Modeling Protein Stability and Structural Preference in Ionic Liquids

Patrick R. Burney

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

Doctor of Philosophy

University of Washington
2014

Reading Committee:

Jim Pfaendtner, Chair

David Beck

Shaoyi Jiang

Program Authorized to Offer Degree:
Chemical Engineering

University of Washington

Abstract

Modeling Protein Stability and Structural Preference in Ionic Liquids

Patrick R. Burney

Chair of the Supervisory Committee:
Professor Jim Pfaendtner, Ph.D.
Chemical Engineering

Ionic liquids have demonstrable suitability as niche alternatives to organic solvents in applied biochemistry. However, many of the known ionic liquids show little to no support for enzyme activity and many cases have been demonstrated where structurally and functionally similar enzymes exhibit discrepant retention of activity in ionic liquids. Although many experimental investigations have attempted to formulate rules that predict enzyme tolerance to ionic liquids, there are few rules that apply broadly. Molecular dynamics is a computational technique that is well suited to provide atomistic insight about the nature and effects of ionic liquid / enzyme interactions. This research has involved development and application of simulation methodology and analysis toward investigating these complex interactions as well as the structural stability of multiple enzymes in ionic liquids. Application of these techniques has indicated that the nature of enzyme deactivation by ionic liquids is just as complicated as the experimental results have suggested; different enzymes and different ionic liquids show indications of competitive inhibition by the solvent or destabilization. Through several independent studies we have observed and described: the dampening of enzyme fluctuations in high concentrations of ionic liquids; the organization of water and ionic liquid substituents about enzymes; and the loss of secondary structure that precedes denaturing and possibly enzyme aggregation. In recognition that future studies in this area will also need to go beyond classical MD simulations, we have applied special multiscale modeling techniques to a separate protein system to quantify how mutations induce changes in the preferred structural state of an important protein fragment.

Table of Contents

Dedication

To my grandfather Morton, who helped me build my first computer and who nurtured my interests in science and my desire to understand the world around me

Preface

Contained herein are the written details of the most interesting research progress made while working toward my PhD. Specifically, the document begins with a brief summary of the key research goals and motivations behind this project. The introduction ends with an overview of the chapters that comprise body of this document; this includes both the state of the articles' publication and the names of collaborators I have worked with while developing them. The body of the document contains a single chapter for each of the four manuscripts written while completing my research objectives. Published and submitted work is included in the form of their latest/final journal submission. Abstracts have been removed from each manuscript to avoid redundancy with the summaries presented in Section 1.2. Finally, the body of this document ends with some concluding remarks about 1) the significance of this research and the contributions we made to the field through it and 2) suggested future research. Supporting materials for Chapters 2 and 3 can be found appended to this document (Appendix 1 and 2, respectively).

Progress through this PhD program has been long and difficult for me, both because of my initial lack of computer programming skills and, much to the frustration of my advisor, highly critical or even skeptical nature when it comes to my own work. Getting this far would not be possible without the patient and hands-on support of my advisor, Professor Jim Pfaendtner. Although I discovered this fairly late in my graduate career, I have begun thoroughly enjoy collaborating both within and outside our group. I must acknowledge my collaborators: Professor Nathan White, Vance Jaeger, Mike Deighan, Eric Nordwald, Professor Joel Kaar and Katie Hickman. I would also like to acknowledge UW-IT. The high-performance computing and data storage services provided and maintained by UW-IT have been absolutely critical for performing the needed computational research in support of my thesis.

1 Introduction

1.1 Motivation

Ionic liquids (ILs), salts that are liquid at temperatures less than 100°C, are a novel alternative to organic solvents in many areas of chemistry. ILs gained attention in academic literature in the late 1990's and early 2000's as "green" reaction media and solvents toward laboratory and industrial applications [1,2] and the growth in scientific interest in ILs since then has been remarkable [3,4]. While their environmental friendliness is still under debate, many of the commercially available ILs have a near-zero vapor pressure, are non-flammable, and are (chemically and thermally) more stable than organic solvents. Typically, ILs consist of an organic cation – usually alkylated amines [3] – paired with a wide variety of organic or inorganic anions. One of the most exciting aspects of ILs is the potential for chemical modification of these ions. By "tailoring" the individual ion's functional groups, such as the length of an alkyl chain, one might fine-tune reaction solvents toward desirable properties, such as product or substrate solubility [5]. When combined with the wide variety of cations and ions capable of forming ILs, this customizability creates a large library of potential solvents.

The aspect of some ILs that has been of more recent interest is their capacity to 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 ILs have the potential to add control over many reaction properties, notably the preference for synthesis over hydrolysis in non-aqueous ILs. Commonly reported applications of ILs in biocatalysis include the pretreatment and breakdown of lignocellulosic biomass [6-8], the degradation of food waste oils [9,10], and the production of pharmaceuticals [11-13]. Not only do many enzymes retain activity in pure ILs [14], but also exhibit enhanced thermal stability [15,16], selectivity [17] and enantioselectivity [18,19] compared to traditional solvents. The potential to

tailor ILs provides an opportunity to engineer these desirable properties towards optimizing product yield and facilitating product separation and enzyme recovery.

