs suitable solvents for chemical processes and catalysis, ILs can also be used in mixtures that stabilize proteins for long-term storage.[183, 226] This opens possibilities to extend the shelf life of therapeutic or industrially relevant proteins. Yet, there is a lack of fundamental understanding of what makes a favorable IL-protein combination whether it is for enzymatic hydrolysis of biomass or for long-term storage of proteins. Our research describes the interactions of ILs and proteins using molecular simulation to give us insight into the interactions that lead to structural and dynamical changes in the protein.

Of the systems that we have previously described, there is a wide range of IL-tolerance displayed by different enzymes. Stability of the enzymes depends on both the species of the IL and the specific enzyme used in the research. Several enzymes have been shown experimentally to tolerate the presence of ILs in concentrations of up to 20% v/v in water before deactivating; others have been shown to tolerate neat ILs and retain high levels of catalytic activity. We have studied several different enzymes including xylanase, lipase, chymotrypsin, and several cellulases.[18, 196, 227, 228] Specific interactions between cations, anions and the proteins of interest were analyzed to determine why certain ILs stabilize or destabilize these enzymes. Our analysis techniques have included root mean square deviation and fluctuation (RMSD and RMSF) of protein structure, principle component analysis (PCA), interaction lifetimes between ions and catalytic residues, surface residue entropy, structuring of water and ions around the protein, secondary structure evolution, and others. One problem that we have encountered frequently is that the proteins tend to become kinetically trapped near their crystallographic structures when they are solvated in very high concentrations (50% w/w or more) of IL mixed with water. Sometimes it takes hundreds of nanoseconds to get the protein to begin to unfold in these concentrations, even when it is known experimentally that the protein unfolds quite readily in ILs. We hypothesize that the high viscosity of the solvent leads the protein to become trapped near its crystallographic structure for longer time periods. Within this paper we describe two methods, metadynamics and essential dynamics coarse graining (EDCG), that we have used to more effectively sample the conformations of a large multi-domain protein in viscous media and to determine the differences in protein structure in IL-containing solvents versus water. Metadynamics adds a time-dependent potential bias to a collective variable of interest within the protein structure. The bias is added in such a way that the free energy landscape of the collective

variable can be uncovered.[229] EDCG is a method by which continuous domains of the protein are assigned based on a minimization of intradomain covariance and maximization of interdomain covariance.[230] A more formal discussion of these methods appears later.

We chose human serum albumin (HSA) as a test system for our methods. We became interested in HSA as a test system for two reasons. First, HSA has several well-defined subdomains that are assigned in literature. We can use this knowledge to compare the domains we find with EDCG to those that were previously published. Second, experimental evidence indicates that a specific loop in the protein denatures in the presence of 1-ethyl-3-methylimmidazolium tetrafluoroborate ([BMIM][$BF_4$]).[231] In previous studies, we have relied on the general observation that the activity of a certain enzyme decreases or that the enzyme aggregates in IL-containing solvents. Our simulations and analysis techniques had to consider many different possible modes of deactivation. In the case of HSA there is a very specific region (loop 1 of domain I) that we can probe in these simulations using metadynamics. Moreover, the same experimental study, upon which we are basing our simulations, indicates that coupling of HSA's domains' motions are affected by [BMIM][$BF_4$]. Covariance analysis, principle component analysis and EDCG can characterize this coupling in simulations. Another interesting study shows that choline dihydrogenphosphate stabilizes the structure HSA, wherein the protein binds fatty acids similarly in the presence of ILs and in crystallographic structures.[232] The imidazolium based IL, [BMIM][$BF_4$], on the other hand severely hampers the ability of HSA to bind fatty acids. We analyzed the differences in the structures of the binding domains over the length of the simulations to determine the extent to which MD can capture this behavior. Ultimately, the results presented herein indicate that our simulations can be used to (a) accurately replicate known IL-induced structural transitions in large proteins, (b) provide additional

molecular-level detail on interactions between ILs and proteins, and (c) suggest which regions of the protein have the largest structural changes when the solvent is varied. This is of particular interest for large proteins and highly viscous solvents that require long timescales to evolve.

## Methods

*Molecular Dynamics*. The molecular structure of HSA was taken from the protein databank (PDB: 4IW2, 1E7I).[233, 234] Protonation states of sidechains were determined using PROPKA3.0[213] at the biological pH of 7.4. The protonation states of these sidechains are specified in the supplemental material. All molecular simulations were performed in the GROMACS 4.5[235] simulation package patched with Plumed 2.0[236]. The AMBER99SB-ildn[200] force field was used to treat protein interactions, water was simulated using the TIP3P model, a combination of GAFF[237] and a published force field[238] was used to treat [BMIM][$BF_4$], and GAFF was used to treat [chol][dhp] and stearic acid. The electrostatic point charges of the ionic liquid, where derived and not taken from literature, were developed using the HF/6-31G(d) level of theory in Gaussian09[25] and the RESP method[151] in the antechamber force field development package. Lennard-Jones forces were shifted to zero beyond 1.2 nm. Electrostatics were treated using particle mesh Ewald summations. A time step of 2 fs was used in all simulations, and the LINCS algorithm was implemented to constrain the bonds between hydrogens and heavy atoms. Pressure was maintained at 1 bar with a Berendsen barostat[35], and the temperature was controlled with a stochastic velocity rescaling technique[34]. Temperature was varied across simulations to observe its effects on the coupling of domain motion for comparison to experimental studies. Molecular dynamics simulations began with 10,000 steps of steepest descent minimization followed by 5 ns of solvent equilibration at 500K in the NVT ensemble

wherein the protein's coordinates were frozen. The subsequent production simulations were of 200 ns in length.

*Metadynamics*. Metadynamics simulations were conducted using the above parameters with a smaller section of the protein (SER5 to PRO113) at a temperature of 310K. Pressure control was removed to sample the canonical ensemble – a requirement of metadynamics. Metadynamics is a method by which we bias simulations away from previously explored conformations so that a full range of structures and conformations can be explored and thereby provide estimates of the relevant free-energies in the system (as projected onto reduced dimensionality descriptors called collective variables – CVs). In contrast, normal MD is limited in the description of certain CVs because the only energy available to overcome barriers is thermal energy on the order of $k_BT$, thereby rendering the timescale to overcome large barriers by thermal fluctuations alone orders of magnitude longer than that accessible with modern computing power. Well-tempered metadynamics, the method that we will employ in this discussion, applies a history-dependent bias to the potential of some collective variable (CV) in the form of a Gaussian hill. The height of the bias hill decays over the simulation depending on how much bias has been applied to the current position of the CV earlier in the simulation. This helps the calculation smoothly converge to a free energy landscape for the CV(s) of interest. The bias potential accumulates to raise the system out of local free energy minima so that the CV of interest can be explored completely. By tracking the position and size of each hill added to the simulation, the free energy of the system as a function of some collective variable can be calculated. The following equations, taken from Barducci, Bussi and Parrinello's original derivation[239], will more thoroughly describe the application of well-tempered metadynamics.

First, we describe a histogram of the states explored by an MD simulation at equilibrium. The histogram should describe all possible states and not be confined in a local minimum. Equation 4.1 describes the histogram, $N$, that is a function of some collective variable of interest, $s$, and time, $t$. Equation 4.2 comes as a solution to Equation 4.1 given $N(s,0) = 0$. By definition of the Helmholtz free energy in the canonical ensemble, we can describe the free energy of the system, $F$, at some temperature, $T$, using Equation 4.3.

(4.1) $$N(s,t) = \int_0^t \delta_{s,s(t')} dt'$$

(4.2) $$\frac{\partial N(s,t)}{\partial t} = \delta_{s,s(t)}$$

(4.3) $$F(s) = -T \lim_{t\to\infty} \ln N(s,t)$$

We then apply the history-dependent bias potential, $V$. The form of the bias in Equation 4.4 contains some constants. There is a temperature, $\Delta T$, and an energy rate, $\omega$. The rate at which bias is applied is important. Low energy rates replicate MD better while high energy rates favor fast sampling. Since the potential decreases as the histogram grows for a given value of the CV, previously visited states are discouraged. Equation 4.5 is the rate at which the bias potential changes.

(4.4) $$V(s,t) = \Delta T \ln\left(1 + \frac{\omega N(s,t)}{\Delta T}\right)$$

(4.5) $$\frac{\partial V(s,t)}{\partial t} = \frac{\omega \Delta T \delta_{s,s(t)}}{\Delta T + \omega N(s,t)} = \omega e^{-\left[\frac{V(s,t)}{\Delta T}\right]} \delta_{s,s(t)}$$

Equation 4.5 can be compared to standard metadynamics. By replacing $\delta(s,s(t))$ with the equation for a Gaussian hill and recognizing ω as an initial hill height, we get standard metadynamics. Equation 4.6 therefore fully describes the potential applied in well-tempered metadynamics. In this equation, $\tau$ is the deposition rate of the hills and $\sigma$ is the width of the

Gaussian bias. We must also recognize that Equation 4.6 describes the case of a one-dimensional free energy landscape and that s is a function of the system's coordinates, $r$, at any given time. There are forms of these equations that account for multidimensional Gaussians as well.

(4.6) $$V(s(r),t) = \sum_{n=1}^{t'=n\tau<t} \omega\tau e^{-[V(s(r),t')/\Delta T]} e^{-\left[\frac{(s(r)-s(r(t')))^2}{2\sigma^2}\right]}$$

At very long times, $V$ will vary little. We assume that it is at equilibrium and that the whole system is at equilibrium at this point in time. In Equation 4.7 we can find the probability distribution from statistical mechanics. We can then combine Equation 4.7 with Equation 4.5 by substitution. We then integrate out *ds* to arrive at equation 4.8.

(4.7) $$P(s,t)ds \propto e^{\left[\frac{-F(s)-V(s,t)}{\mathrm{T}}\right]}ds$$

(4.8) $$\frac{\partial V(s,t)}{\partial t} = \omega e^{-[V(s,t)/\Delta T]} e^{-\left[\frac{F(s)+V(s,t)}{T}\right]} / \int e^{-\left[\frac{F(s)+V(s,t)}{T}\right]}ds$$

We can then imply Equation 4.9 for the case where $t$ is very large, and *c(t)* is an arbitrary constant.

(4.9) $$V(s,t\to\infty) = -\frac{\Delta T}{\Delta T+T}F(s) + c(t), \quad F(s) \cong -\frac{\Delta T+T}{\Delta T}V(s,t) + c(t)$$

We call the prefactor a "bias factor" and give it the symbol γ as in Equation 4.10.

(4.10) $$\gamma = \frac{\Delta T+T}{\Delta T}$$

The collective variable of interest in our simulations was the distance between the centers of mass of two groups of alpha carbons. The first group consisted of GLN32 to PRO35, and the second group consisted of LEU80 to GLU86. The initial height of the bias was 1.0 kJ/mol-nm, and it decayed with a bias factor of 20. The width of the hill was constant at 0.03 nm. Hills were added every 5.0 ps. A harmonic restraint of strength 500 kJ/mol was placed on the upper bound of the CV above 1.8 nm. An additional restraint was placed on the RMSD of the alpha carbons of

the ends of the segment to keep the unbiased portion of the structure near its crystallographic structure. The restraint was a simple harmonic with a prefactor of 500 kJ/mol for any RMSD above 0.14 nm.

*Essential Dynamics Coarse Graining*. Essential dynamics coarse graining (EDCG) is a method by which we define continuous domains of a protein that can act as coarse grain sites[230]. We intend to use this method to understand if the slow modes of motion of HSA are similar in water and in high concentrations of IL. Since RMSD and RMSF are not usually adequate to tell us about long timescale transitions in an IL-solvated protein's structure, we posit that EDCG will be able to predict sites that are most affected by the new solvent environment so that we can focus on those specific areas. By analyzing MD trajectories of the protein in water and in IL-water mixtures, we are able to construct a covariance matrix that quantifies the correlation of atoms' motions. The covariance of atom *i* and atom *j* in dimension *h* and *k* is defined in Equation 4.11, where *C* is the magnitude of covariance, n is the number of frames from the trajectory, *t* is time, and $\Delta r(t)$ is the difference between the current position of the atom and the average position of the atom over all frames of the trajectory. This leads to a matrix of values that is of the size 3N x 3N for a system of *N* particles. In practice, we use only the alpha carbons of the protein backbone.

(4.11) $$C(i_h, j_k) \equiv \frac{1}{n}\sum_{t=1}^{n} \Delta r_{i,h}(t)\Delta r_{j,k}(t)$$

We can then diagonalize the covariance matrix to get eigenvectors that are known as principle component analysis (PCA) modes. We order the modes by the magnitude of their eigenvalues to understand which modes contribute the most to long timescale structural transitions. This method allows us to ignore those motions that are fast and localized and do not contribute to large conformational changes.

Next, we define the number of sites that we want the CG model to have. After trying 6, 12, 18 and 24 sites, we settled on an 18-site model. This produced domains that ranged in size from several residues to tens of residues. Domains of this size have CV-descriptors that can be biased with metadynamics. The alpha carbons included in each CG site are chosen by an algorithm that minimizes, by steepest descent, the intrasite covariance while maximizing the intersite covariance. There are three constraints to the selection of sites. First, each atom is allowed to be in only one site. Second, the CG site is located at the center of mass of the alpha carbons in the group. Third, the atoms are assumed to be part of a continuous protein sequence.

## Results and Discussion

*Structural Evolution*. As a general measure of the stability of solute-free HSA in each solvent, the root mean square deviation (RMSD) of the alpha carbons from the crystallographic structure was measured over the length of the MD simulations. Additionally, the root mean square fluctuation (RMSF) of each alpha carbon from the average simulation structure was calculated. These data are presented in Figures 4.1 and 4.2 respectively for each concentration of solvent at 310 K. For RMSF graphs, we truncated the presentation at residue 530 because of some instability near the C-terminus that causes large fluctuations. The RMSD data demonstrate that at these timescales, HSA remains closer to its crystallographic structure in high concentrations of IL than it does in water. This is a symptom of the limited timescales we can explore with MD. HSA is likely kinetically trapped near its crystallographic structure. The RMSDs of these IL systems demonstrate why we need enhanced sampling techniques to gain insight into structural transitions in these highly viscous media. As for the systems containing 20% IL, we observe that RMSD is similar to that of water. On the timescale of 200 ns, there is no significant denaturation of HSA observed in either water or IL.

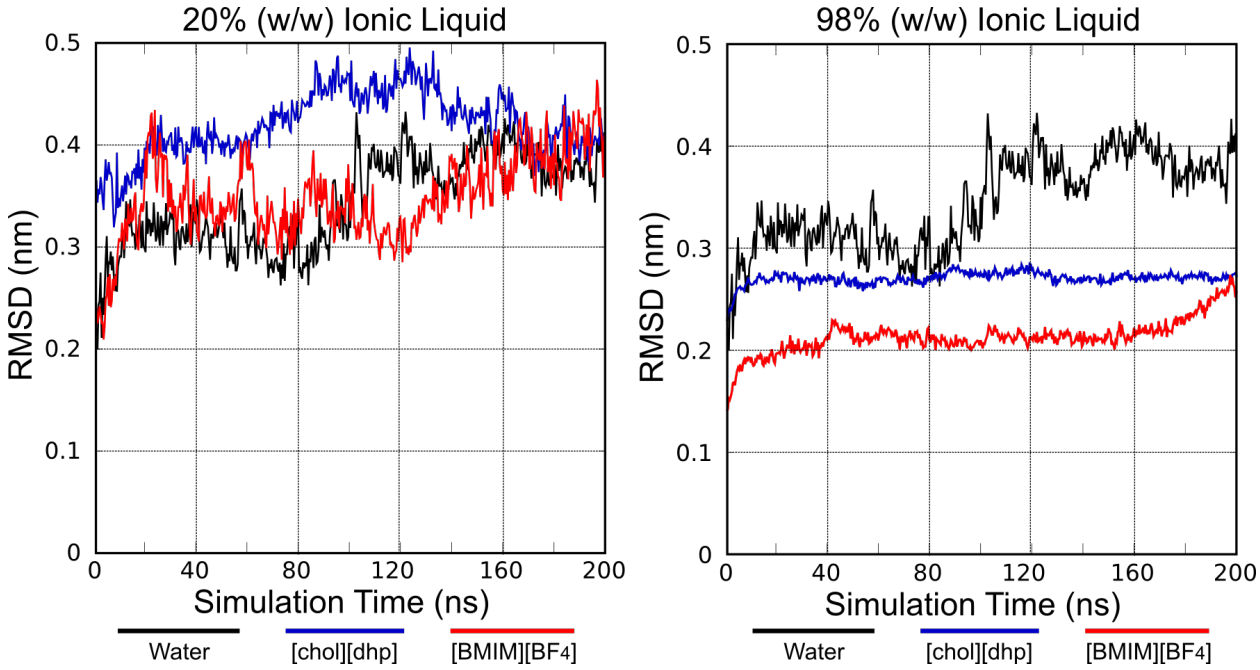



**Figure 4.1:** Alpha carbon RMSD of solute-free HSA at 310 K in various solvents versus crystallographic structure. Water in black for reference.

RMSF data can help us to pinpoint the regions of the protein that are most effected by the presence of IL. We observe that the fluctuations of the alpha carbons are much lower in magnitude in the viscous 98% IL systems than in the water system. This is consistent with our previous simulations of protein-IL systems. We observe higher magnitude fluctuations in domain I relative to domain III for [BMIM][$BF_4$] and the opposite for the water system. This is consistent with the experimental observation of the instability of domain I in this particular IL. Interestingly, the fluctuations in the protein are extremely small for 98% [chol][dhp] and similar to that of water for nearly every portion of the protein. This is consistent with the experimental observations that [chol][dhp] is an IL that tends to stabilize the protein structure especially for long-term storage.

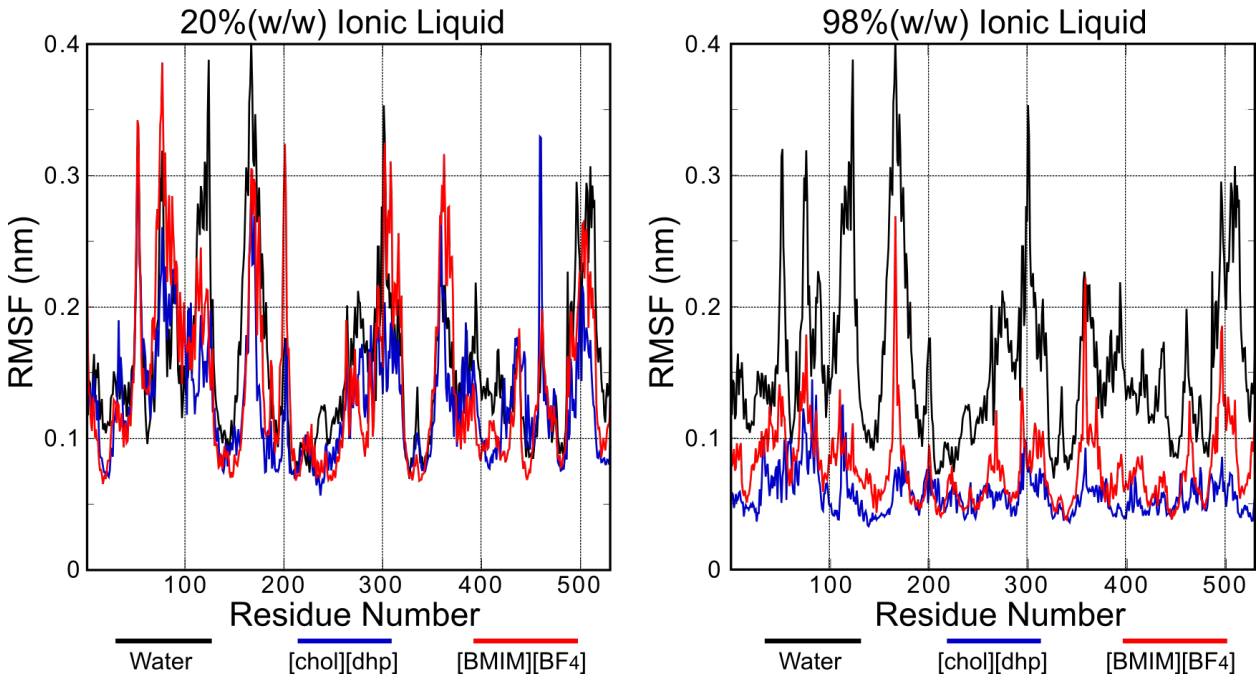



**Figure 4.2:** Alpha carbon RMSF of solute-free HSA at 310 K in various solvents versus crystallographic structure. Water in black for reference.

*Fatty Acid Binding Sites*. HSA transports fatty acids through the body. As such, there are several binding sites on the protein where stearic acid can bind. These sites are known from the crystallographic structure (pdb: 1E7I). By spin labeling stearic acid, experimentalists can determine the distribution of distances between fatty acids in HSA solvated in different environments. From our solute-bound simulations, we can calculate a similar distribution. Figure 4.3 shows the distance distributions of the 5th and 16th carbon of stearic acid bound to HSA in various solvents over the last half of each simulation trajectory. We calculated the distribution for the three most strongly bound stearic acid molecules as described in literature.[232] This is to make our comparison more comparable to the experimental conditions in which there was a 2:1 molar ratio of stearic acid to HSA. This encourages HSA to not become completely saturated with FAs. Thus, there are usually only a few sites occupied by stearic acid, and at equilibrium those sites would tend to be the ones that have the highest binding affinity.

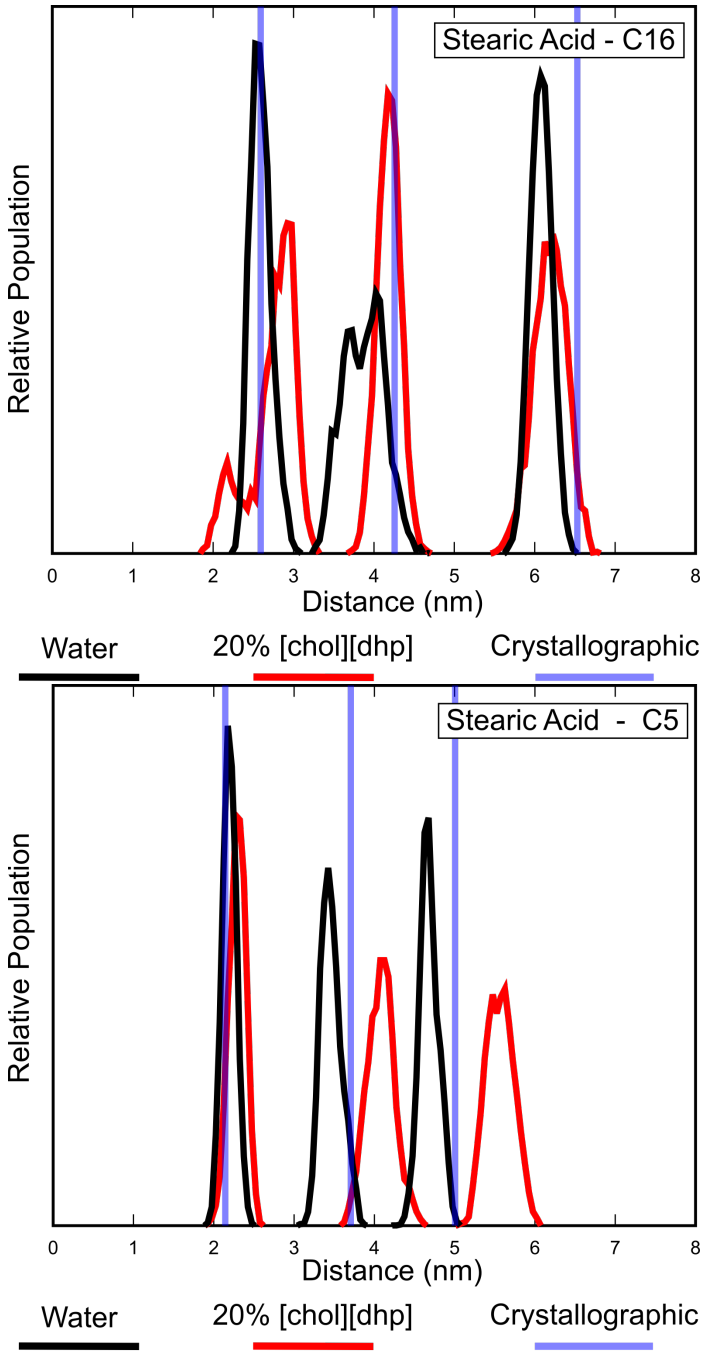



**Figure 4.3:** Stearic acid carbon distance distributions for FA binding sites 2, 4, and 5.

The distance distributions calculated from out simulations indicate that HSA's general structure is similar in water and in ionic liquid. The C5 distribution, which accounts for carbons that are more buried in the binding sites, shows that the stearic acids are closer together in water while they are further apart in 20% [chol][dhp] when compared to the crystallographic structure.

This suggests that protein is swelling in the presence of [chol][dhp]. On the other hand, water tends to shrink the intermolecular distances when compared to the crystallographic structure. The C16 distribution is less conclusive. These carbons are near the ends of the fatty acid molecules, and are thus less buried in the binding sites. The crystallographic data aligns well with water the water case at 2.5 nm, and it aligns well with the 20% [chol][dhp] at 4.2 nm. These data all together would predict that the protein is similarly stable in a buffered solution and in 20% [chol][dhp]. It is difficult to compare directly to the experimental data, because the experiments are able to probe the ensemble of FA distances. This naturally takes into account the relative affinity of the FAs for the binding sites. We, on the other hand, had to assume which sites were highly populated and which were rarely populated. Qualitatively, the double electron-electron resonance (DEER) experiments show peaks in the distribution that are shifted to the right for [chol][dhp] compared to water. This suggests a swelling of the protein in [chol][dhp] compared to water. Our simulations agree with the experimental behavior.

*Denaturation of Loop 1 of Domain I*. The denaturation of loop 1 of domain I of solute-free HSA was examined using metadynamics. By biasing the distance between CYS34 and the nearest neighboring structural features, we are able to uncover the free energy of the transition from folded states to unfolded states in each solvent. Figure 4.4 shows the free energy of domain I of the protein as a function of distance for systems containing 98% IL versus a water case. We notice that the transition from folded to unfolded in water has a low free energy barrier. Still, a tightly packed loop 1 is the preferred structure, where the difference in free energy between the states is about 1-2 kcal/mol. High concentrations of 98% [BMIM][$BF_4$], which are known from experiments to denature this portion of the protein, cause the protein to prefer a state where the cleft near loop 1 is 0.5 nm wider than it is in water. This indicates a swelling of the protein. The

low-distance well is even shifted by a distance of about 0.2 nm. Metadynamics predicts the most thermodynamically stable state of the domain solvated in 98% [BMIM][$BF_4$] is one that has a cleft that is around 1.25 nm wide. This domain is 2-3 kcal/mol lower in free energy than the more tightly-packed low-energy structure.

Preliminary results indicate that this domains behaves differently when solvated in 98% [chol][dhp]. We observe that there are two preferred conformations of the protein. One of these states is similar to the low-energy state observed in the water-solvated case. The other conformation is highly unfolded with a distance of about 1.5 nm. There are many possible ways that the domain can move between these two states, so we cannot say anything conclusive about the energy barrier between these states. Still, we know that the intermediate distances are highly disfavored. The protein may be trapped near its crystallographic structure in 98% [chol][dhp] in both experiments and simulations. This may be why [chol][dhp] has been found to stabilize proteins for long-term storage.

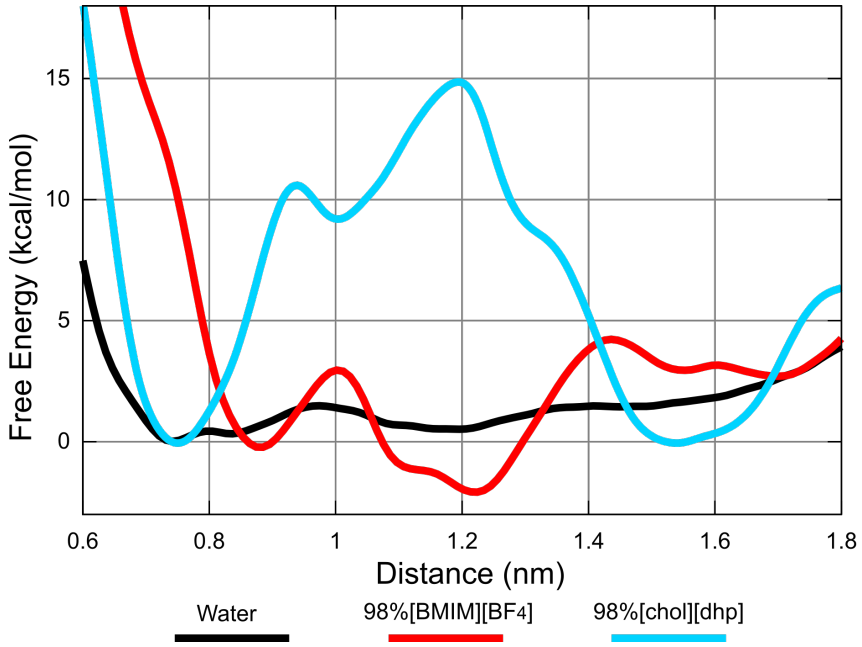



**Figure 4.4:** Free energy of transition for the solvent exposure of CYS34. *Convergence has not yet been rigorously established.*

*Essential Dynamics Coarse Graining*. Essential dynamics coarse graining is a method by which continuous domains of the protein that share correlated motions can be automatically selected. We coarse grained the solute-free protein in high concentrations of IL versus water at elevated temperatures to encourage more dynamic motions from the ostensibly kinetically trapped 98% [BMIM][$BF_4$] system. Figure 4.5 shows the a detailed representation of the sites nearest CYS34, for which we have experimental evidence of denaturation, and a less detailed graph of the domains for the remainder of the protein.

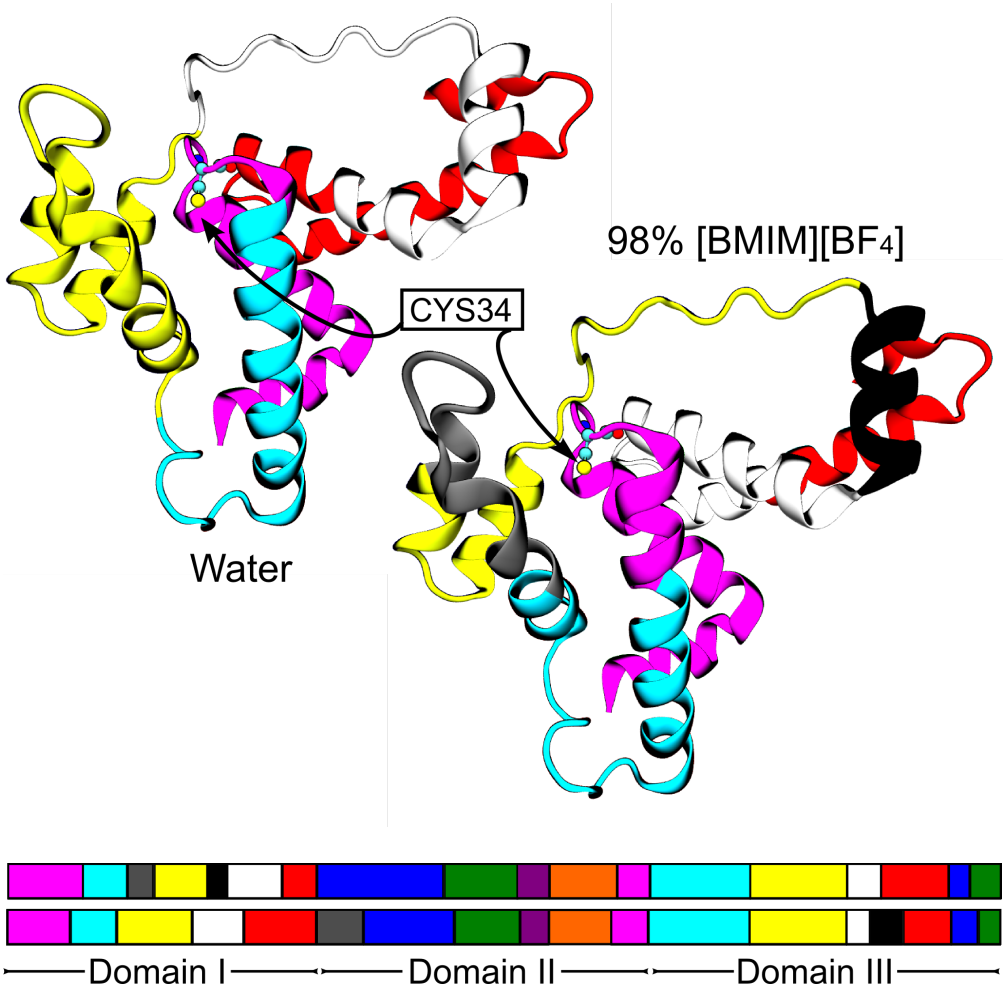



**Figure 4.5:** (A) A rendering of domain I of HSA colored by the domains assigned by EDCG. Notice the two additional domains in 98% [BMIM][$BF_4$] in gray and black. (B) A graphical representation of the domains assigned across the whole protein.

For the system solvated with 98% [BMIM][$BF_4$] we notice that there are two additional domains assigned to domain I. This indicates that domain I has more important fluctuations in domain I than in other domains, which is consistent with the experiments that indicate that the domain denatures. On the other hand, domain III has more important fluctuations for the water solvated case. Ultimately, EDCG shows that there are fundamental differences in the slow modes of motion for HSA solvated in these two different solvents. This agrees with the experimental observation that the correlated motions of the protein are disrupted by the presence of high concentrations of [BMIM][$BF_4$] in water. The two extra domains in domain I for 98% [BMIM][$BF_4$] are colored gray and black in Figure 4.5. Interestingly the additional gray domain (LYS73 to ALA88) contains the region of the protein that we choose to sample in our metadynamics simulations. From this observation, we posit that EDCG might be a way that we can select domains of interest for further metadynamics simulations.

*Charged areas of the protein and their interactions with ILs*. Ionic liquids tend to interact with the charged areas of the surface of a protein. The electrostatic potential of the surface of the solute-free protein was calculated using APBS. This surface charge is shown in Figure 4.6. The negatively charged areas, colored in red, tend to cluster towards the outside of the protein with the highest negative charge being near CYS34. This corresponds to the observation that positively charged ions tend to be attracted to the area. Positively charged imidazolium ions can penetrate the negatively charged cleft and begin the process of denaturing the area around loop 1. The positively charged areas, colored in blue, are clustered in the center of the protein. Because the protein transports negatively charged fatty acids around the body, it makes sense why the interior of the protein is positively charged. Additionally, we have observed that [chol][dhp] tends to swell the protein as in Figure 4.3. The relatively small dihydrogenphosphate ions are to

penetrate the protein. Additionally, having -OH groups allows the formation of hydrogen bonds between the protein and the anion. This close interaction may be the reason why [chol][dhp] is so good at stabilizing certain proteins.

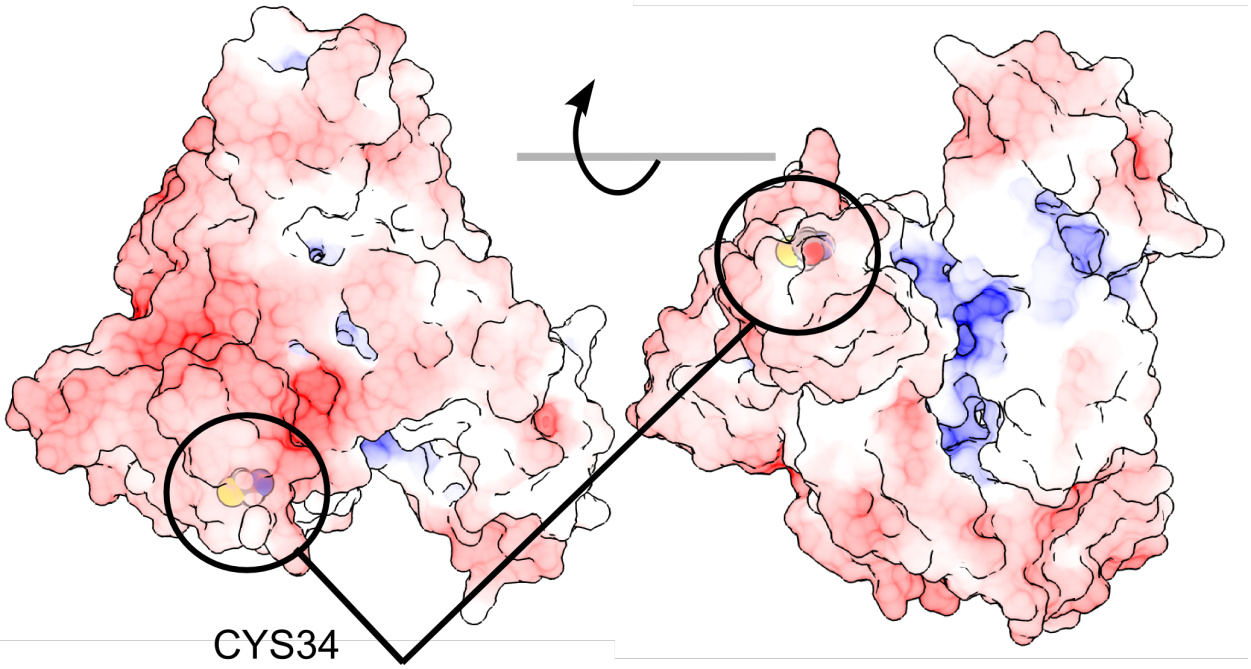



**Figure 4.6:** The electrostatic surface potential of HSA. Red is negative. Blue is positive. Two views of the protein are shown with rotation around the axis shown in gray. CYS34 is circled to guide the eye.

## Conclusions

We used molecular dynamics and metadynamics to analyze the interactions of [BMIM][$BF_4$] and [chol][dhp] with HSA. MD coupled with enhanced sampling techniques have previously been shown to have the ability to replicate experimental results and give insight into the mechanisms by which proteins function. Unfortunately, these techniques are often not predictive, but instead they rely on experimental hypotheses. Our hope is that the techniques we have outlined here will be one step toward more predictive simulations. This is particularly important in non-aqueous solvents where many common experimental techniques are difficult or impossible.

We have shown that limitations of molecular dynamics, with regard to them being kinetically trapped in viscous solvents, can be overcome in certain cases by using enhanced sampling. We posit that EDCG may be used to discover domains of the protein that behave differently in IL vs. water. We suggest that the surface charge of the protein can be used to predict areas that will destabilize in the presence of ILs. Once a region is identified as likely to be affected by the presence of ILs, further metadynamics simulations can be used to focus on those areas of the protein that are most important to structure and function. Our results agree with the findings of two separate experimental studies in a few important ways.

*(A)* Choline dihydrogenphosphate at concentrations of 20 wt% was found to swell the protein.

*(B)* [BMIM][$BF_4$] was found to cause domain I to favor a more unfolded state than compared to the water-solvated case.

*(C)* High concentrations on [chol][dhp] had similar low-energy states to those found in water, and intermediate states between folded and unfolded were highly disfavored.

Ultimately, we believe that the combination of EDCG, metadynamics, and possibly other enhanced sampling techniques are a viable way to gain insight into the interactions of proteins and solvents in cases where standard molecular dynamics is limited in its timescale.

# V. Structure and Free Energy of Leucine Lysine Peptide Aggregates



*Acknowledgement of coauthors:* Experiments were conducted and analyzed by Prof. Joe Baio, Dr. Ariel Zane, Dr. Adrienne Roehrich and Helmut Lutz. My contribution to this publication consists of simulations conducted to support hypotheses put forward by my co-authors. All of the simulations and the analysis thereof were performed solely by me.

## Introduction

Short peptides consisting of leucine and lysine residues exclusively (LK peptides) have been shown by our experimental collaborators to influence the shape and size of silica microparticles precipitated from silica precursor solutions (17% v/v tetramethylorthosilicate in 1 mM HCl). It is hypothesized that the shape of the particles is influenced by the shape of the LK peptide multimers and aggregates that form in high concentrations in solution. We studied three LK peptides with differing secondary structure: LKα14, LKβ15, and $LK_{310}13$. The peptides' secondary structures arise from their different sequences:

LKα14 – Ac-LKKLLKLLKKLLKL-COOH

LKβ15 – Ac-LKLKLKLKLKLKLKL-COOH

$LK_{310}13$ – Ac-LLKLLKLLKLLKL-COOH

It should be noted that there are differences in the ratio of the hydrophobic leucine resides to charged lysine resides. This will play a role in the way the peptides aggregate.

Although our group has published work in the field of LK peptide simulations, we have not studied the aggregation of the peptides. The previous publication by Michael Deighan explored the preferred conformations of atomistic models of LKα14 and LKβ15 at surfaces[240]. Our simulations differ in three ways. First, we use a coarse grained model. This allows us to explore much longer timescales, but we give up some detail in both the solvent and the peptides. Second, we have five peptides in our simulation box instead of the singular peptide in the previous study. Third, we do not have a surface in our simulation box. We are interested in the behavior of the solvated peptides. Ultimately, we are able to use enhanced sampling techniques to efficiently explore a wide range of aggregated conformations for each of the three peptides of interest. Using metadynamics, we are able to discover low energy conformations and speculate on how these conformations template the growth of silica microparticles.

## Methods

*MARTINI Force Field*. The MARTINI force field can be used for coarse grained (CG) modeling of several biologically important molecules such as lipids and proteins[241-244]. MARTINI models the interactions of particles using the same equations as classical MD, but it reduces the detail and complexity of the atomistic model. Up to four atoms are modeled as one site. This helps to speed up calculations in two ways. First, there are fewer particle-particle interactions. Since the calculation of forces between atoms constitutes the largest portion of the computational load in MD, having fewer atoms means a faster simulation. This is compounded by the fact that the cutoff distances for the Lennard-Jones interactions and short-range electrostatic interactions are the same in CG MD as in atomistic MD. This means far fewer forces to calculate. Second, the integration time step can be greatly increased. The time step of a simulation is based on the characteristic motion in the system that has the shortest timescale. For

atomistic MD, this time step is 2 fs. For MARTINI CG, a time step of 40 fs is possible under the right conditions. For our systems, the time step was set at 20 fs, because this time step conserved energy in the NVE ensemble.

MARTINI was originally parameterized for lipids.[244] It was later expanded to proteins.[243] Even later, a polarized protein force field was developed that better described the charged and polar amino acids.[245] When solvating a lipid or protein in water, MARTINI requires a CG water model unlike our atomistic MD simulations that describe water using a standard TIP3P three-site model. The CG water model suggested for proteins is the polarizable MARTINI water model.[246] This models four water molecules as a single entity. This polarized entity contains three sites. The first site is the center of the CG water particle. It is the center from which the Lennard-Jones interactions are calculated. The second and third sites are a positive and a negative site that are connected to the center with bond and angle forces. These sites are allowed to move in order to closely mimic the dielectric constant of water.

There are two main differences in the parameters with which we conduct MARTINI simulations compared to atomistic MD (See Chapter III, Methods for atomistic parameters). The first small difference is in the way the electrostatic forces are calculated. There is an additional parameter in the electrostatic potential, denoted $\varepsilon r$. It is called a relative dielectric constant within the GROMACS manual. In practice, it is a factor that scales the permittivity of free space. Setting $\varepsilon r$ at 2.5 causes the CG polarized water to have a dielectric constant of approximately 80, which is close to that of water. The second difference is in the time constants used for the thermostat and barostat. Since the time step of integration has been increased by a factor of ten compared to atomistic MD, if we do not increase the time constants of the thermostat and the barostat, the simulations will become unstable.

*Simulation Parameters*. LK peptides were simulated using the GROMACS 4[198] molecular dynamics engine patched with plumed 2.0.[236] The peptides were coarse grained (CG) using the MARTINI 2.2 force field[245] with polarizable CG waters.[246] MARTINI does not describe the evolution of a peptide's structure. Instead, the secondary structure of the peptide or protein is calculated using DSSP (Define Secondary Structure of Proteins).[247] The bonded forces between the CG particles are then set in order to maintain the original secondary structure. If there is important tertiary structure such as a disulfide bond, restraints are placed between the cysteine beads to disallow them from moving apart. We do not need to concern ourselves with tertiary structure in the case of LK peptides though. In order to test whether the bonded force field parameters assigned to the beads were sufficient to maintain the experimentally know secondary structures of the peptides, we ran simple CG-MD on the peptides. The 310 helix and the alpha helix maintained structure. On the other hand, the beta peptide began to fold in on itself. This is contrary to what we expect from the experimental data. In order to keep the beta peptide extended, as was found in the experiments, we changed the angle spring constant between the backbone beads from 25 to 100 kJ/mol/rad$^2$ with an equilibrium position of 180° instead of 134°. We also removed the dihedral parameters. These changes kept the peptides extended and the leucines aligned on one side of the chain.

All simulations were conducted in cubic simulation boxes with a side length of 10 nm with a total of five peptides in each box. The box was neutralized with CG chlorine atoms. Electrostatic forces were calculated explicitly for distances of less than 1.4 nm and using particle mesh Ewald summation for long-range interactions. A relative dielectric constant of 2.5 was used in order for the solvent to mimic the dielectric constant of water, as suggested for the polarizable water model. Lennard-Jones interactions were used to calculate non-bonded

interactions and were shifted to zero at 1.2 nm to avoid discontinuities in energy. The CG model was found to conserve energy with a time step of 20 fs, so this time step was used in all simulations. Each simulation was 3.0 μs in length. Simulations were conducted in the canonical ensemble with a stochastic velocity rescaling scheme[204] to maintain the temperature at 310 K.

Well-tempered metadynamics,[239] was employed in order to determine the lowest free energy structures of LK peptide aggregates. The shape of the bias potential added to the CVs was a Gaussian with a height of 2.0 kJ/mol and a width of 0.05 units in CV space. A hill was deposited each picosecond, and the hill height decayed according to the well-tempered metadynamics bias factor of 15. Two CVs were used to promote sampling of a large number of conformations and to describe aggregated states: 1) the radius of gyration (nm) of the centers of mass of the peptides and 2) the coordination number of the centers of mass of the leucine residues. Coordination number is described by the standard sigmoidal function in Plumed 2.0.[236] Structures shown in Figure 5.3 were obtained by extraction from the low free-energy basins of the one-dimensional free-energy surfaces from the metadynamics simulations as projected onto the coordination number collective variable.

*Collective Variables*. We used well-tempered metadynamics to bias two collective variables (See Chapter IV, Methods for a discussion of the theory of metadynamics and collective variables). Our goal was to form and break aggregates of five CG-LK peptides in a simulation box. For this task, we chose radius of gyration of the centers of mass of the peptides and a coordination number among the centers of mass of the leucines in each peptide. The mass-weighted radius of gyration, $R_g$, is described by Equation 5.1. The mass of each peptide is denoted $m_i$, the positions of the centers of mass of the leucines of peptide $i$ is denoted $r_i$, and the center of mass (which is the geometrical average of $r_i$ in this case) of the leucines of the $n$

peptides is denoted $r_{COM}$. In our case of five identical peptides, the masses drop out of the equation.

(5.1) $$R_g = \left(\frac{\sum_i^n m_i(r_i - r_{COM})^2}{\sum_i^n m_i}\right)^{1/2}$$

(5.2) $$s_{pairwise(i,j)}(r_{ij}) = \frac{1-\left(\frac{r_{ij}-d_0}{r_0}\right)^n}{1-\left(\frac{r_{ij}-d_0}{r_0}\right)^m}$$

(5.3) $$s_{system} = \sum_{i>j} s_{pairwise(i,j)}$$

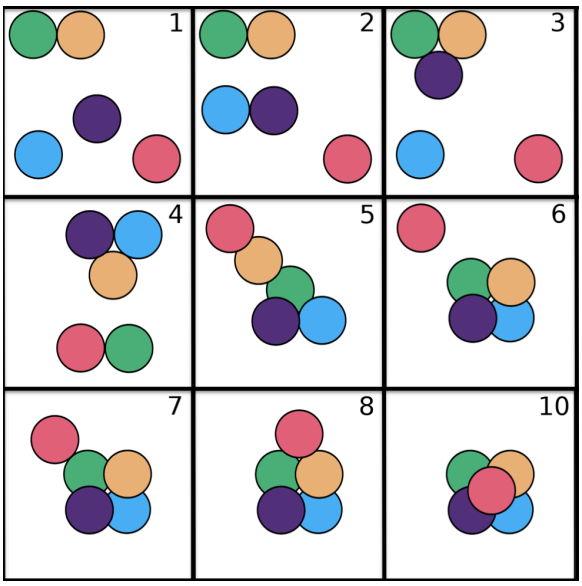



**Figure 5.1:** Illustrations of possible system coordination numbers. These are not exhaustive. There may be degenerate states.

Equation 5.2 describes a sigmoidal function that quantifies the association of any two peptides and thus the coordination number. The coordination number *s* depends on the distance between the centers of mass of the leucines between two separate peptides. The value of *s* decays from one to zero as the radius between the peptides increases. Variables $n$, $m$, and $r_0$ are set to change the sharpness and center of the sigmoidal curve. In our simulations, $n = 8$, $m = 16$, and $r_0 = 1.6$ nm. The curve can be shifted along the x-axis by using $d_0$. We choose to set this variable to

zero. The coordination number that we use to describe our system is the sum of the coordination number for all pairwise interactions as shown in Equation 5.3. For the system as a whole, a coordination number of zero describes a system where all peptides are separated. A coordination number of ten describes a closely packed structure where all of the peptides are interacting with each other. Figure 5.1 illustrates possible states described by each system coordination number described in Equation 5.3.

## Results and Discussion

*Free Energy of Aggregation*. After running each simulation for 3 μs, it was clear that the free energy had converged. We decided to project the free energy onto the coordination number collective variable in order to better understand the physical conformation that each energy-well described. Figure 5.2 shows this projection.

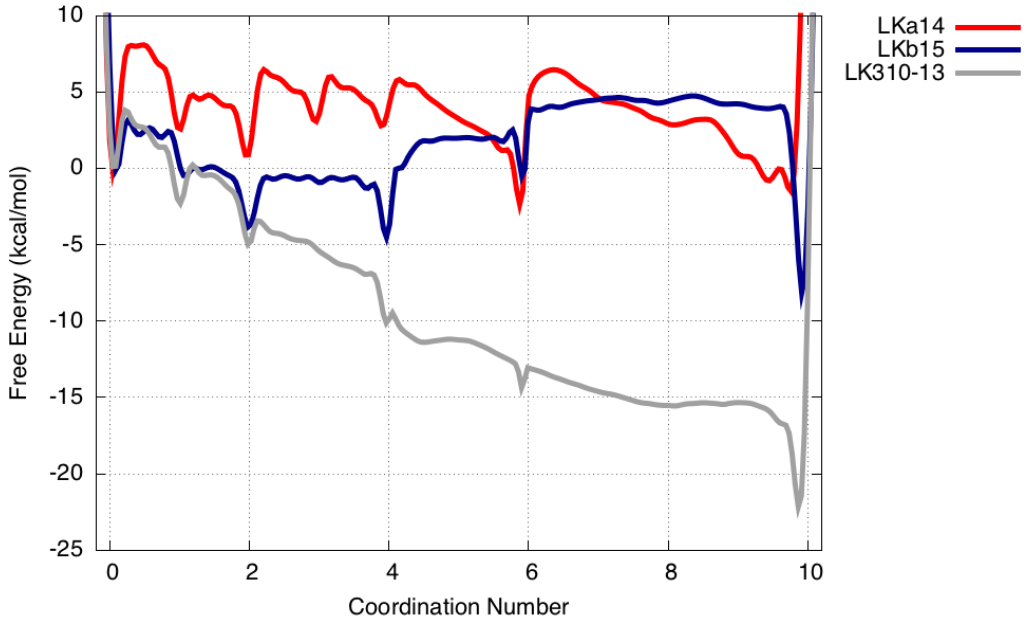



**Figure 5.2:** System coordination number for each peptide sequence.

The shapes of the three peptides' aggregate structures are quite different from each other. It can be seen that for LKβ15 and $LK_{310}13$ the lowest free energy states are those where all five peptides are closely associated with each other. On the other hand, LKα14 has two similarly low energy states. One is a tetramer, and the other is a pentamer. A tetramer is the experimentally confirmed low energy aggregate of LKα14, which is promising for the validity of our techniques.

*Low Energy Structures*. In order to understand the differences in morphology between the lowest energy states of the aggregates, one must visualize the trajectories. The first row of Figure 5.3 illustrates a low energy structure for each of the peptides. The backbone particles are drawn as colored beads. The individual peptides are colored differently, and a vector is drawn between the ends. The second row of Figure 5.3 illustrates the same low energy structures with a different representation.

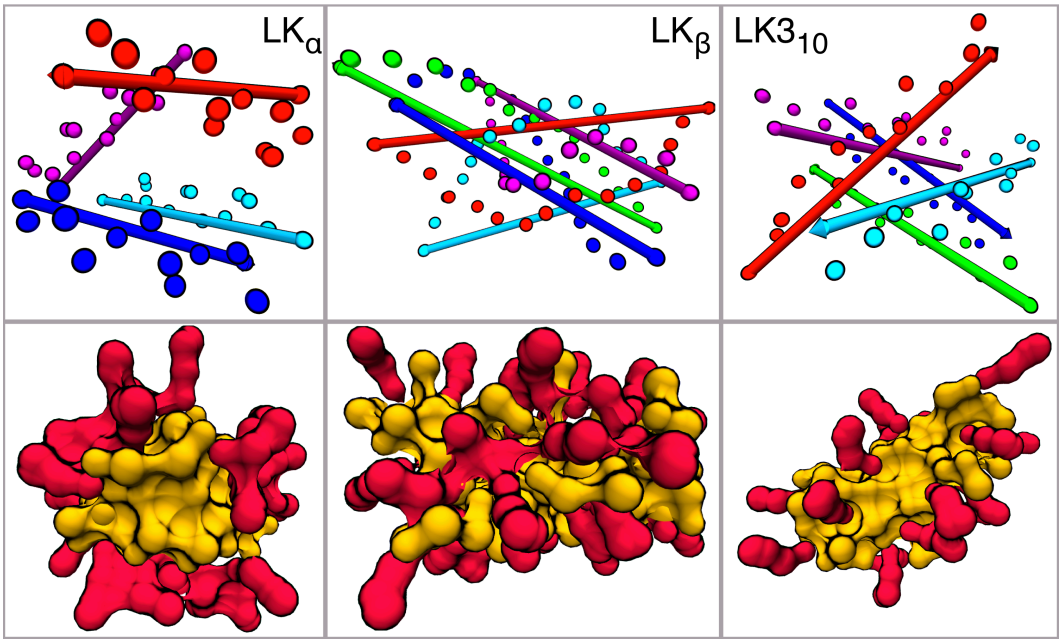



**Figure 5.3:** Illustrations of low energy states. (Row 1) Each peptide is colored uniquely. Beads represent backbone particles. Vectors are drawn between the ends. (Row 2) Surface representation. Hydrophobic leucines are in orange. Charged lysines are in red.

The orange center and the red extrusions are the surfaces of the leucines and the lysines respectively. As expected for each of the three peptides, the hydrophobic leucines provide the cohesive interactions that cause the formation of aggregates while the positively charged lysines are pushing nearby peptides apart. We hypothesize that these low energy aggregates template the formation of silica. Figure 5.4 contains SEM images of the silica particles precipitated from solutions of the LK peptides and a silica precursor. These images were produced by our experimental collaborators. The second row of Figure 5.4 contains representations of surface structures predicted by sum frequency generation spectroscopy.

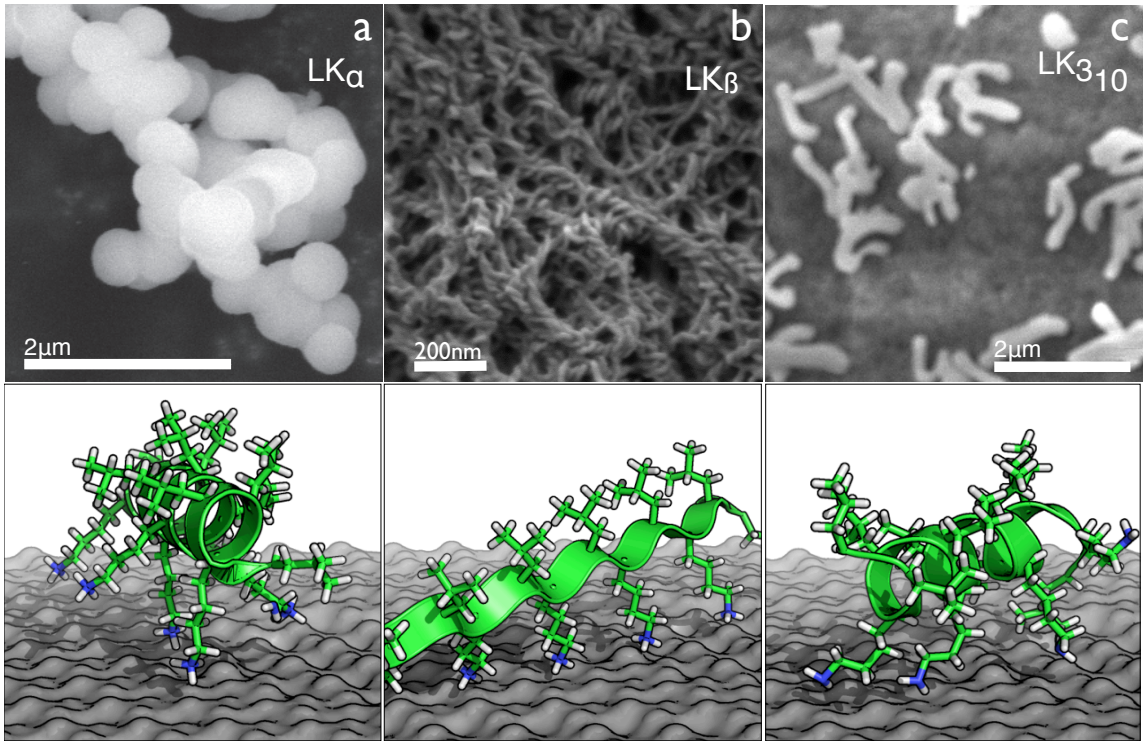



**Figure 5.4:** (Row 1) SEM images of the biosilica nano- and microparticles templated by each of the LK peptides tested. (Row 2) Representations of predicted surface structures for each peptide.

First, we notice spaghetti-like shape of LKβ15 silica. Comparing this structure to the structures seen in Figures 5.3, it can be seen how the simulated aggregates would template the experimentally observed silica nanoparticles. There are two important considerations about the

sequence and structure of LKβ15. This peptide has the highest ratio of K to L. Thus, they are the most highly charged peptides of the three tested. This is coupled with the rod-like shape of the aggregates. The outside of the rod is highly charged. The rods will not interact side-by-side. There are no hydrophobic residues exposed in a tight rod. Instead, if one or two peptides shift a small amount down the length of the rod, hydrophobic residues become exposed at the ends, and the rods can be extended further and further to make the spaghetti-like structure seen with SEM.

$LK_{310}13$ forms a pill-like aggregate. It has the lowest ratio of K to L. There are still exposed hydrophobic residues. If these small pill-like aggregates can interact mediated by the silica precursor, they will grow in all directions equally to form pill-like nano- or microparticles. To some extent, the same argument can be made for LKα14, which forms a more spherical aggregate according to our simulations. LKα14 has an intermediate ratio of K to L, and it has fewer exposed hydrophobic areas than $LK_{310}13$. Nonetheless, we hypothesize a similar mechanism of biosilica formation.

## Conclusions

1. Metadynamics can help us determine the size and morphology of these small peptide aggregates. For more complex systems, this will likely not be the case.

2. Coarse grained models contain enough detail to predict low-energy aggregates that correspond to those found in experiments. We cannot say anything about changes in the peptides' secondary structure during the simulations.

3. The high charge of the system did not affect the application of metadynamics or coarse graining. Yet, one would require coarse grained phosphate ions and silicic acid to simulate the full process of nanoparticle formation.

# VI. Orientation of Leucine Lysine Peptides at Air-Water Interfaces



*Acknowledgement of coauthors:* Experiments were performed by Helmut Lutz and Dr. Rüdiger Berger. With Helmut Lutz, I designed and conducted molecular dynamics simulations of LK peptides at an air-water interface and how they interact with silicic acid mimics in the aqueous solution. I conducted the analysis of the molecular dynamics trajectories.

## Introduction

Previously we explored the morphology of silica nanoparticles whose growth was templated by the presence of LK peptides.[248] Helmut Lutz and Prof. Tobias Weidner discovered that the same LK peptides that template nanoparticle growth in solution can also template the formation of thin silica films at an air-water interface. It is known that LK peptides have specific preferred structures at hydrophobic or hydrophilic interfaces.[249] At the air-water interface the hydrophobic leucine resides orient into the air while the charged-polar lysine residues point into the solution. Silicic acid in the solution orders near the lysine residues, and it is hypothesized that the lysine residues help to catalyze the formation of silica from the silicic acid precursors. Figure 6.1 illustrates the hypothetical mechanism by which the silica thin film forms. The resulting silica has a thickness on the order of nanometers, and it is an order of magnitude smaller than other biosilica thin films. The films can be removed intact from the interface and probed by other techniques such as electron microscopy, atomic force microscopy and x-ray photoelectron spectroscopy.

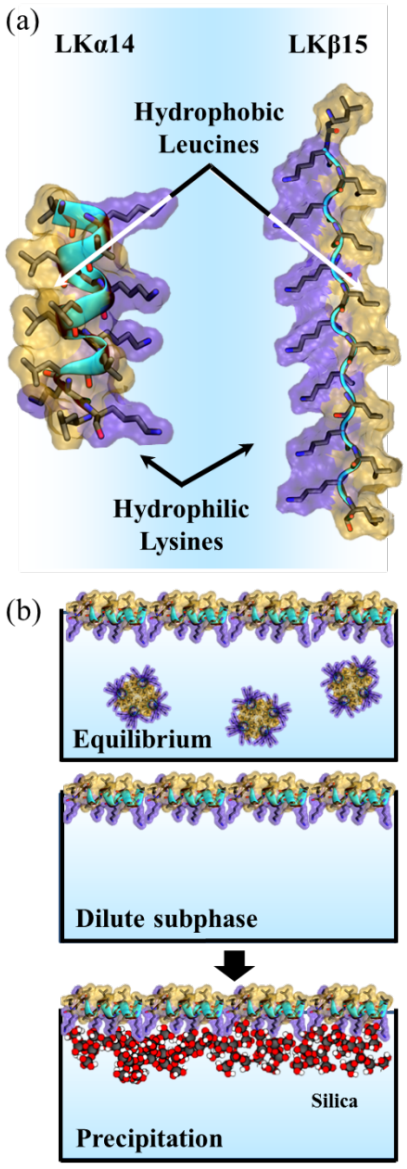



**Figure 6.1:** The hypothesized method of silica thin layer formation. (a) The LK peptides explored in this study. (b) LK peptides are placed into solution and segregate into aggregates and surface-peptides. Aggregates are diluted out of the solution. Silica precursors are added to the solution and interact with the lysine residues.

Since the LK peptides orient at an air-water interface, sum frequency generation spectroscopy (SFG) is a highly appropriate method to probe their ordering in solution. SFG is sensitive to ordering of molecules at interfaces. An infrared laser and a visible laser are directed at the same point on a surface or an interface. The infrared laser frequency is modulated over a range that contains vibrational modes of interest. Meanwhile the narrow-band visible laser is

kept at a constant frequency. When there is ordering of the molecules at the interface, a photon at the sum of the frequencies of the two lasers is emitted. Thus, the method is an ordered-interface-sensitive form of infrared spectroscopy.

## Methods

Protein, phosphate, and silica potentials came from the AMBER99SB-ildn[200], GAFF[237], and modified version of an established force field[250], respectively. Silica dihedral parameters were developed by fitting to B3LYP/6-311G(d) quantum mechanical calculations. Phosphate and silica electrostatic point charges were developed at the HF/6-31G(d) level of theory using the RESP[151] method with Gaussian09[25] and the antechamber package in Ambertools (Figure S6.1). AMBER-type files were converted to GROMACS[235] using acpype[29]. Molecular dynamics simulations were performed using GROMACS 4.6. Freezing the hydrogen-heavy atom bonds and applying the LINCS algorithm allowed for a timestep of 1 fs. Systems of approximately 8.0 x 8.0 x 13.8 nm were built using Packmol[31]. Of the 13.8 nm in the z dimension, about 7.0 nm was vacuous. Periodic boundary conditions were applied in all three dimensions to use particle-mesh Ewald summations for long-range electrostatics. This simulates x-y slabs separated by 7.0 nm of vacuum. Lennard-Jones potentials were shifted to zero at 1.2 nm. Temperature was maintained at 310 K using a stochastic velocity-rescaling thermostat[34].

## Results and Discussion

*Experimental Results (performed and interpreted by H. Lutz and T. Weidner).* Surface spectra of a solution of LKα14 and LKβ15 (0.025 mg/ml), before and after injecting the silica precursor tetramethyl orthosilicate (TMOS), are shown in Figure 6.2. Exchanging the subphase of the solution three times to reduce the bulk solution concentration of LK aggregates prior to the

TMOS injection did not change the spectra. In agreement with previous studies of LK peptide films on hydrophobic surfaces, resonances near 2880, 2904, and 2921 $cm^{-1}$ are observed and can be related to CH-stretching vibrations from ordered leucines pointing towards the air.[251-253] The broad spectral feature seen in Figure 6.2a for LKα14 in the range of 3000-3300 $cm^{-1}$ is the result of stretching modes of differently hydrogen bonded water OH groups as verified by changing the ionic strength of the solution to alter the OH ordering.[254] Remarkably, no substantial water related resonances are observed in the case of LKβ15 (Figure 6.2b).

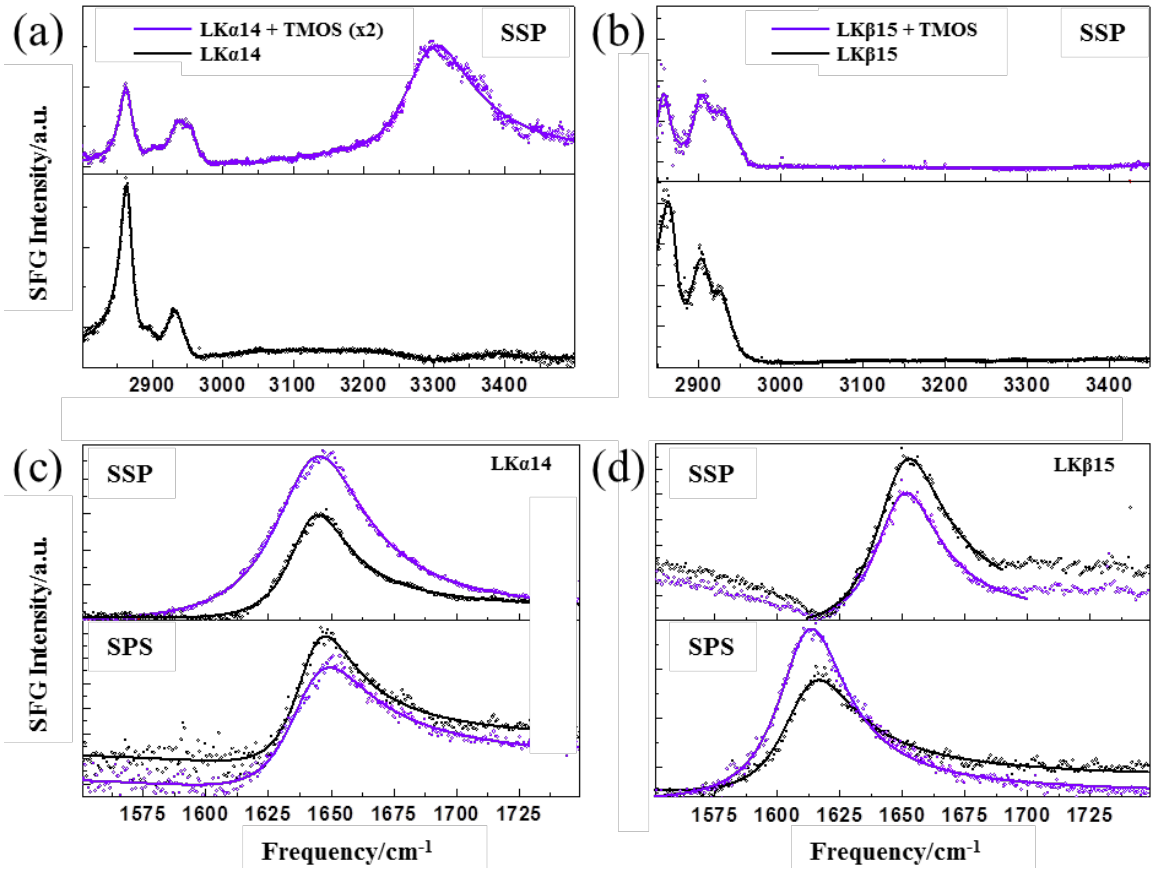



**Figure 6.2:** SFG spectra before (black) and after TMOS addition (violet) in the spectral range of CH and water vibrations: LKα14 (a) and LKβ15 (b) at the air-water interface. In the range of backbone vibrations: (c) ssp and sps spectrum of LKα14. (d) ssp and sps spectrum of LKβ15.

Upon TMOS addition, a peptide-silica film is formed at the surface. In the case of LKα14, a new feature appears at ~3300 $cm^{-1}$ attributed to NH stretching vibrations of the lysine

side chains (Figure 6.2a).[255] The emergence of the NH peak suggests a large increase in the collective molecular order of lysine amine groups, likely due to the peptide-mineral interaction. In contrast, the spectra of LKβ15 before and after TMOS addition (Figure 6.2b) reveal no significant changes in the NH stretch region. This observation indicates that, in contrast to the helical fold, lysine side chains in LK interaction. In contrast, the spectra of LKβ15 before and after TMOS addition (Figure 6.2b) reveal no significant changes in the NH stretch region. This observation indicates that, in contrast to the helical fold, lysine side chains in LKβ15 are not interacting in a collective manner with the silica during the sheet formation. The simulation results shown below indicate that backbone amide modes might also be more involved in the silica interactions.

The Amide I SFG spectral region provides further information about the secondary structures of the two peptides before and after biomineralization. The results for LKα14 (Figure 6.2c) show a resonance centered near 1642 $cm^{-1}$ that is assigned to an α helix, which is in agreement with previously published SFG spectra of LKα14 on hydrophobic surfaces.[256] No significant spectral changes are apparent upon interaction with TMOS, thus we conclude that LKα14 retains its native folding in presence of the silica film. Figure 6.2d shows the results for LKβ15 acquired in ssp and sps polarization combinations, respectively. For ssp, a resonance centered at ~1655 $cm^{-1}$ can be assigned to the B1 β-strand mode, while in sps only the low frequency B2 mode can be observed near 1615 $cm^{-1}$.[257] Chirgadze and Nevskaya characterized the resonance splitting of amide vibrations by transition dipole coupling.[258] We observe the B1 mode in ssp and the B2 mode in sps, indicating the peptide is strongly oriented. The ssp polarization probes modes with transition dipole moments (TDMs) parallel and perpendicular to the surface, while sps probes TDMs oriented in the surface plane. Since the TDM of the B1

mode is oriented along the peptide chain and the B2 mode TDM points in the direction of the C=O bonds, we can conclude that LKβ15 is oriented with the long peptide axis flat on the surfaces with the C=O bonds oriented predominantly parallel to the water surface.

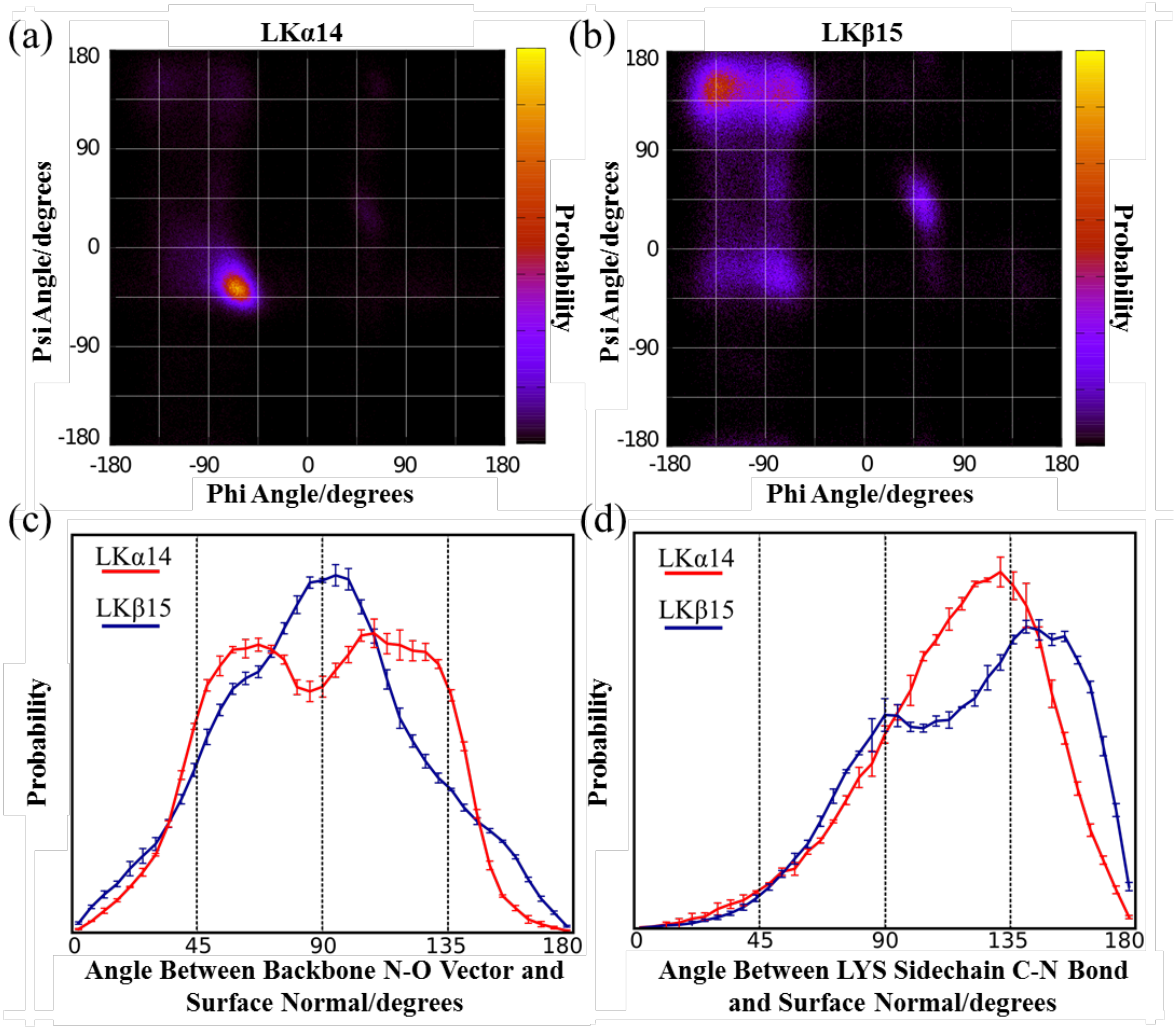



**Figure 6.3:** (a,b) Ramachandran plots for backbone phi and psi angles over the last 100 ns of the simulation. (c) Simulated LK peptide angle distributions of the angle between the N-O vector (backbone nitrogen atoms - nearest carbonyl oxygen) and the surface normal. (d) Simulated angle distributions for the angle between the terminal LYS side chain C-N bond and the surface normal.

*Molecular dynamics simulations (designed and performed by H. Lutz and V. Jaeger, analyzed by V. Jaeger).* Simulation boxes contained 23 LKα14 or LKβ15 peptides at the vacuum–water interface with chlorine ions, phosphate ions and a silica oligomer ($H_{15}Si_7O_{22}^-$)

present in the aqueous phase. Since the intermediate structures formed in the transition from silicic acids to silica are complex and varied, we selected one possible intermediate oligomer that (a) contained functional groups from silicic acid and (b) was small enough and flexible enough to reorient during a 250 ns simulation. Simulations were initialized with LK peptides oriented as a monolayer with leucine residues pointing into the vacuum. The great majority of the peptides remain in a monolayer near the interface throughout the entire simulation, with only two LKβ15 peptides and no LKα14 traveling across the box to find the other vacuum-water interface. Moreover, the secondary structure of the peptides remains stable as measured by Ramachandran plots (Figure 6.3a,b). LKα14 is almost exclusively in the right-handed alpha helix conformation, while LKβ15 heavily favors the beta conformation with small populations in the right and left-handed helix conformations. The high stability of the LK monolayer is in good agreement with the experimental observation that the peptide layer is not affected by removal of peptides from the solution subphase. Within 5 ns of the beginning of the simulations, phosphate and chlorine ions preferentially interact with the lysines near the interface driven by favorable electrostatic forces. Simultaneously, silica oligomers are likewise attracted to the lysine residues, and begin to aggregate into clusters of around 2.0 nm in diameter. The series of events is observed in both the LKα14 and the LKβ15 cases, and the structures that are formed (both the peptide monolayer and the silica clusters) remain stable for the entire 250 ns simulation. Snapshots of the simulations at 5 ns and 100 ns are presented as Figure 6.4.

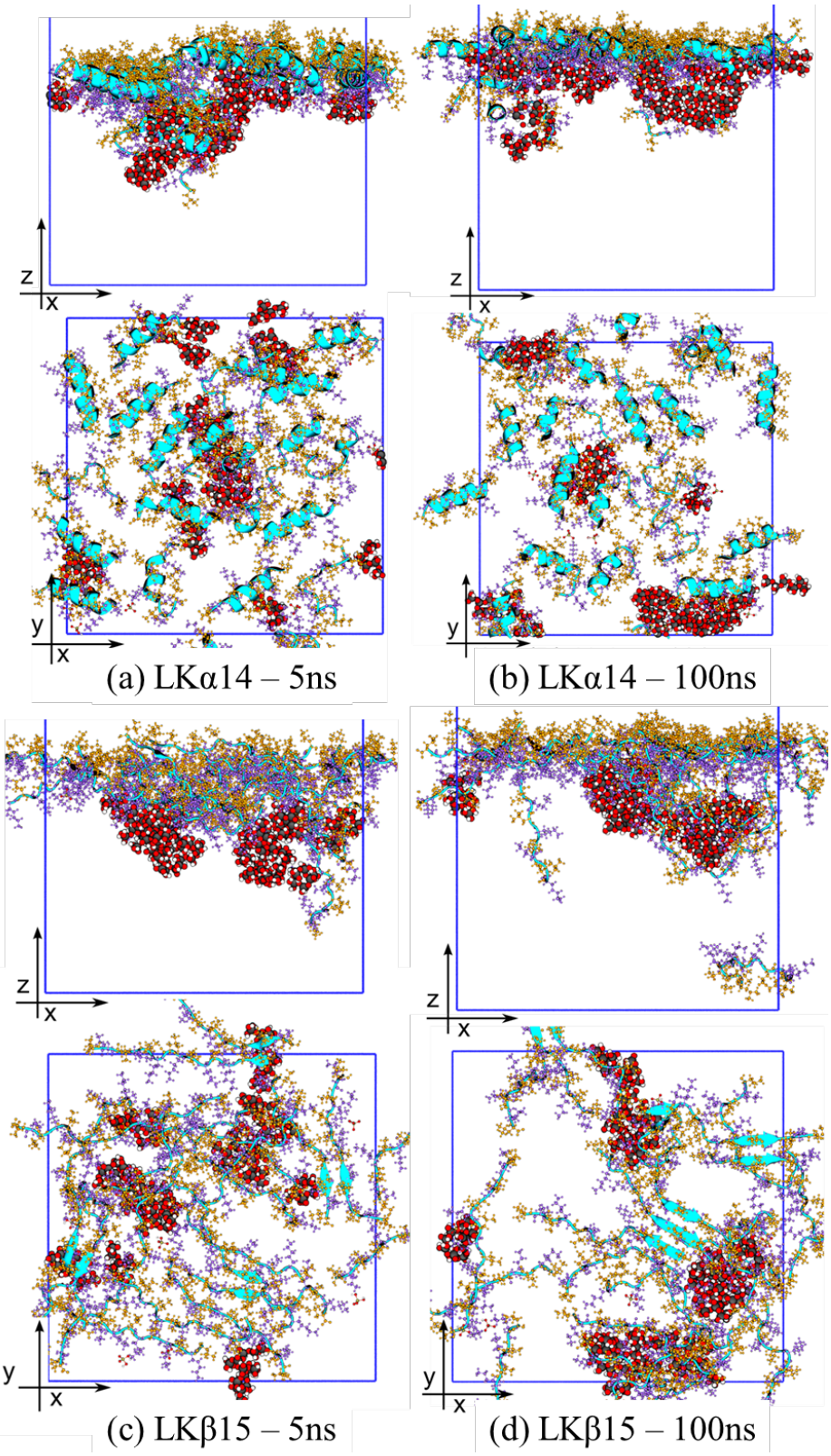



**Figure 6.4:** Snapshots from the MD simulations. (a) LKα14 after 5 ns of simulation. (b) LKα14 after 100 ns. (c) LKβ15 after 5 ns. (d) LKβ15 after 100 ns. Peptides are stable at the interface. Clusters of silica precursor molecules are stable.

Closer inspection of the simulation results reveals several important differences in the interactions of LKα14 and LKβ15 with the silica precursors. LKα14 shows a rather sharp angle distribution for the lysine side chains in contact with silica while LKβ15 shows a random orientation between ~30 and ~100 degrees (Figure 6.3d). This supports the view that the differences in NH signal near 3300 $cm^{-1}$ observed in the SFG spectra are related to different degrees of order at the mineral interface. Analysis of the amide bond orientations within 1.5 nm of the interface shows that while the LKα14 bonds are broadly distributed the peptide bonds within LKβ15 are well aligned with an orientation parallel to the surface (Figure 6.3c) in agreement with SFG results.

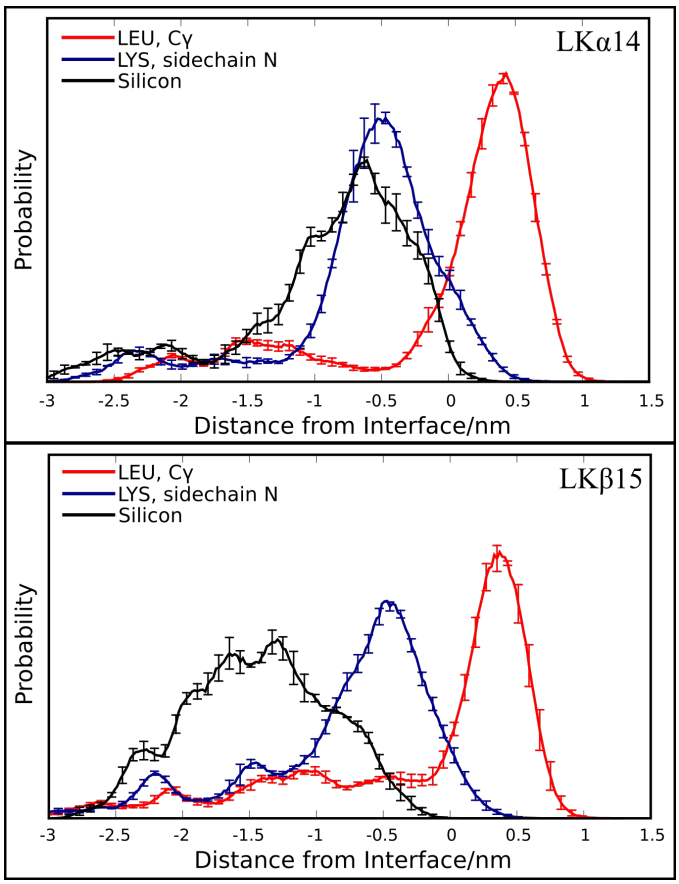



**Figure 6.5:** Simulated number density of certain important atoms with respect to distance from the interface. The interface is defined as the intersection of the LEU and LYS curves. Data from the last 100 ns of the simulation.

Quantitative analysis of the silica-backbone distances shows that LKβ15 interacts with silica more strongly via its backbone amide groups compared with LKα14. In addition, an analysis of the silica-backbone bond length as a function of distance to the air-water interface shows that the LKα14 sheet is comparatively stable and draws the TMOS molecules into the interfacial layer while the silica-backbone bonds in the LKβ15 sheet are less localized and peptides interacting with silica are spread over several nanometers into the bulk solution (Figure 6.5). The selective ordering of the LKa14 side chain is also consistent with the observed XPS results. It is important to note that the experimentally observed sheets are assembled over minutes – the simulations capture only the first 250 ns of the mineralization process.

## Conclusions

To conclude, our experimental collaborators discovered that amphiphilic LK peptides can precipitate nanometer thin self-supported silica sheets. The sheets are stable and can be transferred to solid substrates. They demonstrated that the morphology of the sheets critically depends on the secondary structure of the peptide. SFG analysis indicated that peptides, depending on their folding, control silica biogenesis differently – helices interact via the side chains, while the β-strand motif interacts via backbone groups.

Our simulations supported the experimental findings by giving a molecular level picture of the interaction of silicic acid particles with LK peptides at air-vacuum interfaces. While these simulations did not contain ionic liquids, the methods that we used to parameterize the silicic acid and phosphate force fields are similar to those we employ for ionic liquids. In fact, the phosphate ion we used in these simulations is the same as the one used in Chapter IV with choline dihydrogenphosphate. The results presented for these atomistic simulations of LK

peptides at the air-vacuum interface demonstrate that our techniques can be used to parameterize force fields that interact in feasible and experimentally verifiable ways with peptides. Of particular note is that the lysine residues are among the most important amino acids for the interactions of salts with proteins. This adds to our confidence in the ability of our force fields to accurately describe the interactions of ionic liquids and biomolecules.

# VII. Structure of Metal-Surfactant Complexes in Organic Solvents



*Acknowledgement of coauthors:* Dr. Pablo de la Iglesia, Yuyin Xi and Prof. Lilo Pozzo performed and analyzed the experiments. I designed, executed and analyzed molecular simulations of metal-surfactant complexes in organic solvents to determine if the experimental results were consistent with physical models.

## Introduction

Surfactants are known to modify the properties of apolar solvents in ways that increase their uses as lubricants and dispersants, and they increase the conductivity of apolar media.[259] Surfactants in apolar media form inverse micelles. These inverse micelles, in contrast to those created in polar or aqueous media, form with the long apolar tails exposed to the solvent and a charged core.[259-262] The size, shape and properties of inverse micelles depends highly on the species of surfactant, the presence and species of counterions, and the amount of residual water present in the solvent.[263] One possible type of inverse micelle that has been described in the literature contains multivalent metal ions in the core.[264] Several multivalent metal surfactant complexes have been investigated including $Mg^{2+}$, $Ca^{2+}$, $Co^{2+}$, $Cu^{2+}$, $Al^{3+}$, $Gd^{3+}$, $Fe^{3+}$, and $Ho^{3+}$.[265-275] The metal ion affects the morphology of the micelles; spherical, ellipsoidal, and cylindrical micelles have been known to form. Metal ions also alter the properties of the systems enabling

use in applications such as templates for metallic nanoparticles,[276-278] catalysis,[279] luminescence,[272] and nanomagnets,[265, 280, 281] among others.

The experimental work by Dr. de la Iglesia focused on the micelles formed around the trivalent meal ions of vanadium, iron, and aluminum. The surfactants studied were dioctyl sulfosuccinate (AOT) and dodecylbenzene sulfonate (DBSo). The solvents were toluene and several alcohols. The simulation work performed by me focused on a selection of the micelles produced from the above ingredients [(a) Fe:AOT in toluene, (b) Fe:DBSo in toluene, (c) Fe:AOT in methanol, (d) Fe:DBSo in methanol]. The structures of metal-surfactant complexes were studied with small angle neutron and X-ray scattering (SANS, SAXS) as well as with molecular dynamic (MD) simulations. The combination of these experimental and computational techniques allowed us to understand the molecular structure of these micelles and the ways in which they form.

## Methods

*Molecular Dynamics Simulations*. Molecular dynamics (MD) simulations were conducted to gain insight into the atomic scale structure of the organometallic complexes. All systems contained 8 $Fe^{3+}$ ions and 24 surfactant molecules in a cubic box with a side length of 8.0 nm filled with organic solvent. Four systems were considered: (a) Fe:AOT in toluene, (b) Fe:AOT in methanol, (c) Fe:DBSo in toluene and (d) Fe:DBSo in methanol. The initial configuration of the simulation was randomly generated with Packmol.[202] System details appear in Table S7.1. The general amber force field[237] was used to describe surfactant and solvent molecules.[237] Electrostatic point charges were developed by fitting to HF/6-31G* quantum calculations in Gaussian09[25] with the RESP[151] method in antechamber. The $Fe^{3+}$ Lennard-Jones parameters were taken from the literature,[282] and the charge was assumed to be +3*e*.

GROMACS 4.6[198] was used to perform MD calculations. Lennard-Jones potentials were shifted to zero at 1.2 nm. Particle mesh Ewald summation was used for long-range electrostatics with periodic boundary conditions. Bonds were constrained with the LINCS algorithm in order to allow for an integration timestep of 2 fs. All simulations were kept at 310 K by the application of a stochastic velocity-rescaling thermostat.[204] System were first minimized over 10,000 steps with the steepest descent method, pressure was equilibrated to 1 bar with a 20 ns simulation with a Berendsen barostat, and subsequent production runs of 200 ns in length used the Parrinello-Rahman barostat at the same pressure.[205, 283] Production simulations for DBS in methanol were extended by 100 ns to ensure convergence.

*Forming the metal-surfactant complexes* (performed and interpreted by P. de la Iglesia, Y. Xi, and L. Pozzo). The metal-organic complexes were formed by solvating surfactants in deionized water and then slowly adding metal chlorides to the solution. The precipitate formed by the addition of the metal chlorides was filtered and dried under vacuum for 24 hours. The resulting solid or paste was stored under vacuum to prevent oxidation. Interestingly, the different metals-surfactant precipitates had a wide variety of colors and sizes (Figure 7.1).

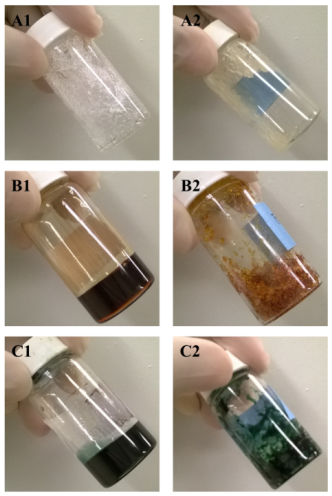



**Figure 7.1:** Pictures of the purified and dried organometallic complexes. The samples in row A are aluminum, row B iron and row C vanadium complexes. Column 1 contains AOT, column 2 is DBSo.

## Results and Discussion

*Neutron and X-ray Small-Angle Scattering of Organometallic Complexes (performed and analyzed by P. de la Iglesia and L. Pozzo).* Small angle scattering was used to probe the structures formed by the organometallic complexes dispersed in organic solvents. In order to obtain a clear analysis of the conformation of the samples, small angle neutron (SANS) and X-ray (SAXS) scattering was performed on all samples. For SAXS, scattering contrast depends on the local electron density. For this reason, SAXS primarily probes the size and shape of the metal core and the surfactant head groups. On the other hand, SANS can probe the characteristics of the surfactant tails when they are surrounded by deuterated solvents. An accurate description of the structure of these complexes can be obtained by simultaneously analyzing the SANS and SAXS scattering data. Both profiles are fit with a core and shell form factor while constraining the size, shape and volume fraction to be equal.

The best fit for most samples in toluene was obtained from a core and shell ellipsoid or cylinder model. Scattering data are presented in Figure 7.2. For a more detailed description, Table 7.1 lists parameters obtained from simultaneous SAXS/SANS fits of iron complexes in toluene. For any of the metals studied, the equatorial radius of the ellipsoid is always 12-15 Å, while the polar radius remains around 5-8 Å. The size of the shell depends on the size of the tail group. DBSo has a long, unbranched tail while AOT has a shorter, branched tail.

**Table 7.1:** Parameters obtained from the simultaneous fitting of the SANS and SAXS data of the iron organometallic complexes dispersed in toluene using the core and shell ellipsoid model.

| Metal: Surfactant | Fe:AOT | Fe:DBSo |
|---|---|---|
| Equatorial Radius Core(Å) | 11.7 | 13.6 |
| Equatorial Thickness (Å) | 7.8 | 11.1 |
| Polar Radius Core (Å) | 6.4 | 4.9 |
| Polar Thickness (Å) | 7.2 | 10.6 |

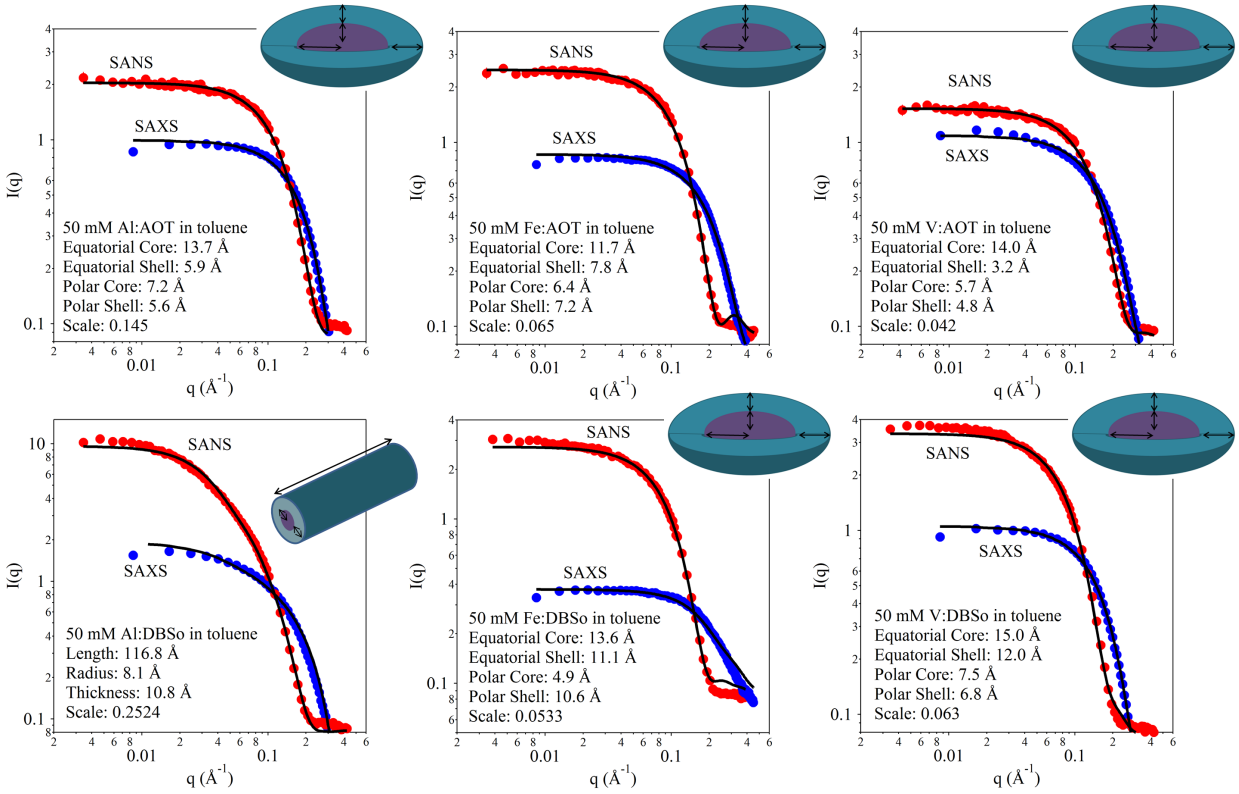



**Figure 7.2:** Scattering data for all of the metal surfactant complexes. Data are fit with a core-shell model with ellipsoids or cylinders.

For organometallic complexes dispersed in methanol the Guinier regions are shifted to higher-q, indicating the presence of smaller features when compared to toluene. The Guinier region is very close to the maximum-q that is measurable with this neutron scattering instrument. This means that small features are barely visible in some of the SANS profiles but they are still very apparent in the SAXS profiles due to the lower incident wavelength and a larger maximum detector angle. Due to the limited resolution of the SANS profiles, simultaneous fits are not attempted on samples in methanol. Instead a simple sphere form factor is fit to the SAXS data to model the dimensions of the core as in Figure 7.3. The size of the spheres are found to be 3.4 Å for Fe:AOT and 5.3 Å for Fe:DBSo.

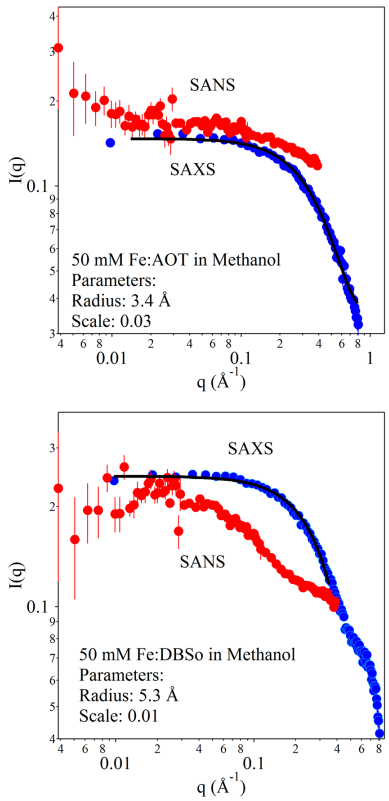



**Figure 7.3:** Scattering data for iron surfactant complexes. Data are fit with spheres for the electron-dense core.

*MD Simulations of Organometallic Complex Dispersions*. MD simulations of the systems were performed in order to further corroborate the existence of the structures that are projected from small angle scattering experiments. Figure 7.4 shows the conformation of Fe:DBSo (Figure 7.4a) and Fe:AOT (Figure 7.4b) dispersed in toluene. Snapshots were taken from a converged final structure of the trajectory. The structures observed in the Figure 7.4 were formed from their random initial configurations within 20-170 ns and are stable for at least 80 ns. Convergence times were as follows. Fe:AOT methanol - 20 ns. Fe:AOT toluene - 80ns. Fe:DBSo methanol - 170 ns. Fe:DBSo toluene - 120 ns.

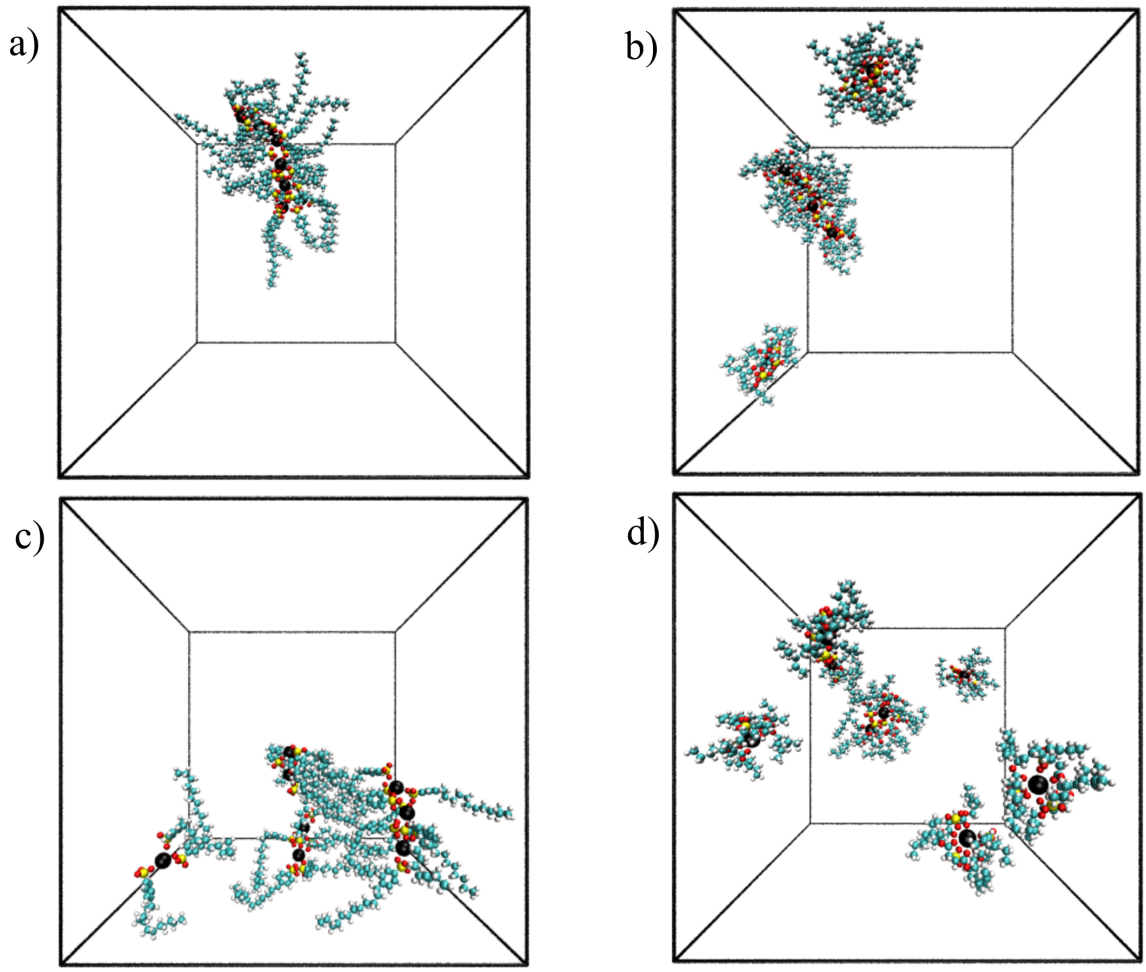



**Figure 7.4:** MD simulation of different iron complexes. a) Fe:DBSo in toluene, b)Fe:AOT in toluene, c)Fe:DBSo in methanol, d) Fe:AOT in methanol. Iron III ions are shown with a larger sphere size (black) in order to facilitate their identification.

By visual inspection, it can be seen that all of the metal complexes are somewhat elongated but their small aspect ratio more accurately defines them as ellipsoids rather than cylinders. This result agrees very well with the small angle scattering models. Fe:AOT samples in toluene also aggregate into smaller domains when compared to Fe:DBSo. Yet, all structures in toluene contain multiple metal cores. For Fe:AOT, only about five iron cores are found to complex together out of the eight that were included in the simulation box. The dimensions of the metallic complex aggregates are estimated by measuring the distances between sulfurs to estimate the core size. The distance between the oxygen atom that is bonded to the first carbon in

the sulfate group and the last carbon of the tail provides an estimate of the 'shell' thickness. These dimensions are measured on different snapshots obtained from the last 50 ns of the simulation and at least 10 measurements of each geometric parameter are averaged. The values are found in Table 7.2, where the uncertainty reported in the table corresponds to one standard deviation from the average. The dimensions of the Fe:AOT aggregates are 11.3 ± 1.2 Å in the equatorial radius and 8.8 ± 2.6 Å for the polar radius of the core, the AOT shell thickness is 8.4 ± 1.2 Å. These sizes are consistent with the ones obtained from the SAS data fitting (Table 7.1) which yielded 11.7 and 6.4 Å for the equatorial and polar radii respectively and a shell thickness of 7.2-7.8 Å. In contrast, all iron cores in the simulation box (eight) form part of the same DBSo micelle. The dimensions from the simulations (Table 7.2) reveal an ellipsoid of 12.8 ± 1.5 Å equatorial radius and 4.9 ± 0.6 Å of polar radius and a shell thickness of 14.3 ± 1.6 Å. This again agrees well with the SAS data fitting which show dimension of 13.6 and 4.9 Å equatorial and polar radii with a shell thickness of 10.6-11.6 Å. Radial distribution functions for Fe ions in the simulations are also provided in the supplemental information. The surfactant tails in the MD simulation are extended for both Fe:DBSo and Fe:AOT complexes in toluene. This shows that the differences in shell thickness obtained from SAS fits is directly related to the size of the surfactant tail.

A fully dispersed conformation is observed in the simulation of Fe:AOT complexes in methanol (Figure 7.4d). The metal cores are isolated from each other but still form stable complexes and the structures are similar to that suggested from the SAS models. In contrast, Fe:DBSo (Figure 7.4c) complexes show aggregates of 2 to 3 ion cores in the MD simulation but appear to be dispersed according to SAS experiments of dilute samples (50 mM). DBSo tails also tend to interact with each other forming a sheet-like structure with intercalated DBSo tails in the

simulation. Additional experiments with a concentration of 300 mM rather than 50 mM showed that higher concentrations of the complexes tend to cause the formation of larger structures among the hydrocarbon tails. The dimension of the MD aggregates once again matches the SAXS results. Fe:AOT simulations reveal single complex with a core radius if 3.3 ± 0.1 Å while the radius of the features found on the SAXS are 3.4 Å. The consistently larger radii of DBSo aggregates, when compared to AOT, observed in the scattering of the metallic complexes dispersed in methanol suggest that small multi-ion complexes may form for all metal:DBSo complexes.

**Table 7.2:** Dimension of the different structures formed by the metallic complexes obtained from the molecular dynamic simulations. The parameters are named the same way as in the small angle scattering models in order to obtain a direct comparison.

| **Metal:Surfactant Solvent** | **Fe:DBSo Toluene** | **Fe:DBSo Methanol** | **Fe:AOT Toluene** | **Fe:AOT Methanol** |
|---|---|---|---|---|
| **Core/Shell Model** | Ellipsoid | Sphere | Ellipsoid | Sphere |
| **Equatorial Radius Core (Å)** | 12.8 ± 1.5 | --- | 11.3 ± 1.2 | --- |
| **Polar Radius Core (Å)** | 4.9 ± 0.6 | --- | 8.8 ± 2.6 | --- |
| **Equatorial Thickness (Å)** | 14.3 ± 1.6 | --- | 8.4 ± 1.2 | --- |
| **Radius (Å)** | --- | 3.2 ± 0.6 | --- | 3.3 ± 0.1 |
| **Radius (Two Iron Complex) (Å)** | --- | 5.6 ± 1.2 | --- | --- |
| **Thickness (Å)** | --- | 15.6 ± 0.6 | --- | 8.2 ± 1.1 |

Finally, the MD trajectories were analyzed using Crysol and Cryson.[284, 285] These programs use the trajectory to simulate small angle scattering. The data are presented in Figure 7.5. We selected 10 frames from each simulation and cut out the dominant complex structure. The scattering curves generally agree well with the data presented in Figures 7.2 and 7.3. What is important in these curves is the shape and position of the knee rather than the magnitude of the curve.

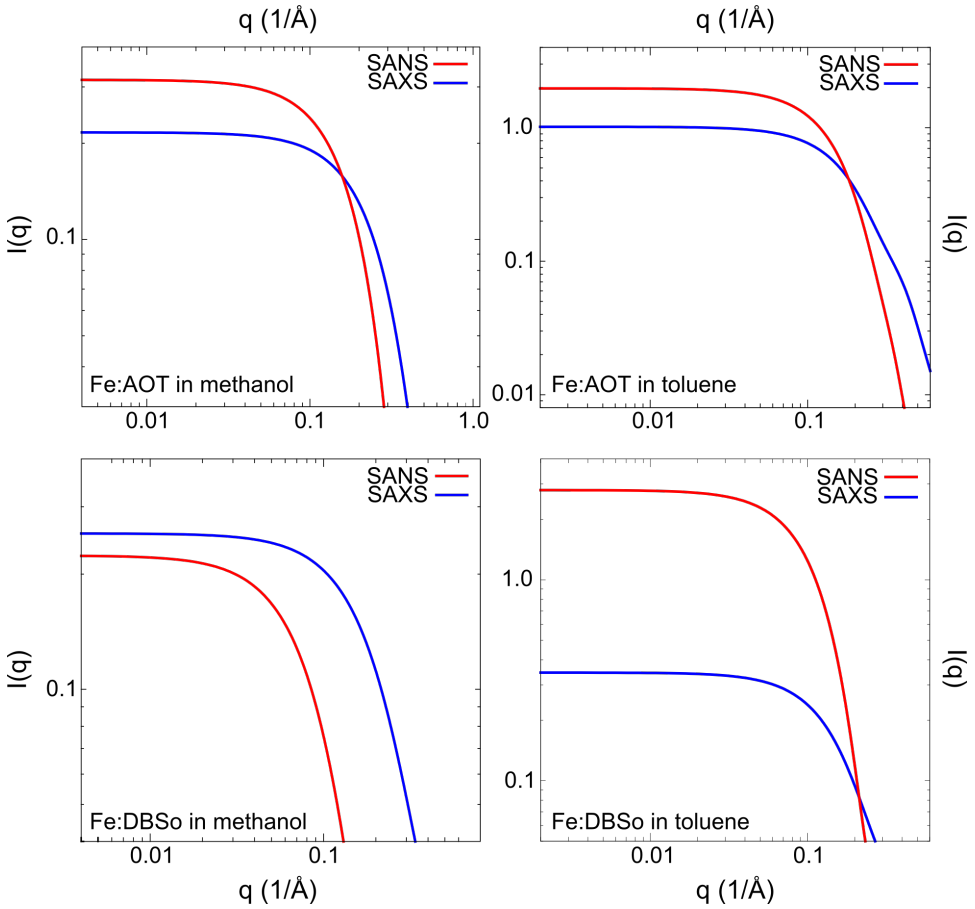



**Figure 7.5:** Crysol and Cryson scattering curves for each of the FeIII systems simulated.

## Conclusions

This set of simulations and analysis demonstrates the ability of our MD techniques to simulate the aggregation of charged molecules in organic solvents. We employed the same parameterization techniques for the surfactants and the organic solvents as we did for the ionic liquids in the previous chapters. The ability of these parameterization of simulation techniques to model the formation of large complexes gives additional credibility to our techniques.

## Contributions of this Research

When we began this research, there were very few studies of the interactions of ILs with enzymes. Those studies that were previously published were of much shorter timescales (below 10 ns), and the conclusions that they reached about IL-protein interactions did not give us large enough ensembles to make any conclusions about their nature. With our longer simulation times and the application of enhanced sampling techniques, we can now make experimentally verifiable predictions about the interactions of ILs and nearly any other molecule of interest. My first IL-protein study about GH11 shows that some of our predictions (about competitive inhibition) are consistent with later experimental evidence produced by other groups. My second IL-protein study teaches us that we can explore timescales that are long enough to show large changes in secondary structure as with the cellulase from TMA. We are also able to show that the most IL-tolerant enzyme, PHO, maintains water-like character in high concentrations of IL. The third IL-protein study shows that we might be able to predict *a priori* those areas of the protein that have different dynamics in IL-containing solvents. Additionally, we can characterize the free energy of important structural transitions of the protein in these exotic solvents by using metadynamics. In conjunction with other studies published by members of our group, our work has shown that our simulation techniques are viable meaningfully describing biomolecules' structure and dynamics in non-aqueous and mixed solvents.

The protocol that we established in Chapter I has been used extensively in our group to parameterize all sorts of small solvent molecules. The techniques we describe in that paper have been automated to a great extent. When a new project presents itself, our automated system allows us begin a wide variety of simulations within a matter of minutes. Our simulations have shown that these methods can be useful to predict the properties of a wide range of ILs, and that

the calculations are quick and accurate for a few of the important properties tested. We have independently arrived at a similar solution to Prof. Maginn's group for the problem of making IL force fields quickly.

In Chapter V, I described a method by which we can predict the preferred conformations of protein aggregates relatively quickly and accurately. Many different systems can be characterized as the interaction of several biomolecules. These methods should be generalizable to study many of those interactions. Since the publication of this work, other members of the group have begun atomistic simulations of the same system. As was stated in Chapter V, the atomistic simulations are about forty times slower than the coarse-grained ones. So, even with improved computing power from a new generation of processors and the removal of most of the water molecules, the atomistic simulations of peptide aggregation are still at the limits of our abilities.

Chapters VI and VII demonstrate that our method of creating force fields for charged molecules can be applied to other atomistic systems. In this case, the phosphate, silicic acid, and surfactant molecules were parameterized using similar methods to those we use for ILs. We were able to analyze the interactions of these charged molecules with positively charged lysine sidechains and multivalent metal ions. Our simulation results agree with many of the observations from sum frequency generation spectroscopy and small angle scattering. Taken together, these two chapters demonstrate that our simulations can probe the formation of complex structures in highly charged environments.

# Bibliography


1. Bara, J.E., et al., *Synthesis and performance of polymerizable room-temperature ionic liquids as gas separation membranes*. Industrial & Engineering Chemistry Research, 2007. **46**(16): p. 5397-5404.
2. Karadas, F., M. Atilhan, and S. Aparicio, *Review on the use of ionic liquids (ils) as alternative fluids for co2 capture and natural gas sweetening*. Energy & Fuels, 2010. **24**: p. 5817-5828.
3. Kamimura, A. and S. Yamamoto, *An efficient method to depolymerize polyamide plastics: A new use of ionic liquids*. Organic Letters, 2007. **9**(13): p. 2533-2535.
4. Kim, Y.H., et al., *Ionic liquid-mediated extraction of lipids from algal biomass*. Bioresource Technology, 2012. **109**: p. 312-315.
5. Armand, M., et al., *Ionic-liquid materials for the electrochemical challenges of the future*. Nature Materials, 2009. **8**(8): p. 621-629.
6. Bridges, N.J., A.E. Visser, and E.B. Fox, *Potential of nanoparticle-enhanced ionic liquids (neils) as advanced heat-transfer fluids*. Energy & Fuels, 2011. **25**(10): p. 4862-4864.
7. Li, H., et al., *Thermostabilization of extremophilic dictyoglomus thermophilum gh11 xylanase by an n-terminal disulfide bridge and the effect of ionic liquid [emim]oac on the enzymatic performance (vol 53, pg 414, 2013)*. Enzyme and Microbial Technology, 2014. **55**: p. 170-170.
8. Keskin, S., et al., *A review of ionic liquids towards supercritical fluid applications*. The Journal of Supercritical Fluids, 2007. **43**(1): p. 150-180.
9. Mora-Pale, M., et al., *Room temperature ionic liquids as emerging solvents for the pretreatment of lignocellulosic biomass*. Biotechnology and Bioengineering, 2011. **108**(6): p. 1229-1245.
10. Jorgensen, W.L., D.S. Maxwell, and J. TiradoRives, *Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids*. Journal of the American Chemical Society, 1996. **118**(45): p. 11225-11236.
11. Vanommeslaeghe, K., et al., *Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields*. Journal of Computational Chemistry, 2010. **31**(4): p. 671-690.
12. Rappe, A.K., et al., *Uff, a full periodic-table force-field for molecular mechanics and molecular-dynamics simulations*. Journal of the American Chemical Society, 1992. **114**(25): p. 10024-10035.
13. Wang, J.W., R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., *Development and testing of a general amber force field*. Journal of Computational Chemistry, 2004. **25**: p. 1157-1174.
14. Zhang, Y. and E.J. Maginn, *A simple aimd approach to derive atomic charges for condensed phase simulation of ionic liquids*. The Journal of Physical Chemistry B, 2012. **116**(33): p. 10036-10048.
15. Men, S., K.R.J. Lovelock, and P. Licence, *X-ray photoelectron spectroscopy of pyrrolidinium-based ionic liquids: Cation-anion interactions and a comparison to imidazolium-based analogues*. Physical Chemistry Chemical Physics, 2011. **13**(33): p. 15244-15255.
16. Hurisso, B.B., K.R.J. Lovelock, and P. Licence, *Amino acid-based ionic liquids: Using xps to probe the electronic environment via binding energies*. Physical Chemistry Chemical Physics, 2011. **13**(39): p. 17737-17748.
17. Youngs, T.G.A. and C. Hardacre, *Application of static charge transfer within an ionic-liquid force field and its effect on structure and dynamics*. Chemphyschem, 2008. **9**(11): p. 1548-1558.
18. Jaeger, V.W. and J. Pfaendtner, *Structure, dynamics, and activity of xylanase solvated in binary mixtures of ionic liquid and water*. Acs Chemical Biology, 2013. **8**(6): p. 1179-1186.
19. Jarin, Z. and J. Pfaendtner, *Ionic liquids can selectively change the conformational free-energy landscape of sugar rings*. Journal of Chemical Theory and Computation, 2014. **10**(2): p. 507-510.
20. Burney, P.R. and J. Pfaendtner, *Structural and dynamic features of candida rugosa lipase 1 in water, octane, toluene, and ionic liquids bmim-pf6 and bmim-no3*. The Journal of Physical Chemistry B, 2013. **117**(9): p. 2662-2670.
21. Nordwald, E.M. and J.L. Kaar, *Mediating electrostatic binding of 1-butyl-3-methylimidazolium chloride to enzyme surfaces improves conformational stability*. The Journal of Physical Chemistry B, 2013. **117**(30): p. 8977-8986.
22. Fiebig, O.C., et al., *Quantitative evaluation of myoglobin unfolding in the presence of guanidinium hydrochloride and ionic liquids in solution*. The Journal of Physical Chemistry B, 2013. **118**(2): p. 406-412.

23. Nordwald, E.M., G.S. Armstrong, and J.L. Kaar, *Nmr-guided rational engineering of an ionic-liquid-tolerant lipase*. ACS Catalysis, 2014. **4**(11): p. 4057-4064.
24. Yu, C.-Y., et al., *Using ionic liquid in a biphasic system to improve asymmetric hydrolysis of styrene oxide catalyzed by cross-linked enzyme aggregates (cleas) of mung bean epoxide hydrolases*. Industrial & Engineering Chemistry Research, 2014. **53**(19): p. 7923-7930.
25. Frisch, M.J., et al., *Gaussian 09, revision b.01*. 2009: Wallingford CT.
26. Wang, J.W., W.; Kollman, P. A.; Case, D. A., *Automatic atom type and bond type perception in molecular mechanical calculations*. Journal of Molecular Graphics and Modelling, 2006. **25**.
27. Cornell, W.D., et al., *Application of resp charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation*. Journal of the American Chemical Society, 1993. **115**(21): p. 9620-9631.
28. Heaps, C.G., *Long-range energy alternatives planning (leap) system [software version 2014.0.1.20]*. 2012, Stockholm Environment Institute.
29. Sousa da Silva, A.W.V., W. F., *Acpype - antechamber python parser interace*. BMC Research Notes, 2012. **367**(5).
30. Tenney, C.M., et al., *A computational and experimental study of the heat transfer properties of nine different ionic liquids*. Journal of Chemical & Engineering Data, 2014. **59**(2): p. 391-399.
31. Martínez, L.A., R.; Birgin, E. G.; Martínez, J. M., *Packmol: A package for building initial configurations for molecular dynamics simulations*. Journal of Computational Chemistry, 2009. **30**(13): p. 2157-2164.
32. Berendsen, H.J.C.v.d.S., D.; van Drunen, R., *Gromacs: A message-passing parallel molecular dynamics implementation*. Comp. Phys. Comm. , 1995. **91**: p. 43-56.
33. Lindahl, E.H., B.; van der Spoel, D., *Gromacs 3.0: A package for molecular simulation and trajectory analysis*. J. Mol. Mod. , 2001. **7**: p. 306-317.
34. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. The Journal of Chemical Physics, 2007. **126**(1): p. -.
35. Berendsen, H.J.C., et al., *Molecular dynamics with coupling to an external bath*. The Journal of Chemical Physics, 1984. **81**(8): p. 3684.
36. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*. Journal of Applied Physics, 1981. **52**(12): p. 7182-7190.
37. Gosling, E.M., I.R. McDonald, and K. Singer, *On the calculation by molecular dynamics of the shear viscosity of a simple fluid*. Molecular Physics, 1973. **26**(6): p. 1475-1484.
38. Vega, C. and J.L.F. Abascal, *Simulating water with rigid non-polarizable models: A general perspective*. Physical Chemistry Chemical Physics, 2011. **13**(44): p. 19663-19688.
39. Greaves, T.L., et al., *Protic ionic liquids: Solvents with tunable phase behavior and physicochemical properties*. The Journal of Physical Chemistry B, 2006. **110**(45): p. 22479-22487.
40. Almeida, H.F.D., et al., *Thermophysical properties of five acetate-based ionic liquids*. Journal of Chemical & Engineering Data, 2012. **57**(11): p. 3005-3013.
41. Araújo, J.M.M., et al., *Nucleic acid bases in 1-alkyl-3-methylimidazolium acetate ionic liquids: A thermophysical and ionic conductivity analysis*. The Journal of Chemical Thermodynamics, 2013. **57**(0): p. 1-8.
42. Shiflett, M.B., et al., *Solubility and diffusivity of difluoromethane in room-temperature ionic liquids*. Journal of Chemical & Engineering Data, 2006. **51**(2): p. 483-495.
43. Pereiro, A.B., et al., *Physical properties of 1-butyl-3-methylimidazolium methyl sulfate as a function of temperature*. Journal of Chemical & Engineering Data, 2007. **52**(2): p. 377-380.
44. Singh, T. and A. Kumar, *Temperature dependence of physical properties of imidazolium based ionic liquids: Internal pressure and molar refraction*. Journal of Solution Chemistry, 2009. **38**(8): p. 1043-1053.
45. García-Miaja, G., J. Troncoso, and L. Romaní, *Excess properties for binary systems ionic liquid + ethanol: Experimental results and theoretical description using the eras model*. Fluid Phase Equilibria, 2008. **274**(1–2): p. 59-67.
46. Domańska, U., A. Pobudkowska, and A. Wiśniewska, *Solubility and excess molar properties of 1,3-dimethylimidazolium methylsulfate, or 1-butyl-3-methylimidazolium methylsulfate, or 1-butyl-3-methylimidazolium octylsulfate ionic liquids with n-alkanes and alcohols: Analysis in terms of the pfp and fbt models*. Journal of Solution Chemistry, 2006. **35**(3): p. 311-334.
47. Iglesias-Otero, M.A., et al., *Density and refractive index in mixtures of ionic liquids and organic solvents: Correlations and predictions*. The Journal of Chemical Thermodynamics, 2008. **40**(6): p. 949-956.

48. Harris, K.R., M. Kanakubo, and L.A. Woolf, *Temperature and pressure dependence of the viscosity of the ionic liquids 1-hexyl-3-methylimidazolium hexafluorophosphate and 1-butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide*. Journal of Chemical & Engineering Data, 2007. **52**(3): p. 1080-1085.
49. Nieto de Castro, C.A., et al., *Studies on the density, heat capacity, surface tension and infinite dilution diffusion with the ionic liquids [$c_4$mim][$ntf_2$], [$c_4$mim][dca], [$c_2$mim][$etoso_3$] and [aliquat][dca]*. Fluid Phase Equilibria, 2010. **294**(1–2): p. 157-179.
50. Krummen, M., P. Wasserscheid, and J. Gmehling, *Measurement of activity coefficients at infinite dilution in ionic liquids using the dilutor technique*. Journal of Chemical & Engineering Data, 2002. **47**(6): p. 1411-1417.
51. Vranes, M., et al., *Physicochemical characterization of 1-butyl-3-methylimidazolium and 1-butyl-1-methylpyrrolidinium bis(trifluoromethylsulfonyl)imide*. Journal of Chemical & Engineering Data, 2012. **57**(4): p. 1072-1077.
52. Troncoso, J., et al., *Thermodynamic properties of imidazolium-based ionic liquids: Densities, heat capacities, and enthalpies of fusion of [bmim][$pf_6$] and [bmim][$ntf_2$]*. Journal of Chemical & Engineering Data, 2006. **51**(5): p. 1856-1859.
53. Freire, M.G., et al., *Thermophysical characterization of ionic liquids able to dissolve biomass*. Journal of Chemical & Engineering Data, 2011. **56**(12): p. 4813-4822.
54. Fröba, A.P., et al., *Thermal conductivity of ionic liquids: Measurement and prediction*. International Journal of Thermophysics, 2010. **31**(11-12): p. 2059-2077.
55. Rosenboom, J.-G., W. Afzal, and J.M. Prausnitz, *Solubilities of some organic solutes in 1-ethyl-3-methylimidazolium acetate. Chromatographic measurements and predictions from cosmo-rs*. The Journal of Chemical Thermodynamics, 2012. **47**(0): p. 320-327.
56. Schmidt, H., et al., *Experimental study of the density and viscosity of 1-ethyl-3-methylimidazolium ethyl sulfate*. The Journal of Chemical Thermodynamics, 2012. **47**(0): p. 68-75.
57. Gaciño, F.M., et al., *Influence of molecular structure on densities and viscosities of several ionic liquids*. Journal of Chemical & Engineering Data, 2011. **56**(12): p. 4984-4999.
58. Seki, S., et al., *Comprehensive refractive index property for room-temperature ionic liquids*. Journal of Chemical & Engineering Data, 2012. **57**(8): p. 2211-2216.
59. Gómez, E., et al., *Physical properties of pure 1-ethyl-3-methylimidazolium ethylsulfate and its binary mixtures with ethanol and water at several temperatures*. Journal of Chemical & Engineering Data, 2006. **51**(6): p. 2096-2102.
60. Vercher, E., et al., *Refractive indices and deviations in refractive indices of trifluoromethanesulfonate-based ionic liquids in water*. Journal of Chemical & Engineering Data, 2011. **56**(12): p. 4499-4504.
61. Olivier, E., et al., *Activity coefficients at infinite dilution of organic solutes in the ionic liquid 1-ethyl-3-methylimidazolium trifluoromethanesulfonate using gas–liquid chromatography at t = (313.15, 323.15, and 333.15) k*. The Journal of Chemical Thermodynamics, 2010. **42**(1): p. 78-83.
62. Dzida, M., et al., *Speed of sound and adiabatic compressibility of 1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide under pressures up to 100 mpa*. Journal of Chemical & Engineering Data, 2013. **58**(6): p. 1571-1576.
63. González, B., et al., *Synthesis and physical properties of 1-ethyl 3-methylpyridinium ethylsulfate and its binary mixtures with ethanol and water at several temperatures*. Journal of Chemical & Engineering Data, 2008. **53**(8): p. 1824-1828.
64. Calvar, N., et al., *Experimental determination, correlation, and prediction of physical properties of the ternary mixtures ethanol and 1-propanol + water + 1-ethyl-3-methylpyridinium ethylsulfate at 298.15 k*. Journal of Chemical & Engineering Data, 2009. **54**(8): p. 2229-2234.
65. González, E.J., et al., *(liquid+liquid) equilibria for ternary mixtures of (alkane+benzene+[empy] [eso4]) at several temperatures and atmospheric pressure*. The Journal of Chemical Thermodynamics, 2009. **41**(11): p. 1215-1221.
66. González, B., S. Corderí, and A.G. Santamaría, *Application of 1-alkyl-3-methylpyridinium bis(trifluoromethylsulfonyl)imide ionic liquids for the ethanol removal from its mixtures with alkanes*. The Journal of Chemical Thermodynamics, 2013. **60**(0): p. 9-14.
67. Gómez, E., et al., *Dynamic viscosities of a series of 1-alkyl-3-methylimidazolium chloride ionic liquids and their binary mixtures with water at several temperatures*. Journal of Chemical & Engineering Data, 2006. **51**(2): p. 696-701.

68. He, R.-H., et al., *Solubility of hydrogen chloride in three 1-alkyl-3-methylimidazolium chloride ionic liquids in the pressure range (0 to 100) kpa and temperature range (298.15 to 363.15) k*. Journal of Chemical & Engineering Data, 2012. **57**(11): p. 2936-2941.
69. Kenneth, R.S., S. Annegret, and T. María-José, *Viscosity and density of 1-alkyl-3-methylimidazolium ionic liquids*, in *Clean solvents*. 2002, American Chemical Society. p. 34-49.
70. Huddleston, J.G., et al., *Characterization and comparison of hydrophilic and hydrophobic room temperature ionic liquids incorporating the imidazolium cation*. Green Chemistry, 2001. **3**(4): p. 156-164.
71. Widegren, J.A. and J.W. Magee, *Density, viscosity, speed of sound, and electrolytic conductivity for the ionic liquid 1-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide and its mixtures with water*. Journal of Chemical & Engineering Data, 2007. **52**(6): p. 2331-2338.
72. Kato, R. and J. Gmehling, *Systems with ionic liquids: Measurement of vle and $\gamma \infty$ data and prediction of their thermodynamic behavior using original unifac, mod. Unifac(do) and cosmo-rs(ol)*. The Journal of Chemical Thermodynamics, 2005. **37**(6): p. 603-619.
73. Muhammad, A., et al., *Thermophysical properties of 1-hexyl-3-methyl imidazolium based ionic liquids with tetrafluoroborate, hexafluorophosphate and bis(trifluoromethylsulfonyl)imide anions*. The Journal of Chemical Thermodynamics, 2008. **40**(9): p. 1433-1438.
74. McHale, G., et al., *Density−viscosity product of small-volume ionic liquid samples using quartz crystal impedance analysis*. Analytical Chemistry, 2008. **80**(15): p. 5806-5811.
75. Oliveira, F.S., et al., *Structural and positional isomerism influence in the physical properties of pyridinium $ntf_2$-based ionic liquids: Pure and water-saturated mixtures*. Journal of Chemical & Engineering Data, 2010. **55**(10): p. 4514-4520.
76. Liu, Q.-S., et al., *Physicochemical properties of ionic liquids $[c_3py][ntf_2]$ and $[c_6py][ntf_2]$*. Journal of Chemical & Engineering Data, 2011. **56**(11): p. 4094-4101.
77. Arce, A., O. Rodríguez, and A. Soto, *Experimental determination of liquid−liquid equilibrium using ionic liquids: Tert-amyl ethyl ether + ethanol + 1-octyl-3-methylimidazolium chloride system at 298.15 k*. Journal of Chemical & Engineering Data, 2004. **49**(3): p. 514-517.
78. David, W., et al., *Activity coefficients of hydrocarbon solutes at infinite dilution in the ionic liquid, 1-methyl-3-octyl-imidazolium chloride from gas–liquid chromatography*. The Journal of Chemical Thermodynamics, 2003. **35**(8): p. 1335-1341.
79. Nebig, S. and J. Gmehling, *Measurements of different thermodynamic properties of systems containing ionic liquids and correlation of these properties using modified unifac (dortmund)*. Fluid Phase Equilibria, 2010. **294**(1–2): p. 206-212.
80. Papaiconomou, N., et al., *Synthesis and properties of seven ionic liquids containing 1-methyl-3-octylimidazolium or 1-butyl-4-methylpyridinium cations*. Journal of Chemical & Engineering Data, 2006. **51**(4): p. 1389-1393.
81. Alonso, L., et al., *(liquid + liquid) equilibria of [c8mim][ntf2] ionic liquid with a sulfur-component and hydrocarbons*. The Journal of Chemical Thermodynamics, 2008. **40**(2): p. 265-270.
82. Tokuda, H., et al., *How ionic are room-temperature ionic liquids? An indicator of the physicochemical properties*. The Journal of Physical Chemistry B, 2006. **110**(39): p. 19593-19600.
83. Liu, Q.-S., et al., *Density and surface tension of ionic liquids $[c_npy][ntf_2]$ (n = 2, 4, 5)*. Journal of Chemical & Engineering Data, 2010. **55**(11): p. 4928-4930.
84. Sambasivarao, S.V. and O. Acevedo, *Development of opls-aa force field parameters for 68 unique ionic liquids*. Journal of Chemical Theory and Computation, 2009. **5**(4): p. 1038-1050.
85. Damm, W., et al., *Opls all-atom force field for carbohydrates*. Journal of Computational Chemistry, 1997. **18**(16): p. 1955-1970.
86. Strechan, A.A., et al., *Low-temperature heat capacity of hydrophilic ionic liquids [bmim][$cf_3coo$] and [bmim][$ch_3coo$] and a correlation scheme for estimation of heat capacity of ionic liquids*. The Journal of Chemical Thermodynamics, 2008. **40**(4): p. 632-639.
87. Rocha, M.A.A., et al., *Heat capacities at 298.15 k of the extended $[c_nc_1im][ntf_2]$ ionic liquid series*. The Journal of Chemical Thermodynamics, 2012. **53**(0): p. 140-143.
88. Shimizu, Y., et al., *Effects of thermal history on thermal anomaly in solid of ionic liquid compound, $[c_4mim][tf_2n]$*. Chemistry Letters, 2007. **36**(12): p. 1484-1485.
89. Zhang, Z.-H., et al., *Thermodynamic investigation of room temperature ionic liquid: The heat capacity and standard enthalpy of formation of emies*. Thermochimica Acta, 2006. **447**(2): p. 141-146.

90. Ge, R., et al., *Heat capacities of ionic liquids as a function of temperature at 0.1 mpa. Measurement and prediction*. Journal of Chemical & Engineering Data, 2008. **53**(9): p. 2148-2153.
91. García-Miaja, G., J. Troncoso, and L. Romaní, *Excess enthalpy, density, and heat capacity for binary systems of alkylimidazolium-based ionic liquids+water*. The Journal of Chemical Thermodynamics, 2009. **41**(2): p. 161-166.
92. Paulechka, Y.U., et al., *Thermodynamic properties and polymorphism of 1-alkyl-3-methylimidazolium bis(triflamides)*. The Journal of Chemical Thermodynamics, 2007. **39**(6): p. 866-877.
93. Waliszewski, D., et al., *Heat capacities of ionic liquids and their heats of solution in molecular liquids*. Thermochimica Acta, 2005. **433**(1–2): p. 149-152.
94. Crosthwaite, J.M., et al., *Phase transition and decomposition temperatures, heat capacities and viscosities of pyridinium ionic liquids*. The Journal of Chemical Thermodynamics, 2005. **37**(6): p. 559-568.
95. Calvar, N., et al., *Thermal analysis and heat capacities of pyridinium and imidazolium ionic liquids*. Thermochimica Acta, 2013. **565**(0): p. 178-182.
96. Verevkin, S.P., et al., *Does alkyl chain length really matter? Structure–property relationships in thermochemistry of ionic liquids*. Thermochimica Acta, 2013. **562**(0): p. 84-95.
97. Blokhin, A.V., Y.U. Paulechka, and G.J. Kabo, *Thermodynamic properties of [$c_6$mim][$ntf_2$] in the condensed state*. Journal of Chemical & Engineering Data, 2006. **51**(4): p. 1377-1388.
98. Shimizu, Y., et al., *Low-temperature heat capacity of room-temperature ionic liquid, 1-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide*. The Journal of Physical Chemistry B, 2006. **110**(28): p. 13970-13975.
99. Maginn, E.J., *Atomistic simulation of the thermodynamic and transport properties of ionic liquids*. Accounts of Chemical Research, 2007. **40**(11): p. 1200-1207.
100. Cadena, C., et al., *Molecular modeling and experimental studies of the thermodynamic and transport properties of pyridinium-based ionic liquids*. The Journal of Physical Chemistry B, 2006. **110**(6): p. 2821-2832.
101. Earle, M.J., et al., *The distillation and volatility of ionic liquids*. Nature, 2006. **439**(7078): p. 831-834.
102. Santos, L.M.N.B.F., et al., *Ionic liquids: First direct determination of their cohesive energy*. Journal of the American Chemical Society, 2006. **129**(2): p. 284-285.
103. Kelkar, M.S. and E.J. Maginn, *Calculating the enthalpy of vaporization for ionic liquid clusters*. The Journal of Physical Chemistry B, 2007. **111**(32): p. 9424-9427.
104. Armstrong, J.P., et al., *Vapourisation of ionic liquids*. Physical Chemistry Chemical Physics, 2007. **9**(8): p. 982-990.
105. Zaitsau, D.H., et al., *Experimental vapor pressures of 1-alkyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imides and a correlation scheme for estimation of vaporization enthalpies of ionic liquids*. The Journal of Physical Chemistry A, 2006. **110**(22): p. 7303-7306.
106. Zaitsau, D., et al., *Structure-property relationships in ils: A study of the alkyl chain length dependence in vaporisation enthalpies of pyridinium based ionic liquids*. Science China Chemistry, 2012. **55**(8): p. 1525-1531.
107. Tokuda, H., et al., *Physicochemical properties and structures of room temperature ionic liquids. 2. Variation of alkyl chain length in imidazolium cation*. The Journal of Physical Chemistry B, 2005. **109**(13): p. 6103-6110.
108. Hess, B., *Determining the shear viscosity of model liquids from molecular dynamics simulations*. The Journal of Chemical Physics, 2002. **116**(1): p. 209-217.
109. Hu, Z. and C.J. Margulis, *On the response of an ionic liquid to external perturbations and the calculation of shear viscosity*. The Journal of Physical Chemistry B, 2007. **111**(18): p. 4705-4714.
110. Fendt, S., et al., *Viscosities of acetate or chloride-based ionic liquids and some of their mixtures with water or other common solvents*. Journal of Chemical & Engineering Data, 2010. **56**(1): p. 31-34.
111. Andreatta, A.E., et al., *Isobaric vapour–liquid equilibria and physical properties for isopropyl acetate + isopropanol + 1-butyl-3-methyl-imidazolium bis(trifluoromethylsulfonyl)imide mixtures*. Fluid Phase Equilibria, 2011. **300**(1–2): p. 162-171.
112. Quijada-Maldonado, E., et al., *Experimental densities, dynamic viscosities and surface tensions of the ionic liquids series 1-ethyl-3-methylimidazolium acetate and dicyanamide and their binary and ternary mixtures with water and ethanol at t = (298.15 to 343.15 k)*. The Journal of Chemical Thermodynamics, 2012. **51**(0): p. 51-58.
113. Arce, A., H. Rodriguez, and A. Soto, *Use of a green and cheap ionic liquid to purify gasoline octane boosters*. Green Chemistry, 2007. **9**(3): p. 247-253.

114. Calvar, N., et al., *Experimental determination, correlation, and prediction of physical properties of the ternary mixtures ethanol + water with 1-octyl-3-methylimidazolium chloride and 1-ethyl-3-methylimidazolium ethylsulfate*. Journal of Chemical & Engineering Data, 2007. **52**(6): p. 2529-2535.
115. González, B., et al., *Density and viscosity experimental data of the ternary mixtures 1-propanol or 2-propanol + water + 1-ethyl-3-methylimidazolium ethylsulfate. Correlation and prediction of physical properties of the ternary systems*. Journal of Chemical & Engineering Data, 2008. **53**(3): p. 881-887.
116. Rodríguez, H. and J.F. Brennecke, *Temperature and composition dependence of the density and viscosity of binary mixtures of water + ionic liquid*. Journal of Chemical & Engineering Data, 2006. **51**(6): p. 2145-2155.
117. Schreiner, C., et al., *Fractional walden rule for ionic liquids: Examples from recent measurements and a critique of the so-called ideal kcl line for the walden plot*. Journal of Chemical & Engineering Data, 2009. **55**(5): p. 1784-1788.
118. Yao, H., et al., *Densities and viscosities of the binary mixtures of 1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide with n-methyl-2-pyrrolidone or ethanol at t = (293.15 to 323.15) k*. Journal of Chemical & Engineering Data, 2012. **57**(3): p. 875-881.
119. Kandil, M.E., K.N. Marsh, and A.R.H. Goodwin, *Measurement of the viscosity, density, and electrical conductivity of 1-hexyl-3-methylimidazolium bis(trifluorosulfonyl)imide at temperatures between (288 and 433) k and pressures below 50 mpa*. Journal of Chemical & Engineering Data, 2007. **52**(6): p. 2382-2387.
120. Ahosseini, A. and A. Scurto, *Viscosity of imidazolium-based ionic liquids at elevated pressures: Cation and anion effects*. International Journal of Thermophysics, 2008. **29**(4): p. 1222-1243.
121. González, E.J., L. Alonso, and Á. Domínguez, *Physical properties of binary mixtures of the ionic liquid 1-methyl-3-octylimidazolium chloride with methanol, ethanol, and 1-propanol at t = (298.15, 313.15, and 328.15) k and at p = 0.1 mpa*. Journal of Chemical & Engineering Data, 2006. **51**(4): p. 1446-1452.
122. Alonso, L., et al., *Liquid−liquid equilibria for [$c_8$mim][$ntf_2$] + thiophene + 2,2,4-trimethylpentane or + toluene*. Journal of Chemical & Engineering Data, 2008. **53**(8): p. 1750-1755.
123. Andreatta, A.E., et al., *Physical properties of binary and ternary mixtures of ethyl acetate, ethanol, and 1-octyl-3-methyl-imidazolium bis(trifluoromethylsulfonyl)imide at 298.15 k*. Journal of Chemical & Engineering Data, 2009. **54**(3): p. 1022-1028.
124. Garrote, G., H. Domínguez, and J.C. Parajó, *Mild autohydrolysis: An environmentally friendly technology for xylooligosaccharide production from wood*. Journal of Chemical Technology & Biotechnology, 1999. **74**(11): p. 1101–1109.
125. Swatloski, R.P., et al., *Dissolution of cellose with ionic liquids*. Journal of the American Chemical Society, 2002. **124**(18): p. 4974–4975.
126. Fort, D.A., et al., *Can ionic liquids dissolve wood? Processing and analysis of lignocellulosic materials with 1-n-butyl-3-methylimidazolium chloride*. Green Chemistry, 2007. **9**(1): p. 63–69.
127. Li, C.L., et al., *Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification*. Bioresource Technology, 2010. **101**(13): p. 4900–4906.
128. Wang, Y., et al., *Compatible ionic liquid-cellulases system for hydrolysis of lignocellulosic biomass*. Biotechnology and Bioengineering, 2011. **108**(5): p. 1042–1048.
129. Klibanov, A.M., *Improving enzymes by using them in organic solvents*. Nature, 2001. **409**(6817): p. 241–246.
130. Datta, S., et al., *Ionic liquid tolerant hyperthermophilic cellulases for biomass pretreatment and hydrolysis*. Green Chemistry, 2010. **12**(2): p. 338–345.
131. Park, S. and R.J. Kazlauskas, *Biocatalysis in ionic liquids – advantages beyond green technology*. Current Opinion in Biotechnology, 2003. **14**(4): p. 432-437.
132. Baker, G.A. and W.T. Heller, *Small-angle neutron scattering studies of model protein denaturation in aqueous solutions of the ionic liquid 1-butyl-3-methylimidazolium chloride*. Chemical Engineering Journal, 2009. **147**(1): p. 6–12.
133. De Diego, T., et al., *Understanding structure - stability relationships of candida antartica lipase b in ionic liquids*. Biomacromolecules, 2005. **6**(3): p. 1457–1464.
134. Mostofian, B., J.C. Smith, and X. Cheng, *The solvation structures of cellulose microfibrils in ionic liquids*. Interdisciplinary Sciences: Computational Life Sciences, 2011. **3**(4): p. 308–320.
135. Gross, A.S., A.T. Bell, and J.-W. Chu, *Thermodynamics of cellulose solvation in water and the ionic liquid 1-butyl-3-methylimidazolim chloride*. The Journal of Physical Chemistry B, 2011. **115**(46): p. 13433–13440.

136. Gupta, K.M., Z. Hu, and J. Jiang, *Mechanistic understanding of interactions between cellulose and ionic liquids: A molecular simulation study*. Polymer, 2011. **52**(25): p. 5904–5911.
137. Micaelo, N.M. and C.M. Soares, *Protein structure and dynamics in ionic liquids. Insights from molecular dynamics simulation studies*. The Journal of Physical Chemistry B, 2008. **112**(9): p. 2566–2572.
138. Yon, J.M., D. Perahia, and C. Ghélis, *Conformational dynamics and enzyme activity*. Biochimie, 1998. **80**(1): p. 33–42.
139. Benkovic, S.J. and S. Hammes-Schiffer, *A perspective on enzyme catalysis*. Science, 2003. **301**(5637): p. 1196–1202.
140. Kamerlin, S.C.L. and A. Warshel, *At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis?* Proteins-Structure Function and Bioinformatics, 2010. **78**(6): p. 1339–1375.
141. Paës, G., et al., *New insights into the role of the thumb-like loop in gh-ii xylanases*. Protein Engineering Design & Selection, 2007. **20**(1): p. 15–23.
142. Pollet, A., et al., *Crystallographic and activity-based evidence for thumb flexibility and its relevance in glycoside hydrolase family 11 xylanases*. Proteins-Structure Function and Bioinformatics, 2009. **77**(2): p. 395–403.
143. Macleod, A.M., et al., *The acid/base catalyst in the exoglucanase/xylanase from cellulomonas-fimi is glutamic-acid-127 - evidence from detailed kinetic-studies of mutants*. Biochemistry, 1994. **33**(20): p. 6371–6376.
144. Turunen, O., et al., *A combination of weakly stabilizing mutations with a disulfide bridge in the œ±-helix region of trichoderma reesei endo-1,4-œ≤-xylanase ii increases the thermal stability through synergism*. Journal of Biotechnology, 2001. **88**(1): p. 37–46.
145. Hakulinen, N., et al., *Three-dimensional structures of thermophilic beta-1,4-xylanases from chaetomium thermophilum and nonomuraea flexuosa - comparison of twelve xylanases in relation to their thermal stability*. European Journal of Biochemistry, 2003. **270**(7): p. 1399–1412.
146. Purmonen, M., et al., *Molecular dynamics studies on the thermostability of family 11 xylanases*. Protein Engineering Design & Selection, 2007. **20**(11): p. 551–559.
147. Phillips, J.C., et al., *Scalable molecular dynamics with namd*. Journal of Computational Chemistry, 2005. **26**(16): p. 1781–1802.
148. Mahoney, M.W. and W.L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions*. Journal of Chemical Physics, 2000. **112**(20): p. 8910–8922.
149. Hornak, V., et al., *Comparison of multiple amber force fields and development of improved protein backbone parameters*. Proteins-Structure Function and Bioinformatics, 2006. **65**(3): p. 712–725.
150. Wang, J.M., et al., *Development and testing of a general amber force field*. Journal of Computational Chemistry, 2004. **25**(9): p. 1157–1174.
151. Bayly, C.I., et al., *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the resp model*. Journal of Physical Chemistry, 1993. **97**(40): p. 10269–10280.
152. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes*. Journal of Computational Physics, 1977. **23**(3): p. 327–341.
153. Sugahara, M., Y. Kageyama-Morikawa, and N. Kunishima, *Packing space expansion of protein crystallization screening with synthetic zeolite as a heteroepitaxic nucleant*. Crystal Growth & Design, 2010. **11**(1): p. 110–120.
154. Bernstein, F.C., et al., *Protein data bank - computer-based archival file for macromolecular structures*. Journal of Molecular Biology, 1977. **112**(3): p. 535–542.
155. Xu, F., Y.-C. Shi, and D. Wang, *Enhanced production of glucose and xylose with partial dissolution of corn stover in ionic liquid, 1-ethyl-3-methylimidazolium acetate*. Bioresource Technology, 2012. **114**(0): p. 720-724.
156. Quijada-Maldonado, E., et al., *Experimental densities, dynamic viscosities and surface tensions of the ionic liquids series 1-ethyl-3-methylimidazolium acetate and dicyanamide and their binary and ternary mixtures with water and ethanol at t = (298.15 to 343.15 k)*. Journal of Chemical Thermodynamics, 2012. **51**: p. 51–58.
157. Xiong, H., et al., *Engineering the thermostability of trichoderma reesei endo-1,4-ß-xylanase ii by combination of disulphide bridges*. Extremophiles, 2004. **8**(5): p. 393–400.

158. Bihari, M., T.P. Russell, and D.A. Hoagland, *Dissolution and dissolved state of cytochrome c in a neat, hydrophilic ionic liquid*. Biomacromolecules, 2010. **11**(11): p. 2944–2948.
159. Sate, D., et al., *Enzyme aggregation in ionic liquids studied by dynamic light scattering and small angle neutron scattering*. Green Chemistry, 2007. **9**(8): p. 859–867.
160. Ficke, L.E., H.c. Rodríguez, and J.F. Brennecke, *Heat capacities and excess enthalpies of 1-ethyl-3-methylimidazolium-based ionic liquids and water*. Journal of Chemical & Engineering Data, 2008. **53**(9): p. 2112-2119.
161. Menjoge, A., et al., *Influence of water on diffusion in imidazolium-based ionic liquids: A pulsed field gradient nmr study*. The Journal of Physical Chemistry B, 2009. **113**(18): p. 6353-6359.
162. White, A.D., et al., *Decoding nonspecific interactions from nature*. Chemical Science, 2012. **3**(12): p. 3488–3494.
163. Pollet, A., et al., *Mutagenesis and subsite mapping underpin the importance for substrate specificity of the aglycon subsites of glycoside hydrolase family 11 xylanases*. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2010. **1804**(4): p. 977–985.
164. Vieira, D.S., L. Degrève, and R.J. Ward, *Characterization of temperature dependent and substrate-binding cleft movements in bacillus circulans family 11 xylanase: A molecular dynamics investigation*. Biochimica Et Biophysica Acta-General Subjects, 2009. **1790**(10): p. 1301–1306.
165. Vieira, D.S. and R.J. Ward, *Conformation analysis of a surface loop that controls active site access in the gh11 xylanase a from bacillus subtilis*. Journal of molecular modeling, 2012. **18**(4): p. 1473–1479.
166. Henzler-Wildman, K.A., et al., *Intrinsic motions along an enzymatic reaction trajectory*. Nature, 2007. **450**(7171): p. 838–U13.
167. da Costa Sousa, L., et al., *'Cradle-to-grave' assessment of existing lignocellulose pretreatment technologies*. Current Opinion in Biotechnology, 2009. **20**(3): p. 339-347.
168. Várnai, A., M. Siika-aho, and L. Viikari, *Restriction of the enzymatic hydrolysis of steam-pretreated spruce by lignin and hemicellulose*. Enzyme and Microbial Technology, 2010. **46**(3-4): p. 185-193.
169. Nakagame, S., R.P. Chandra, and J.N. Saddler, *The effect of isolated lignins, obtained from a range of pretreated lignocellulosic substrates, on enzymatic hydrolysis*. Biotechnology and Bioengineering, 2010. **105**(5): p. 871-879.
170. Agbor, V.B., et al., *Biomass pretreatment: Fundamentals toward application*. Biotechnology Advances, 2011. **29**(6): p. 675-685.
171. Mok, W.S.L., M.J. Antal, and G. Varhegyi, *Productive and parasitic pathways in dilute acid-catalyzed hydrolysis of cellulose*. Industrial & Engineering Chemistry Research, 1992. **31**(1): p. 94-100.
172. Ramos, L.P., *The chemistry involved in the steam treatment of lignocellulosic materials*. Quimica Nova, 2003. **26**(6): p. 863-871.
173. Fort, D.A., et al., *Can ionic liquids dissolve wood? Processing and analysis of lignocellulosic materials with 1-n-butyl-3-methylimidazolium chloride*. Green Chemistry, 2007. **9**(1): p. 63.
174. Kilpeläinen, I., et al., *Dissolution of wood in ionic liquids*. Journal of Agricultural and Food Chemistry, 2007. **55**(22): p. 9142-9148.
175. Zhang, H., et al., *1-allyl-3-methylimidazolium chloride room temperature ionic liquid: A new and powerful nonderivatizing solvent for cellulose*. Macromolecules, 2005. **38**(20): p. 8272-8277.
176. Shill, K., et al., *Ionic liquid pretreatment of cellulosic biomass: Enzymatic hydrolysis and ionic liquid recycle*. Biotechnology and Bioengineering, 2011. **108**(3): p. 511-520.
177. Engel, P., et al., *Point by point analysis: How ionic liquid affects the enzymatic hydrolysis of native and modified cellulose*. Green Chemistry, 2010. **12**(11): p. 1959.
178. Swatloski, R.P., et al., *Dissolution of cellose with ionic liquids*. Journal of the American Chemical Society, 2002. **124**(18): p. 4974-4975.
179. Dadi, A.P., C.A. Schall, and S. Varanasi, *Mitigation of cellulose recalcitrance to enzymatic hydrolysis by ionic liquid pretreatment*. Applied Biochemistry and Biotechnology, 2007. **137**: p. 407-421.
180. Mansfield, S.D., C. Mooney, and J.N. Saddler, *Substrate and enzyme characteristics that limit cellulose hydrolysis*. Biotechnology Progress, 1999. **15**(5): p. 804-816.
181. Lozano, P., et al., *Stabilization of α-chymotrypsin by ionic liquids in transesterification reactions*. Biotechnology and Bioengineering, 2001. **75**(5): p. 563-569.
182. Lozano, P., et al., *Enzymatic ester synthesis in ionic liquids*. Journal of Molecular Catalysis B: Enzymatic, 2003. **21**(1-2): p. 9-13.

183. Ulbert, O., et al., *Thermal stability enhancement of candida rugosa lipase using ionic liquids*. Biocatalysis and Biotransformation, 2005. **23**(3-4): p. 177-183.
184. Lozano, P., et al., *Over-stabilization of candida antarctica lipase b by ionic liquids in ester synthesis*. Biotechnology Letters, 2001. **23**(18): p. 1529-1533.
185. Park, S. and R.J. Kazlauskas, *Improved preparation and use of room-temperature ionic liquids in lipase-catalyzed enantio- and regioselective acylations*. Journal of Organic Chemistry, 2001. **66**(25): p. 8395-8401.
186. Ulbert, O., et al., *Enhanced enantioselectivity of candida rugosa lipase in ionic liquids as compared to organic solvents*. Journal of Molecular Catalysis B: Enzymatic, 2004. **31**(1-3): p. 39-45.
187. Zhang, T., et al., *Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance*. Green Chemistry, 2011. **13**(8): p. 2083-2090.
188. Datta, S., et al., *Ionic liquid tolerant hyperthermophilic cellulases for biomass pretreatment and hydrolysis*. Green Chemistry, 2010. **12**(2): p. 338-345.
189. Roberts, E., et al., *Multiseq: Unifying sequence and structure data for evolutionary analysis*. Bmc Bioinformatics, 2006. **7**.
190. Vieille, C. and G.J. Zeikus, *Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability*. Microbiology and Molecular Biology Reviews, 2001. **65**(1): p. 1-43.
191. Shao, Q., *On the influence of hydrated imidazolium-based ionic liquid on protein structure stability: A molecular dynamics simulation study*. Journal of Chemical Physics, 2013. **139**(11): p. 115102.
192. Kim, H.S., et al., *The relationship between enhanced enzyme activity and structural dynamics in ionic liquids: A combined computational and experimental study*. Physical Chemistry Chemical Physics, 2014. **16**(7): p. 2944-2953.
193. Figueiredo, A.M., et al., *Protein destabilisation in ionic liquids: The role of preferential interactions in denaturation*. Physical Chemistry Chemical Physics, 2013. **15**(45): p. 19632-19643.
194. Haberler, M., C. Schroder, and O. Steinhauser, *Solvation studies of a zinc finger protein in hydrated ionic liquids*. Physical Chemistry Chemical Physics, 2011. **13**(15): p. 6924-6938.
195. Badieyan, S., D.R. Bevan, and C.M. Zhang, *Study and design of stability in gh5 cellulases*. Biotechnology and Bioengineering, 2012. **109**(1): p. 31-44.
196. Burney, P.R. and J. Pfaendtner, *Structural and dynamic features of candida rugosa lipase 1 in water, cctane, toluene, and ionic liquids bmim-pf6 and bmim-no3*. Journal of Physical Chemistry B, 2013. **117**(9): p. 2662-2670.
197. Li, H., et al., *Thermostabilization of extremophilic dictyoglomus thermophilum gh11 xylanase by an n-terminal disulfide bridge and the effect of ionic liquid [emim]oac on the enzymatic performance*. Enzyme and Microbial Technology, 2013. **53**(6-7): p. 414-419.
198. Hess, B., et al., *Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation*. Journal of Chemical Theory and Computation, 2008. **4**(3): p. 435–447.
199. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
200. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the amber ff99sb protein force field*. Proteins-Structure Function and Bioinformatics, 2010. **78**(8): p. 1950-1958.
201. Zhang, Y. and E.J. Maginn, *A simple aimd approach to derive atomic charges for condensed phase simulation of ionic liquids*. Journal of Physical Chemistry B, 2012. **116**(33): p. 10036-10048.
202. Martinez, L., et al., *Packmol: A package for building initial configurations for molecular dynamics simulations*. Journal of Computational Chemistry, 2009. **30**(13): p. 2157-2164.
203. Hess, B., et al., *Lincs: A linear constraint solver for molecular simulations*. Journal of Computational Chemistry, 1997. **18**(12): p. 1463-1472.
204. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. Journal of Chemical Physics, 2007. **126**(1).
205. Berendsen, H.J.C., et al., *Molecular-dynamics with coupling to an external bath*. Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.
206. Trivedi, N., et al., *Detection of ionic liquid stable cellulase produced by the marine bacterium pseudoalteromonas sp. Isolated from brown alga sargassum polycystum c. Agardh*. Bioresource Technology, 2013. **132**: p. 313-319.
207. Quijada-Maldonado, E., et al., *Experimental densities, dynamic viscosities and surface tensions of the ionic liquids series 1-ethyl-3-methylimidazolium acetate and dicyanamide and their binary and ternary mixtures*

*with water and ethanol at t = (298.15 to 343.15 k)*. Journal of Chemical Thermodynamics, 2012. **51**: p. 51-58.
208. Berman, H.M., et al., *The protein data bank*. Nucleic Acids Research, 2000. **28**(1): p. 235-242.
209. Lee, T.M., et al., *A structural study of hypocrea jecorina cel5a*. Protein Science, 2011. **20**(11): p. 1935-1940.
210. Li, X.H., et al., *Expression of trichoderma viride endoglucanase iii in the larvae of silkworm, bombyx mori l. And characteristic analysis of the recombinant protein*. Molecular Biology Reports, 2011. **38**(6): p. 3897-3902.
211. Wu, T.H., et al., *Diverse substrate recognition mechanism revealed by thermotoga maritima cel5a structures in complex with cellotetraose, cellobiose and mannotriose*. Biochimica Et Biophysica Acta-Proteins and Proteomics, 2011. **1814**(12): p. 1832-1840.
212. Kim, H.W. and K. Ishikawa, *Functional analysis of hyperthermophilic endocellulase from pyrococcus horikoshii by crystallographic snapshots*. Biochemical Journal, 2011. **437**: p. 223-230.
213. Olsson, M.H.M., et al., *Propka3: Consistent treatment of internal and surface residues in empirical pk(a) predictions*. Journal of Chemical Theory and Computation, 2011. **7**(2): p. 525-537.
214. Li, H., A.D. Robertson, and J.H. Jensen, *Very fast empirical prediction and rationalization of protein pk(a) values*. Proteins: Structure, Function, and Bioinformatics, 2005. **61**(4): p. 704-721.
215. Voss, N.R. and M. Gerstein, *Calculation of standard atomic volumes for rna and comparison with proteins: Rna is packed more tightly*. Journal of Molecular Biology, 2005. **346**(2): p. 477-92.
216. Kaar, J.L., et al., *Impact of ionic liquid physical properties on lipase activity and stability*. Journal of the American Chemical Society, 2003. **125**(14): p. 4125–4131.
217. Erbeldinger, M., A.J. Mesiano, and A.J. Russell, *Enzymatic catalysis of formation of z-aspartame in ionic liquid - an alternative to enzymatic catalysis in organic solvents*. Biotechnology Progress, 2000. **16**(6): p. 1129-1131.
218. Nordwald, E.M. and J.L. Kaar, *Stabilization of enzymes in ionic liquids via modification of enzyme charge*. Biotechnology and Bioengineering, 2013. **110**(9): p. 2352-2360.
219. Nordwald, E.M. and J.L. Kaar, *Mediating electrostatic binding of 1-butyl-3-methylimidazolium chloride to enzyme surfaces improves conformational stability*. Journal of Physical Chemistry B, 2013. **117**(30): p. 8977-8986.
220. Joosten, R.P., et al., *A series of pdb related databases for everyday needs*. Nucleic Acids Research, 2010. **39**(Database): p. D411-D419.
221. Weingartner, H., C. Cabrele, and C. Herrmann, *How ionic liquids can help to stabilize native proteins*. Physical Chemistry Chemical Physics, 2012. **14**(2): p. 415-426.
222. Lau, R.M., et al., *Dissolution of candida antarctica lipase b in ionic liquids: Effects on structure and activity*. Green Chemistry, 2004. **6**(9): p. 483-487.
223. Constatinescu, D., C. Herrmann, and H. Weingartner, *Patterns of protein unfolding and protein aggregation in ionic liquids*. Physical Chemistry Chemical Physics, 2010. **12**(8): p. 1756-1763.
224. Byrne, N., et al., *Reversible folding-unfolding, aggregation protection, and multi-year stabilization, in high concentration protein solutions, using ionic liquids*. Chemical Communications, 2007(26): p. 2714-2716.
225. Stark, A.C., C.T. Andrews, and A.H. Elcock, *Toward optimized potential functions for protein-protein interactions in aqueous solutions: Osmotic second virial coefficient calculations using the martini coarse-grained force field*. Journal of Chemical Theory and Computation, 2013. **9**(9): p. 4176-4185.
226. Gloria, D.E., K. Regina, and R.M. Douglas, *The development of ionic liquids for biomedical applications ? Prospects and challenges*, in *Ionic liquids: From knowledge to application*. 2009, American Chemical Society. p. 95-105.
227. Burney, P.R., et al., *Molecular dynamics investigation of the ionic liquid/enzyme interface: Application to engineering enzyme surface charge*. Proteins-Structure Function and Bioinformatics, 2015. **83**(4): p. 670-680.
228. Jaeger, V., P. Burney, and J. Pfaendtner, *Comparison of three ionic liquid-tolerant cellulases by molecular dynamics*. Biophysical Journal, 2015. **108**(4): p. 880-892.
229. Laio, A. and M. Parrinello, *Escaping free-energy minima*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(20): p. 12562-12566.
230. Zhang, Z.Y., et al., *A systematic methodology for defining coarse-grained sites in large biomolecules*. Biophysical Journal, 2008. **95**(11): p. 5073-5083.
231. Page, T.A., et al., *Dynamics of loop 1 of domain i in human serum albumin when dissolved in ionic liquids*. Journal of Physical Chemistry B, 2009. **113**(38): p. 12825-12830.

232. Akdogan, Y., M.J.N. Junk, and D. Hinderberger, *Effect of ionic liquids on the solution structure of human serum albumin*. Biomacromolecules, 2011. **12**(4): p. 1072-1079.
233. Wang, Y., et al., *Structural mechanism of ring-opening reaction of glucose by human serum albumin*. Journal of Biological Chemistry, 2013. **288**(22): p. 15980-15987.
234. Bhattacharya, A.A., T. Grune, and S. Curry, *Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin*. Journal of Molecular Biology, 2000. **303**(5): p. 721-732.
235. Hess, B., et al., *Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation*. Journal of Chemical Theory and Computation, 2008. **4**(3): p. 435-447.
236. Tribello, G.A., et al., *Plumed 2: New feathers for an old bird*. Computer Physics Communications, 2014. **185**(2): p. 604-613.
237. Wang J, et al., *Development and testing of a general amber forced field*. Journal of Computational Chemistry, 2005. **25**(9): p. 1157-1174.
238. Liu, Z.P., S.P. Huang, and W.C. Wang, *A refined force field for molecular simulation of imidazolium-based ionic liquids*. Journal of Physical Chemistry B, 2004. **108**(34): p. 12978-12989.
239. Barducci, A., G. Bussi, and M. Parrinello, *Well-tempered metadynamics: A smoothly converging and tunable free-energy method*. Physical Review Letters, 2008. **100**(2).
240. Deighan, M. and J. Pfaendtner, *Exhaustively sampling peptide adsorption with metadynamics*. Langmuir, 2013. **29**(25): p. 7999-8009.
241. Uusitalo, J.J., H.I. Ingolfsson, and S.J. Marrink, *Covering all the bases: A martini coarse-grained force field for DNA*. Biophysical Journal, 2013. **104**(2): p. 169A-169A.
242. Lopez, C.A., et al., *Martini coarse-grained force field: Extension to carbohydrates*. Journal of Chemical Theory and Computation, 2009. **5**(12): p. 3195-3210.
243. Monticelli, L., et al., *The martini coarse-grained force field: Extension to proteins*. Journal of Chemical Theory and Computation, 2008. **4**(5): p. 819-834.
244. Marrink, S.J., et al., *The martini force field: Coarse grained model for biomolecular simulations*. Journal of Physical Chemistry B, 2007. **111**(27): p. 7812-7824.
245. de Jong, D.H., et al., *Improved parameters for the martini coarse-grained protein force field*. Journal of Chemical Theory and Computation, 2013. **9**(1): p. 687-697.
246. Yesylevskyy, S.O., et al., *Polarizable water model for the coarse-grained martini force field*. Plos Computational Biology, 2010. **6**(6).
247. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure - pattern-recognition of hydrogen-bonded and geometrical features*. Biopolymers, 1983. **22**(12): p. 2577-2637.
248. Baio, J.E., et al., *Diatom mimics: Directing the formation of biosilica nanoparticles by controlled folding of lysine-leucine peptides*. Journal of the American Chemical Society, 2014. **136**(43): p. 15134-15137.
249. Degrado, W.F. and J.D. Lear, *Induction of peptide conformation at apolar water interfaces .1. A study with model peptides of defined hydrophobic periodicity*. Journal of the American Chemical Society, 1985. **107**(25): p. 7684-7689.
250. Emami, F.S., et al., *Prediction of specific biomolecule adsorption on silica surfaces as a function of ph and particle size*. Chemistry of Materials, 2014. **26**(19): p. 5725-5734.
251. Ji, N. and Y.R. Shen, *Sum frequency vibrational spectroscopy of leucine molecules adsorbed at air-water interface*. Journal of Chemical Physics, 2004. **120**(15): p. 7107-7112.
252. Weidner, T., et al., *Assembly and structure of alpha-helical peptide films on hydrophobic fluorocarbon surfaces*. Biointerphases, 2010. **5**(1): p. 9-16.
253. Mermut, O., et al., *In situ adsorption studies of a 14-amino acid leucine-lysine peptide onto hydrophobic polystyrene and hydrophilic silica surfaces using quartz crystal microbalance, atomic force microscopy, and sum frequency generation vibrational spectroscopy*. Journal of the American Chemical Society, 2006. **128**(11): p. 3598-3607.
254. Du, Q., E. Freysz, and Y.R. Shen, *Surface vibrational spectroscopic studies of hydrogen-bonding and hydrophobicity*. Science, 1994. **264**(5160): p. 826-828.
255. Weidner, T., et al., *Amide or amine: Determining the origin of the 3300 cm(-1) nh mode in protein sfg spectra using n-15 isotope labels*. Journal of Physical Chemistry B, 2009. **113**(47): p. 15423-15426.
256. Weidner, T. and D.G. Castner, *Sfg analysis of surface bound proteins: A route towards structure determination*. Physical Chemistry Chemical Physics, 2013. **15**(30): p. 12516-12524.
257. Nguyen, K.T., J.T. King, and Z. Chen, *Orientation determination of interfacial beta-sheet structures in situ*. Journal of Physical Chemistry B, 2010. **114**(25): p. 8291-8300.

258. Chirgadze, Y.N. and N.A. Nevskaya, *Infrared-spectra and resonance interaction of amide-one vibration of anti-parallel-chain pleated sheet*. Biopolymers, 1976. **15**(4): p. 607-625.
259. Beunis, F., et al., *Inverse micelles as charge carriers in nonpolar liquids: Characterization with current measurements*. Current Opinion in Colloid & Interface Science, 2013. **18**(2): p. 129-136.
260. Mathews, M.B. and E. Hirschhorn, *Solubilization and micelle formation in a hydrocarbon medium*. Journal of Colloid Science, 1953. **8**(1): p. 86-96.
261. Piper, J.D., et al., *Electrical insulating materials - liquid dielectrics - chemical, physical, electrical properties of systems containing lead or copper soaps in liquid paraffin*. Industrial and Engineering Chemistry, 1939. **31**: p. 307-317.
262. Berg, J.C., *An introduction to interfaces and colloids: The bridge to nanoscience*. 2010, Singapore: World Scientific Publishing Co. Pte. Ltd.
263. Lyklema, J., *Principles of interactions in non-aqueous electrolyte solutions*. Current Opinion in Colloid & Interface Science, 2013. **18**(2): p. 116-128.
264. Eastoe, J., et al., *Variation of surfactant counterion and its effect on the structure and properties of aerosol-ot-based water-in-oil microemulsions*. Journal of the Chemical Society-Faraday Transactions, 1992. **88**(3): p. 461-471.
265. Brown, P., et al., *Microemulsions as tunable nanomagnets*. Soft Matter, 2012. **8**(46): p. 11609-11612.
266. Burrows, H.D. and M.J. Tapia, *Lanthanide ion binding in aot/water/isooctane microemulsions*. Langmuir, 2002. **18**(17): p. 6706-6708.
267. Caboi, F., et al., *Microstructure of ca-aot/water/decane w/o microemulsions*. Journal of Physical Chemistry B, 1997. **101**(49): p. 10205-10212.
268. Dunn, C.M., B.H. Robinson, and F.J. Leng, *Photon-correlation spectroscopy applied to the size characterization of water-in-oil microemulsion systems stabilized by aerosol-ot effect of change in counterion*. Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 1990. **46**(6): p. 1017-1025.
269. Eastoe, J., et al., *Small-angle neutron scattering from novel bis-2-ethylenesulphosuccinate microemulsion - evidence for nonspherical structures*. Physica B, 1992. **180**: p. 555-557.
270. Eastoe, J., et al., *Structure of cobalt aerosl-ot reversed micelles studied by small-angle scattering methods*. Journal of the Chemical Society-Faraday Transactions, 1994. **90**(17): p. 2497-2504.
271. Haynes, J.S., J.R. Sams, and R.C. Thompson, *Sythesis and structural studies of iron(ii) and iron(iii) sulfonates*. Canadian Journal of Chemistry-Revue Canadienne De Chimie, 1981. **59**(4): p. 669-678.
272. Longo, A., A. Ruggirello, and V.T. Liveri, *Physicochemical investigation of nanostructures in liquid phases: Ytterbium nitrate ionic clusters confined in ytterbium bis(2-ethylhexyl) sulfosuccinate reversed micelles and liquid crystals*. Chemistry of Materials, 2007. **19**(5): p. 1127-1133.
273. Marques, M.F.F., et al., *Analysis and modelling of positron lifetime spectroscopy data for the characterisation of water-in-oil microemulsion systems stabilised by aerosol ot - effect of changing the counter-ion*. Journal of the Chemical Society-Faraday Transactions, 1997. **93**(21): p. 3827-3831.
274. Towey, T.F., et al., *Esr spectroscopy of metal bis(2-ethylhexyl) sulfoccinate aggregates in cyclohexane* Journal of Physical Chemistry, 1995. **99**(12): p. 3939-3942.
275. Eastoe, J., et al., *Structures of metal bis(2-ethylhexyl) sulfosuccinate aggregates in cyclohexane*. Journal of Physical Chemistry, 1993. **97**(7): p. 1459-1463.
276. de Oliveira, R.J., et al., *Photoreactive surfactants: A facile and clean route to oxide and metal nanoparticles in reverse micelles*. Langmuir, 2011. **27**(15): p. 9277-9284.
277. Eastoe, J., et al., *Preparation of colloidal cobalt using reversed micelles*. Colloids and Surfaces a-Physicochemical and Engineering Aspects, 1996. **119**(2-3): p. 123-131.
278. Rees, G.D., et al., *Formation and morphology of calcium sulfate nanoparticles and nanowires in water-in-oil microemulsions*. Langmuir, 1999. **15**(6): p. 1993-2002.
279. Rao, C.R.K., et al., *Synthesis and properties of polypyrrole obtained from a new fe(iii) complex as oxidizing agent*. Indian Journal of Chemistry Section a-Inorganic Bio-Inorganic Physical Theoretical & Analytical Chemistry, 2013. **52**(6): p. 744-748.
280. Brown, P., et al., *Magnetic emulsions with responsive surfactants*. Soft Matter, 2012. **8**(29): p. 7545-7546.
281. Brown, P., et al., *Magnetizing DNA and proteins using responsive surfactants*. Advanced Materials, 2012. **24**(46): p. 6244-6247.
282. Guardia E and P. JA, *Molecular dyanmics simulation of ferrous and ferric ions in water*. Chemical Physics, 1990. **144**(3): p. 353-362.

283. Parrinello, M. and A. Rahman, *Polymorphic transitions in single-crystals - a new molecular-dynamics method*. Journal of Applied Physics, 1981. **52**(12): p. 7182-7190.
284. Svergun, D.I., et al., *Protein hydration in solution: Experimental observation by x-ray and neutron scattering*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(5): p. 2267-2272.
285. Svergun, D., C. Barberato, and M.H.J. Koch, *Crysol - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates*. Journal of Applied Crystallography, 1995. **28**: p. 768-773.
286. Case, D.A., et al., *The amber biomolecular simulation programs*. Journal of Computational Chemistry, 2005. **26**(16): p. 1668–1688.
287. Tai, K., et al., *Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase*. Biophysical Journal, 2001. **81**(2): p. 715–724.
288. Kelkar, M.S., W. Shi, and E.J. Maginn, *Determining the accuracy of classical force fields for ionic liquids: Atomistic simulation of the thermodynamic and transport properties of 1-ethyl-3-methylimidazolium ethylsulfate ([emim][etso(4)]) and its mixtures with water*. Industrial & Engineering Chemistry Research, 2008. **47**(23): p. 9115–9126.
289. Liu, H.B., et al., *Understanding the interactions of cellulose with ionic liquids: A molecular dynamics study*. Journal of Physical Chemistry B, 2010. **114**(12): p. 4293–4301.
290. Martínez, L., et al., *Packmol: A package for building initial configurations for molecular dynamics simulations*. Journal of Computational Chemistry, 2009. **30**(13): p. 2157–2164.
291. Garcia-Miaja, G., J. Troncoso, and L. Romani, *Excess enthalpy, density, and heat capacity for binary systems of alkylimidazolium-based ionic liquids plus water*. Journal of Chemical Thermodynamics, 2009. **41**(2): p. 161–166.
292. Freire, M.G., et al., *Thermophysical characterization of ionic liquids able to dissolve biomass*. Journal of Chemical and Engineering Data, 2011. **56**(12): p. 4813–4822.
293. McQuarrie, D.A., *Statistical thermodynamics*. 1973, Mill Valley, CA: University Science Books. 343.

# Supporting Information

## Supporting Information for Chapter II

*Simulation Details*. Simulation boxes for purely aqueous systems were constructed in the tleap package of Ambertools from AMBER 11.[286] A minimum distance of 12.5 Å was left between the edge of the protein and the periodic boundary. Simulation boxes that contained ionic liquid were constructed by stacking smaller equilibrated boxes of IL or IL-water solutions together. For the IL simulations, the dimensions of the boxes were extended further by about 8 Å in each positive dimension since possible unfolding behavior of the protein could not be predicted *a priori*. The protein was placed in the stacked box, and all organic ions and non-crystallographic waters that possessed an atom within 2 Å of the protein were removed. At random, anions were selected and removed in order to equalize molar concentrations of the counterions. Chloride ions were added and randomly placed in the box to neutralize the charge of the protein.

All systems were minimized for a total of 10,000 steps using the standard conjugate gradient method in NAMD. The minimization consisted of three parts: protein fixed with solvent relaxing, solvent fixed with protein relaxing, and all atoms free to relax. The systems were heated to their simulation temperatures by a ramp over 100 ps. Simulations were initially conducted for 50 ns each with some simulations extended to 500 ns in order to study longer time behavior of the enzyme structure and fluctuations. Table S2.2 gives a full description of the composition, temperature, and simulation time of each system.

*Principal Component Analysis*. In order to extract correlated motions from the molecular dynamics trajectories, the g_covar and g_anaeig packages from GROMACS 4.5.4 were used.[198]

For each 500 ns molecular dynamics run, three separate 100 ns trajectories containing only protein alpha carbons were selected from equilibrated sections of the full all-atom trajectories. In our analysis we directly used the calculated covariance matrices as well as the primary and secondary eigenvectors of the covariance matrix to construct porcupine plots for visualization.[287] The vectors are scaled in order to improve the visualization on small figures. In each case, the porcupine plots of the 100 ns subsets of any given full trajectory were very similar. Therefore, only one porcupine plot for any molecular dynamics run is presented herein.

*IL Verification*. Force field parameters for the ionic liquids were developed using the antechamber package within ambertools. All the bonded and Lennard-Jones parameters were taken from GAFF. The electrostatic point charges of each individual ion were developed in antechamber using a restrained electrostatic potential (RESP) fitting of the electron density.[151] All calculations of the electronic structure and minimizations were performed in Gaussian09A using the Hartree-Fock method and a 6-31G* basis set in vacuo with a charge of +1 for cations and -1 for anions. The derived point charges are provided in Table S2.1. These point charges and force field parameters can be compared to those of Kelkar and Maginn for [EMIM][$EtSO_4$][288] and to those of Liu and Singh for [EMIM][OAc][289] who more thoroughly assess the thermophysical properties of these ionic liquids with similar point charge parameters.

Equilibrium boxes of ionic liquids at different concentrations were produced in Packmol[290] by setting the volume of a 200 ion-pair box to achieve the experimental density. In order to make the experimental methods easier and more comparable to the simulations, the concentrations cited in all of the computational studies are weight percent of water in the solvent. To validate the accuracy of the GAFF-based force fields, the boxes were simulated in the isobaric-isothermal (NPT) ensemble at a pressure of 1 bar and a temperature of 300 K for 10 ns

in order to equilibrate density. The heat capacity was then tested in the canonical (NVT) ensemble for 10 ns by setting the box dimensions to achieve the mean density of the equilibrated NPT simulations. In a similar fashion to how common tool solvate a protein in water, the final structures of the NPT runs were saved for later use in building systems for protein simulations.

In the case of [EMIM][$EtSO_4$], the calculated density was 1.27 +/- 0.01g $cm^{-3}$, [291] which is within 2.4% of the experimental density of 1.24 g $cm^{-3}$. Similarly, the density of [EMIM][OAc] was calculated to be 1.12 +/- 0.01g $cm^{-3}$, [292] which represents an error of 1.9% versus the experimental density of 1.10 g $cm^{-3}$. Since no parameterization was performed, it is reassuring to see reasonable prediction of density.

The heat capacity was also calculated using the standard definition of heat capacity in the canonical ensemble[293] in order to confirm the validity of a thermodynamic property of the force field. In this case, the force field did not agree as well with experimental data, but still within the range of operability and near the range of other published force fields of similar liquids. For [EMIM][OAc], the simulations predict a heat capacity of 288.8 J $mol^{-1}$ $K^{-1}$, where the experimental heat capacity is 321.9 kJ $mol^{-1}$ $K^{-1}$.[292] The second solvent, [EMIM][$EtSO_4$] performs slightly more poorly giving a heat capacity of 342.4 J $mol^{-1}$ $K^{-1}$ compared to an experimental heat capacity of 394.4 kJ $mol^{-1}$ $K^{-1}$.[291] The experimental values are for a temperature of 298.15 K while the simulations are at 300 K. These parameters are similar to Kalkar and Maginn's published force field for pure [EMIM][$EtSO_4$] for the two properties tested herein.[288] Their simulations indicate that several thermophysical properties of mixtures containing low concentrations of ionic liquid and high concentrations of water are quite accurate when compared with experimental properties.

*Xylose Yield*. Xylose concentration was measured using a Shimadzu HPLC equipped with a RID-20A refractive index detector, a SPD-20A UV-visual detector, a LC-20AD liquid chromatograph, and a Rezex ROA column. The column temperature was set to 60°C, and the detector temperature was set to 40°C with a mobile phase of 0.005 N sulfuric acid. A set of standards showed that the ionic liquids did not affect the ability of the equipment to accurately measure sugar concentrations for up to 20 wt% [EMIM][$EtSO_4$] and up to 50 wt% [EMIM][OAc]. At higher concentrations, the xylose peak merged with a neighboring peak. For these higher concentrations, a YSI 7100 Multiparameter Bioanalytical System equipped with a xylose membrane was used to determine the concentration of xylose. As with the HPLC, a set of standards indicated that the ionic liquid had no effect on the ability of the equipment to determine xylose concentration, and the xylose concentrations measured with this method matched those of the HPLC. All yield experiments were conducted in triplicate, and errors are reported as standard deviation over the total of six samples.

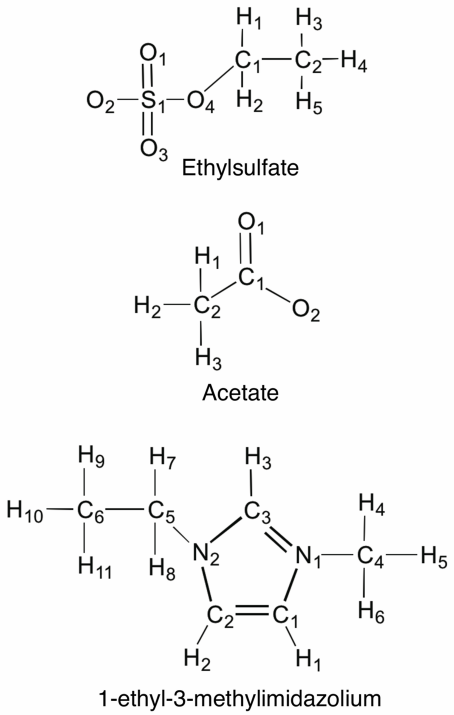



**Figure S2.1:** Naming scheme of the ionic liquids.

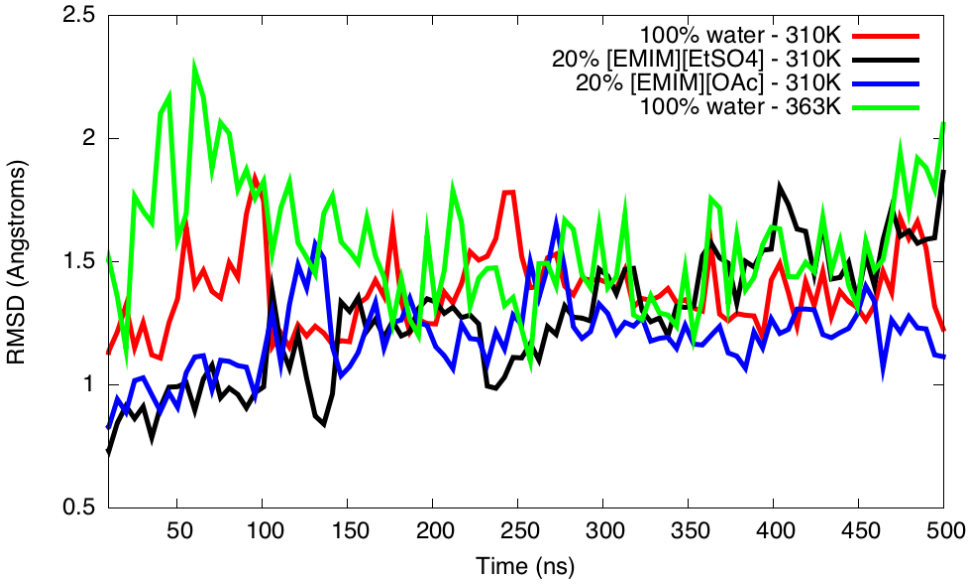



**Figure S2.2:** RMSD of systems extended to 500 ns. Lines are Bezier smoothed.

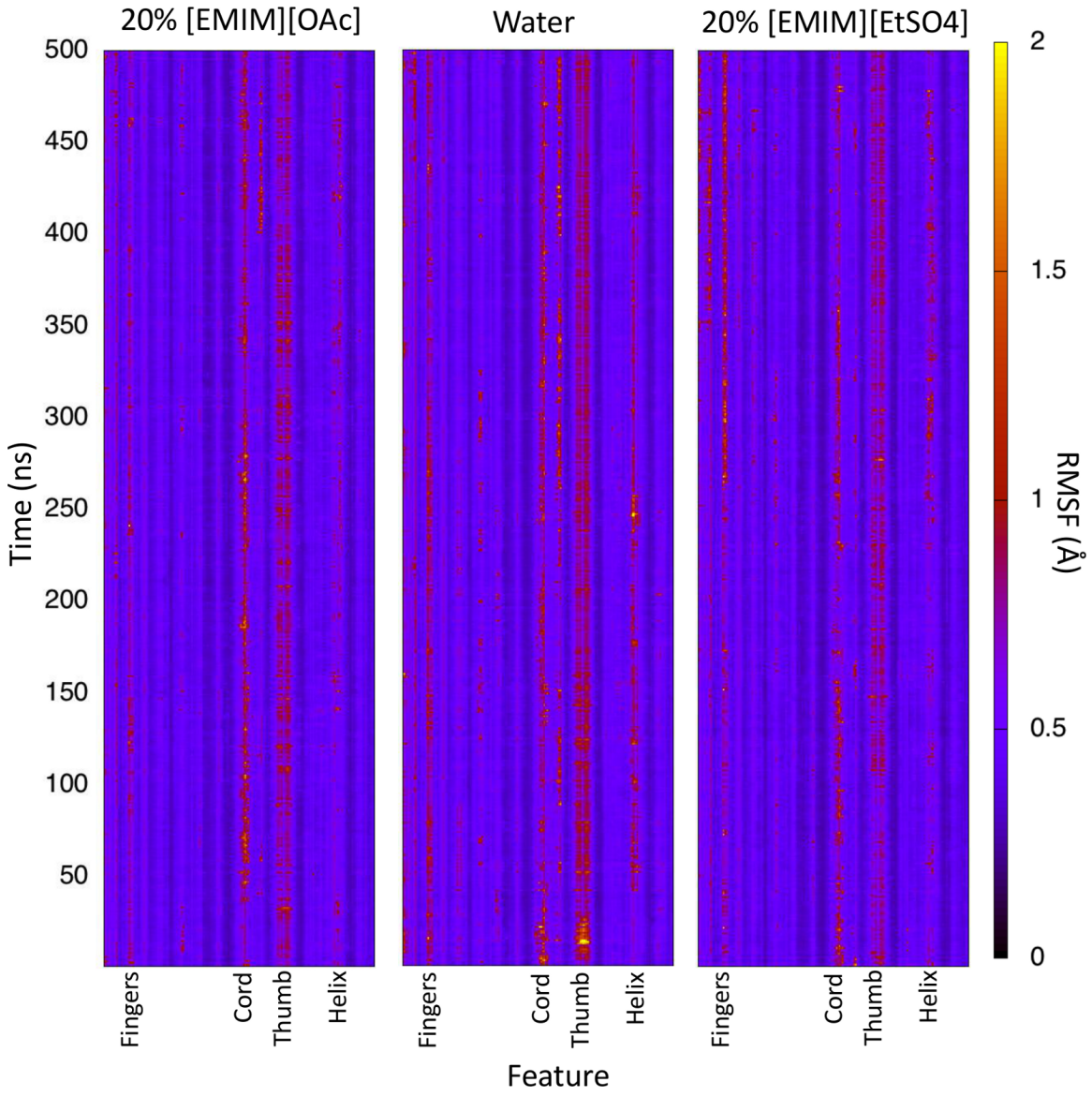



**Figure S2.3:** Time windowed xylanase alpha carbon root mean square fluctuations. Each pixel represents one reside and one nanosecond. Values were capped at a maximum of 2.0 Å to ensure resolution in the lower range of RMSF. Data were collected at 310 K.

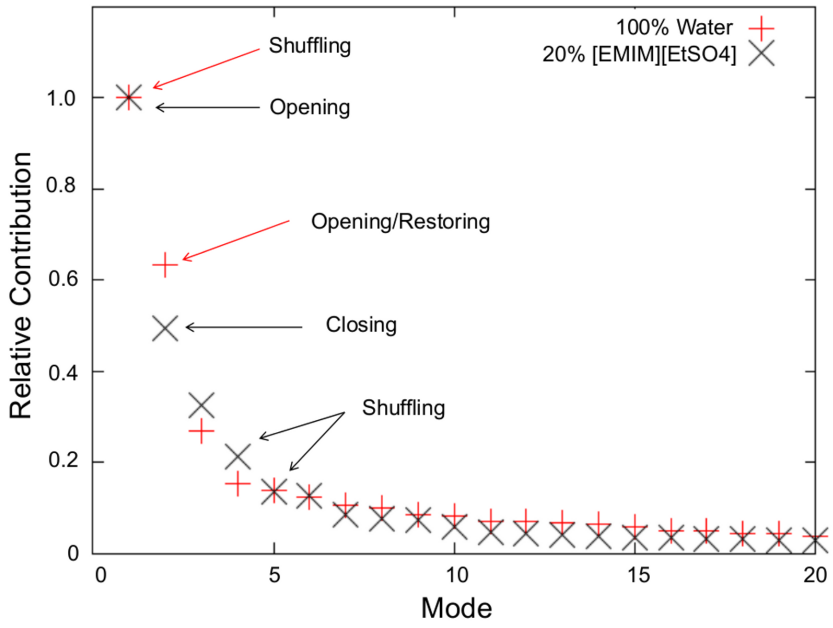



**Figure S2.4:** Contribution of PCA modes. The relative contribution of each of the first 20 PCA modes is plotted. The type of motion is described. It can be seen that the first few modes contribute most significantly.

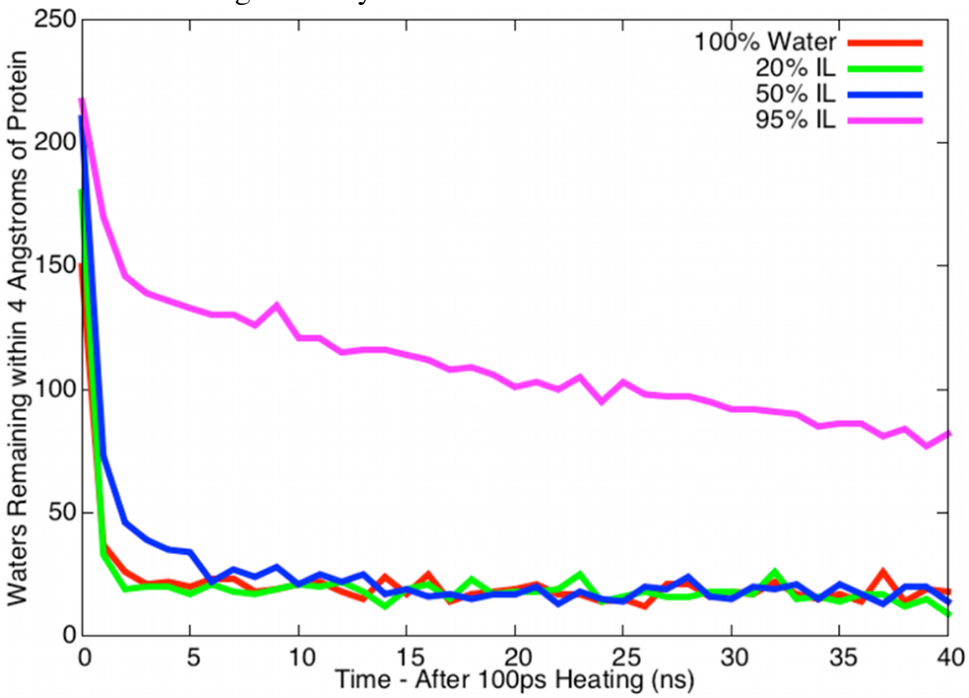



**Figure S2.5:** Fate of crystallographic waters. The number of crystallographic waters within 4 Å of the surface was tracked over the first 40 ns of each of the simulations for [EMIM][OAc] at 310K. The 95% IL system does not converge over this time period.

**Table S2.1:** Electrostatic point charges of the ionic liquid models.

| | Name | Charge (e) |
|---|---|---|
| Ethyl sulfate | S1 | 1.272518 |
| | O1 | –0.676331 |
| | O2 | –0.676331 |
| | O3 | –0.676331 |
| | O4 | –0.493129 |
| | C1 | 0.489427 |
| | H1 | –0.06064 |
| | H2 | –0.06064 |
| | C2 | –0.253901 |
| | H3 | 0.045119 |
| | H4 | 0.045119 |
| | H5 | 0.045119 |
| Acetate | C1 | 0.943591 |
| | C2 | –0.331169 |
| | H1 | 0.033195 |
| | H2 | 0.033195 |
| | H3 | 0.033195 |
| | O1 | –0.856003 |
| | O2 | –0.856003 |
| Imidazolium | N1 | 0.086098 |
| | C1 | –0.139338 |
| | H1 | 0.243557 |
| | C2 | –0.229391 |
| | H2 | 0.247313 |
| | N2 | 0.061656 |
| | C3 | 0.041447 |
| | H3 | 0.213804 |
| | C4 | –0.277769 |
| | H4 | 0.158439 |
| | H5 | 0.158439 |
| | H6 | 0.158439 |
| | C5 | –0.026788 |
| | H7 | 0.123268 |
| | H8 | 0.123268 |
| | C6 | –0.238261 |
| | H9 | 0.098607 |
| | H10 | 0.098607 |
| | H11 | 0.098607 |

**Table S2.2:** Composition, temperature, and simulation time of each system. An asterisk (*) indicates a 500 ns simulation while all others were 50 ns in length. Each system contained 2,818 enzyme atoms and two chloride ions. Concentration indicates the concentration of the constituent equilibrium boxes used to construct the full simulation box.

| Solvent | T(K) | Conc. (wt%) | Water Molecules | Cation Molecules | Anion Molecules | Total Atoms |
|---|---|---|---|---|---|---|
| Water* | 310 | 100 | 9517 | - | - | 31371 |
| Water | 323 | 100 | 9517 | - | - | 31371 |
| Water | 343 | 100 | 9517 | - | - | 31371 |
| Water* | 363 | 100 | 9517 | - | - | 31371 |
| [EtSO4]* | 310 | 20 | 15878 | 286 | 286 | 59320 |
| [EtSO4] | 323 | 20 | 15878 | 286 | 286 | 59320 |
| [EtSO4] | 343 | 20 | 15878 | 286 | 286 | 59320 |
| [EtSO4] | 363 | 20 | 15878 | 286 | 286 | 59320 |
| [OAc]* | 310 | 20 | 11075 | 270 | 270 | 43065 |
| [OAc] | 323 | 20 | 11075 | 270 | 270 | 43065 |
| [OAc] | 343 | 20 | 11075 | 270 | 270 | 43065 |
| [OAc] | 363 | 20 | 11075 | 270 | 270 | 43065 |
| [EtSO4] | 310 | 50 | 9874 | 698 | 698 | 54080 |
| [EtSO4] | 323 | 50 | 9874 | 698 | 698 | 54080 |
| [EtSO4] | 343 | 50 | 9874 | 698 | 698 | 54080 |
| [EtSO4] | 363 | 50 | 9874 | 698 | 698 | 54080 |
| [OAc] | 310 | 50 | 7047 | 676 | 676 | 41537 |
| [OAc] | 323 | 50 | 7047 | 676 | 676 | 41537 |
| [OAc] | 343 | 50 | 7047 | 676 | 676 | 41537 |
| [OAc] | 363 | 50 | 7047 | 676 | 676 | 41537 |
| [EtSO4] | 310 | 95 | 1218 | 1341 | 1341 | 48045 |
| [EtSO4] | 323 | 95 | 1218 | 1341 | 1341 | 48045 |
| [EtSO4] | 343 | 95 | 1218 | 1341 | 1341 | 48045 |
| [EtSO4] | 363 | 95 | 1218 | 1341 | 1341 | 48045 |
| [OAc] | 310 | 95 | 933 | 1286 | 1286 | 39055 |
| [OAc] | 323 | 95 | 933 | 1286 | 1286 | 39055 |
| [OAc] | 343 | 95 | 933 | 1286 | 1286 | 39055 |
| [OAc] | 363 | 95 | 933 | 1286 | 1286 | 39055 |

## Supporting Information for Chapter III

This work included analysis of 54 molecular dynamics simulations. The simulated systems are composed of: the enzymes *Trichoderma viride* (VIR), *Thermotoga maritima* (TMA), and *Pyrococcus horikoshii* (PHO); each in water with 0, 15 or 50 wt% [EMIM][OAc]; and run at both 310 and 353 K. Each of these 18 systems are simulated in triplicate to increase statistics and help to verify observations. The list of simulations is provided in Table S3.1. Root-mean-square displacement (RMSD) was used to track departure from the experimental reference structure for each of the 54 simulations, but the number of systems made the figures cluttered. For this reason we have relegated plots of the RMSD from each trajectory to this supporting document and used a table of summarizing statistics in the main article. Bezier curve fitting is also used to help distinguish each replicate trajectory in Figures S3.1 and S3.2 below.

**Table S3.1:** A list of the conditions of each of the 54 simulations performed. VIR, PDB: 3QR3. TMA, PDB: 3AMD. PHO, PDB: 3AXX.

| Enzyme | Temp. (K) | IL (wt%) | Time (ns) | Replicas | pH |
|---|---|---|---|---|---|
| VIR | 310 | 0 | 500 | 3 | 4.5 |
| VIR | 310 | 15 | 500 | 3 | 4.5 |
| VIR | 310 | 50 | 500 | 3 | 4.5 |
| VIR | 353 | 0 | 500 | 3 | 4.5 |
| VIR | 353 | 15 | 500 | 3 | 4.5 |
| VIR | 353 | 50 | 500 | 3 | 4.5 |
| TMA | 310 | 0 | 500 | 3 | 4.8 |
| TMA | 310 | 15 | 500 | 3 | 4.8 |
| TMA | 310 | 50 | 500 | 3 | 4.8 |
| TMA | 353 | 0 | 500 | 3 | 4.8 |
| TMA | 353 | 15 | 500 | 3 | 4.8 |
| TMA | 353 | 50 | 500 | 3 | 4.8 |
| PHO | 310 | 0 | 500 | 3 | 6.4 |
| PHO | 310 | 15 | 500 | 3 | 6.4 |
| PHO | 310 | 50 | 500 | 3 | 6.4 |
| PHO | 353 | 0 | 500 | 3 | 6.4 |
| PHO | 353 | 15 | 500 | 3 | 6.4 |
| PHO | 353 | 50 | 500 | 3 | 6.4 |

**Table S3.2:** List of charged GLU and ASP residues. If not listed, GLU or ASP is protonated. All LYS and ARG are charged. All HIS are uncharged. Residue indices are from the respective PDB files.

| | | |
|---|---|---|
| VIR | GLU | 232, 259 |
| | ASP | 20, 43, 54, 79, 87, 152, 195, 208, 223, 225, 239, 285, 299 |
| TMA | GLU | 7, 27, 37, 78, 99, 105, 106, 127, 144, 151, 152, 172, 180, 188, 202, 209, 212, 215, 234, 239, 240, 253, 263, 275 |
| | ASP | 4, 29, 36, 40, 70, 77, 103, 120, 124, 161, 226, 227, 261, 296, 305 |
| PHO | GLU | 42, 57, 58, 87, 108, 137, 163, 169, 173, 174, 182, 234, 333, 342, 368, 399 |
| | ASP | 46, 88, 120, 126, 131, 145, 170, 175, 197, 214, 228, 262, 282, 302, 317, 321, 352, 364, 382, 385, 392, 393, 400, 410 |

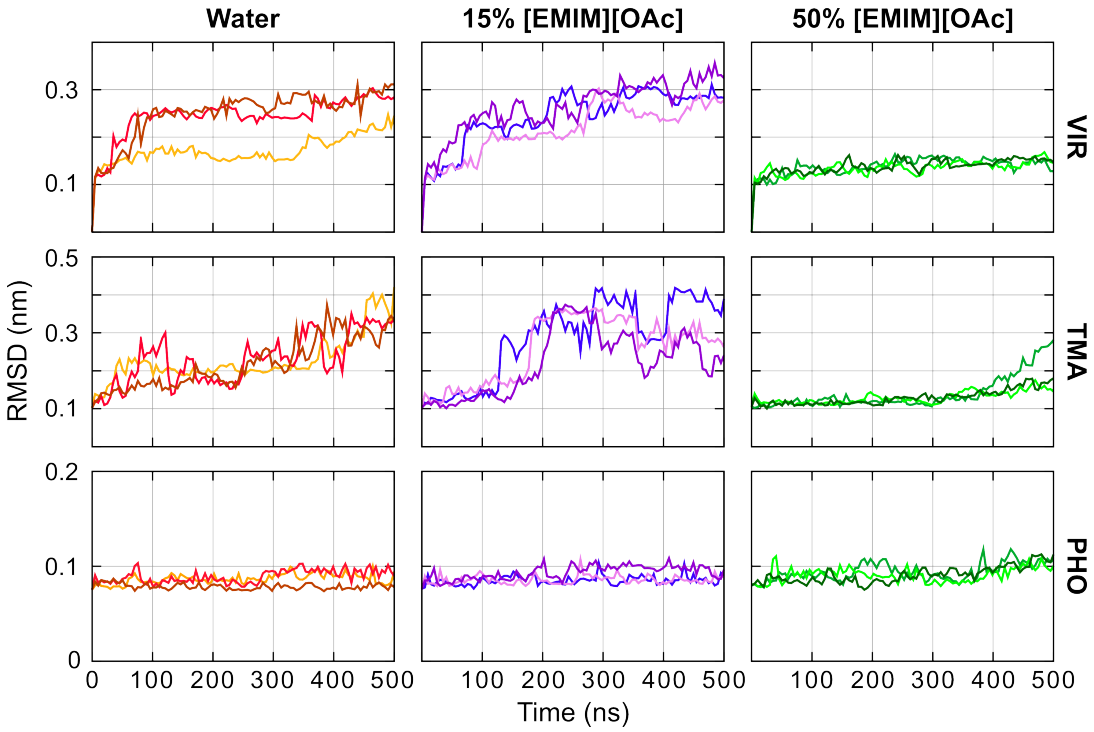



**Figure S3.1:** Bezier curve fitting of the Cα RMSD for each of the simulated systems at 310 K. The RMSD was calculated from (after alignment to) the corresponding experimental structure for each enzyme. Note the change in y-axis scale between each row.

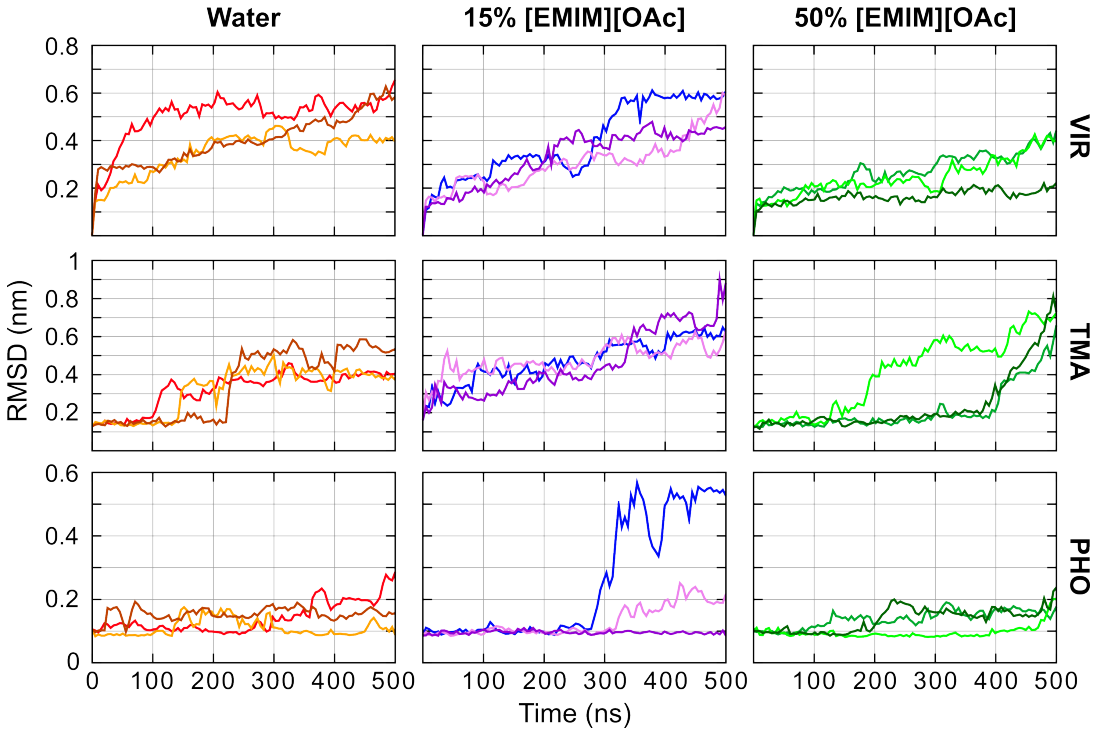



**Figure S3.2:** Bezier curve fitting of the Cα RMSD for each of the simulated systems at 353 K. The RMSD was calculated from (after alignment to) the corresponding experimental structure for each enzyme. Note the change in y-axis scale between each row.

RMSD was calculated using the g_rms tool in GROMACS 4.6 [235]. These data are for the full sequence, whereas those presented in the main document for TMA do not include the first 15 N-terminal carbons. The reference structure for VIR was obtained by modifying GLU53 to ASP in the *Trichoderma reesei* (PDB: 3QR3) [209] x-ray structure in order to match the known sequence for *T. viride* endoglucanase [210]. The reference structures for TMA and PHO were modeled using x-ray structures (PDB: 3AMD, 3AXX) [211, 212].

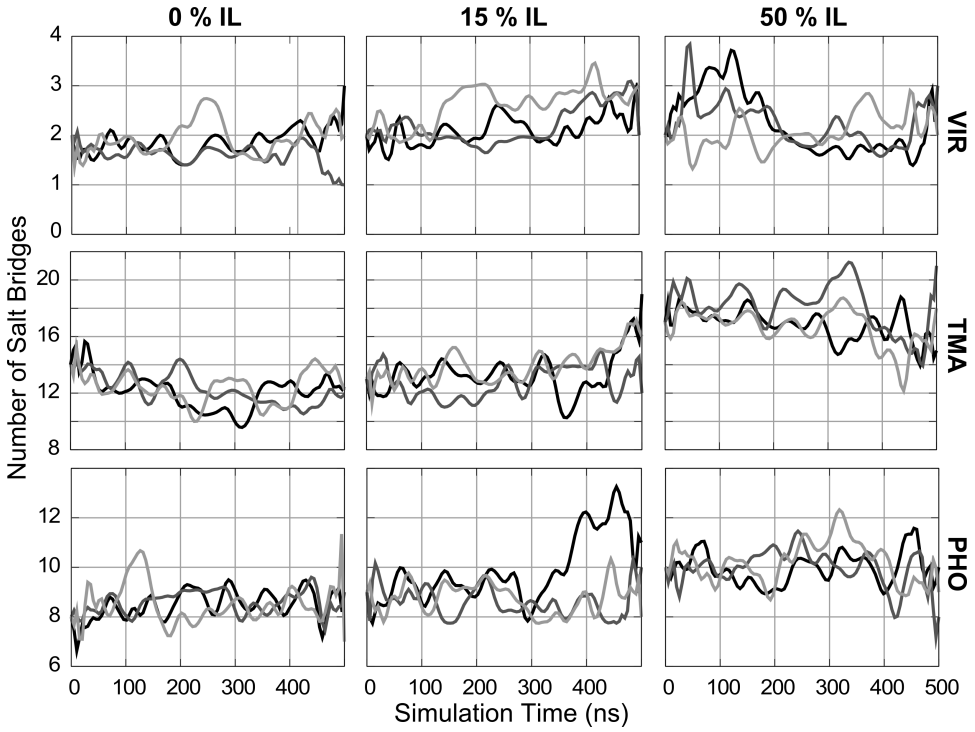



**Figure S3.3:** Bezier curve fitting of number of salt bridges for each of the simulated systems at its optimum temperature. Note the change in y-axis scale.

Salt bridges are calculated by measuring the distance between all salt-bridge-forming pairs of oxygens and nitrogens over the trajectory. This simplifies the calculation compared to calculating each oxygen-hydrogen pair. Plots begin after minimization. The intensity of each salt bridge, *S*, between atom *i* and *j* is calculated by using the sigmoidal function below. When compared to a hard cutoff, this method gives a smoother curve that helps with visualization. The intensities are summed over the whole protein. Bezier smoothing was used to smooth the plots. This removes some detail, but it separates the plots so that they can be distinguished across the three replicas. The smoothing removes the more extreme values, especially if the extremes are short-lived.

$$S = \frac{1 - \left(\frac{r_{ij}}{3.7}\right)^{20}}{1 - \left(\frac{r_{ij}}{3.7}\right)^{40}}$$

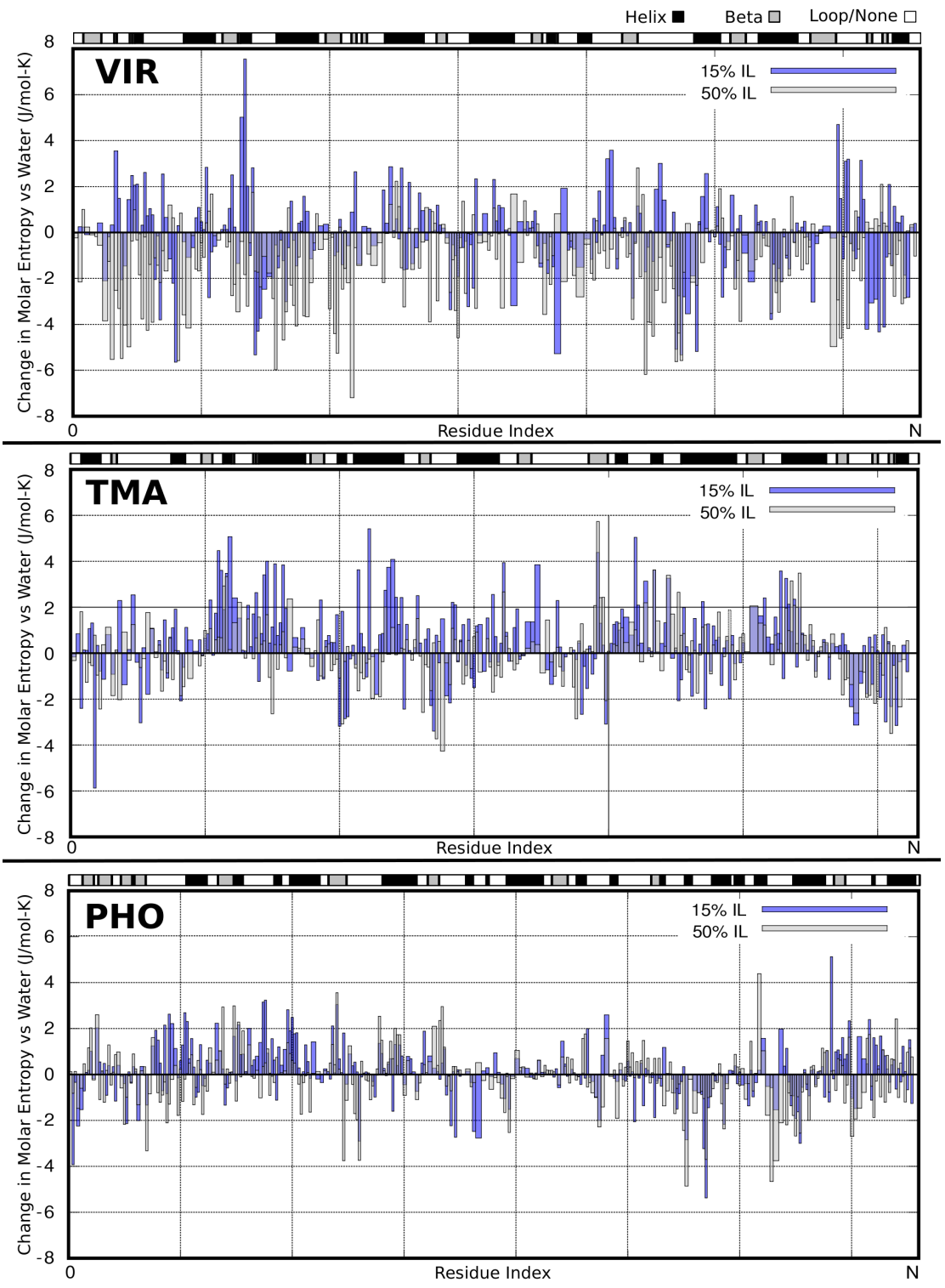



**Figure S3.4:** Relative statistical entropy of sidechains $\chi_1$ angles. Wide bars compensate for alanines that do not have a $\chi_1$ angle. Secondary structure is displayed at the top.

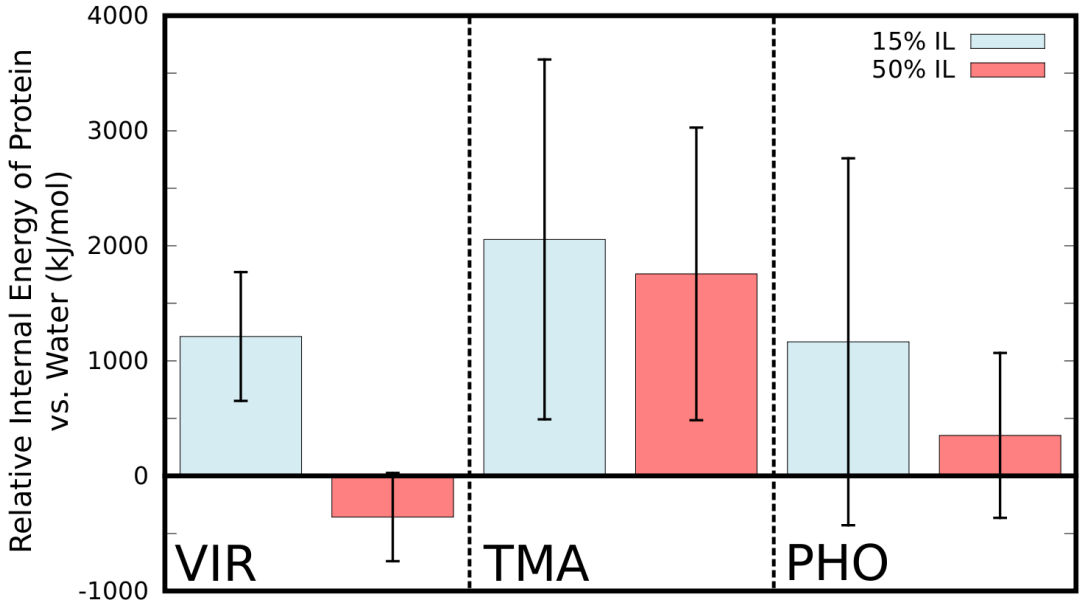
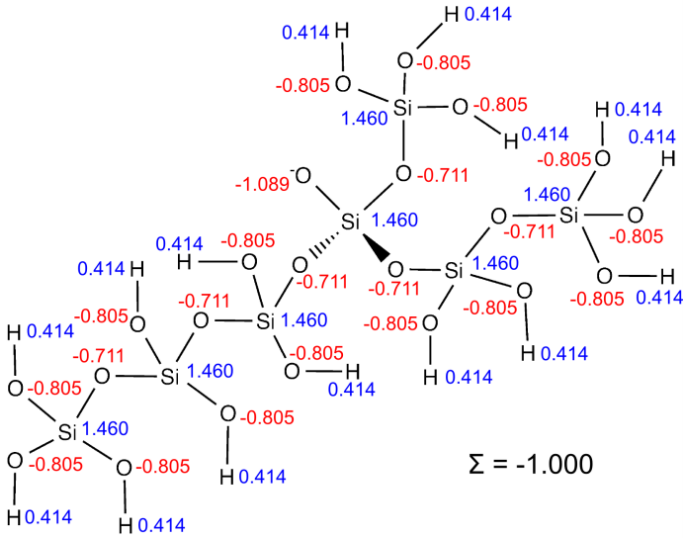



**Figure S3.5:** Relative internal energy of the protein averaged over the last half of the trajectories. Error bars are standard one deviation of the data.

## Supporting Information for Chapter VI



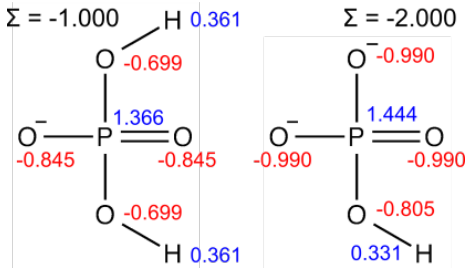




**Figure S6.1:** Chemical structure of the silica model precursor and the phosphate ions with their electrostatic point charges.

The following .frcmod file was used for silica model precursors.

| MASS | | |
|---|---|---|
| Si | 28.085 | 0.000 |
| os | 16.000 | 0.465 |
| oh | 16.000 | 0.465 |
| ho | 1.008 | 0.135 |

| BOND | | |
|---|---|---|
| Si-os | 885.10 | 1.610 |
| Si-oh | 428.00 | 1.420 |
| oh-ho | 545.00 | 0.960 |

| ANGLE | | |
|---|---|---|
| Si-os-Si | 4.660 | 174.220 |
| os-Si-os | 159.570 | 110.930 |
| os-Si-oh | 153.260 | 111.090 |
| Si-oh-ho | 57.500 | 106.000 |
| oh-Si-oh | 89.620 | 116.260 |

| DIHE | | | | |
|---|---|---|---|---|
| Si-os-Si-os | 1 | 0.000 | 0.000 | 1.000 |
| Si-os-Si-oh | 1 | 0.000 | 0.000 | 1.000 |
| oh-Si-oh-ho | 1 | 0.280 | 30.000 | 2.000 |
| ho-oh-Si-os | 1 | 0.630 | 0.000 | -2.000 |
| ho-oh-Si-os | 1 | 0.280 | 180.000 | 1.000 |

| IMPROPER | | | |
|---|---|---|---|
| Si-os-Si-os | 1.1 | 180.0 | 2.0 |
| Si-os-Si-oh | 1.1 | 180.0 | 2.0 |

| NONBON | | |
|---|---|---|
| Si | 2.1475 | 0.3000 |
| os | 1.7500 | 0.1700 |
| oh | 1.7700 | 0.2104 |
| ho | 0.2245 | 0.0460 |

The following .frcmod file was used for phosphate ions.

| MASS | | |
|---|---|---|
| p5 | 30.970 | 1.538 |
| o | 16.000 | 0.434 |
| oh | 16.000 | 0.465 |
| ho | 1.008 | 0.135 |

| BOND | | |
|---|---|---|
| p5-o | 487.70 | 1.481 |
| p5-oh | 321.20 | 1.625 |
| oh-ho | 369.60 | 0.974 |

| ANGLE | | |
|---|---|---|
| p5-oh-ho | 44.150 | 110.140 |
| o-p5-oh | 69.980 | 115.260 |
| o-p5-o | 73.530 | 115.800 |
| oh-p5-oh | 71.250 | 102.450 |

| DIHE | | | | |
|---|---|---|---|---|
| o-p5-oh-ho | 1 | 0.533 | 0.000 | 3.000 |
| oh-p5-oh-ho | 1 | 0.533 | 0.000 | 3.000 |

| IMPROPER |
|---|

| NONBON | | |
|---|---|---|
| p5 | 2.1000 | 0.2000 |
| o | 1.6612 | 0.2100 |
| oh | 1.7210 | 0.2104 |
| ho | 0.2245 | 0.0460 |

## Supporting Information for Chapter VII

**Table S7.1:** Number of molecules of each type in the simulation boxes. After pressure equilibration, cubic box side length ranges from 6.8 to 7.8 nm depending on the system. Density fluctuations are small after equilibration.

| | Number of molecules | | | |
|---|---|---|---|---|
| **Molecule** | **DBSo-Methanol** | **DBSo-Toluene** | **AOT-Methanol** | **AOT-Toluene** |
| **DBSo** | 24 | 24 | - | - |
| **AOT** | - | - | 24 | 24 |
| **$Fe^{3+}$** | 8 | 8 | 8 | 8 |
| **Methanol** | 4470 | - | 4470 | - |
| **Toluene** | - | 2448 | - | 2448 |