However, many studies have revealed that some IL cations or anions can be detrimental to or even eliminate an enzyme's activity. Other studies have indicated that proteins with similar structure and function can have greatly differing degrees of thermal stability or activity in ILs [20,21]. These observations point to an obvious research objective – to determine which properties of ILs contribute to diminished protein stability or capacity to preform catalysis. While many groups have published studies toward this goal – even going so far as to propose Hofmeister series of ions' ability to support enzymatic activity in aqueous IL mixtures [22-24] – there is still no consensus or unifying principle and little understanding of the fundamental mechanism by which ILs induce molecular level changes in an enzyme, especially those changes that lead to altered enzyme activity. The interactions between the IL and the enzyme structure may alter the equilibrium structure or dynamic fluctuations, competively inhibit catalysis by blocking the enzyme active site, or disrupt tertiary structure leading to denaturation or aggregation. The link between these atomic-scale interactions and reaction rates [25] and thermal stability remains an area of great interest.

Obtaining a mechanistic understanding of the interaction between proteins and ILs is an important first step toward effectively engineering ILs and proteins and synergizing their role in industrial and laboratory applications. The current body of literature in this field, both historically and currently, is dominated by more empirical or serendipitous discovery – mixing different enzymes in available ILs and reporting their activity or thermal stability [26]. Additionally, there are many studies describing efforts to fix enzymes on various substrates toward increasing their stability in ILs – mirroring the efforts in past research on biocatalysis in organic solvents [27,28]. Fundamental experimental studies toward relating IL characteristics to enzyme stability are greatly hindered by the

lack of analytical techniques that function broadly in the immense design-scape offered by ILs, especially at higher concentrations of IL. In addition to the difficulty in purifying enzymes, these limitations have restricted many studies to focus on discovery of ideal cation-anion pairs and characterizing enzyme fixation without atomistic insight guiding their search.

This project's main focus is to find and apply computational approaches that identify aspects of protein structure and solvation that explain the detrimental or stabilizing affects of ILs. Secondly, we strive to support or eliminate many of the long-standing hypotheses proposed by experimentalists in the field and facilitate engineering of ILs by improving the understanding of which properties of ILs or enzymes provide synergistic effects. Our approaches toward this goal will combine both studies of protein structural preference in addition to solvent structuring about the enzyme and coordinating with surface residues. These features of the IL/enzyme systems were indicated as important and interesting in our initial studies, including my own work [29] and my colleague Vance Jaeger's [30]. An important theme throughout this project is using simulation to compliment established or concurrent experimental results by providing the missing atomistic details. We accomplish this through collaboration with experimental research groups and by using recent and novel publications in the field to guide our investigations. Finally, this project represents an initial push in a larger study with the long-term goal of establishing computational pre-screening tools that widely and effectively predict IL/enzyme interactions. The tools and methodologies we develop here may become fundamental in this broader effort.

1.2 Chapter Summaries

Chapter 2 is the article I prepared with Professor Jim Pfaendtner and was subsequently published in the *Journal of Physical Chemistry-B* (JPCB) [29] in March, 2013. The supporting materials for the JPCB article can be found near the end of this document in Appendix 1. This article summarizes our

group's first study involving simulation of proteins in ionic liquids (ILs) – a theme that has become fundamental to our groups continued research objectives. In addition to our group's second published article in this area [30], written by my colleague Vance Jaeger, principles learned from this first study greatly influenced the simulation methodology and analysis we applied in subsequent research. In Chapter 2 we report the results for molecular dynamics simulations of *Candida rugosa* lipase (CRL) in ILs 1-ethyl-3-methyl-imidazolium hexafluorophosphate (BMIM-PF_6) and 1-ethyl-3-methyl-imidazolium nitrate as well as organic solvents toluene and octane in an effort to explore the role of solvent on the structure and dynamics of an enzyme known to be active in many non-aqueous media. Simulations of CRL in water were also included for comparison, bringing the aggregate simulation time to over 2.8 μs. We found that at both 310 and 375 K the ILs significantly dampen protein dynamics and trap the system near its starting structure. Structural changes in the enzyme follow the viscosity of the solvent, with the enzyme deviating from its initial structure the most in water and the least in BMIM-PF_6. Interactions between the enzyme surface and the solvent in the IL simulations show that contacts are dominated by the IL anion, which we ascribe to a broader spatial distribution of positively charged protein residues and reduced mobility of the cation due to the size of the imadazolium ring.

Chapter 3 is an article I prepared with Vance Jaeger and Prof. Pfaendtner. We have submitted this manuscript to *Biophysical Journal*. Supporting material for Chapter 3 can be found in Appendix 2. Vance and I collaborated on the design, setup, analysis, and subsequent writing for this project. The research itself consisted of the most extensive, both in terms of number of simulations and simulation time scale, computational work to date concerning enzyme behavior in ionic liquids. The 54 independent simulations used in this study account for 27 μs of net simulation time. Chapter 3 describes our use of molecular dynamics to investigate the differences in ionic liquid tolerance among three distinct family five cellulases from *Trichoderma viride*, *Thermogata 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 computation methodology was modeled after a recent experimental investigation of these three enzymes in aqueous IL mixtures. The analyses we employed demonstrate that the effects of ionic liquids on the enzymes vary in each individual case. Monitoring the enzymes' structures over a 500 ns time scale revealed 1) significant changes in the local enzyme structure near the binding site of the least IL-tolerant cellulase, 2) detrimental loss of secondary structure that intensifies with increasing concentration of ionic liquid, and 3) structural stability and dynamics similar to that observed in water for the most IL-tolerant cellulase. These observations and others are used to propose specific experimentally testable hypotheses regarding the origin of activity loss for each of the cellulases we have investigated.

Chapter 4 is a complete manuscript that is in preparation for journal submission. The work detailed in this manuscript was completed in collaboration with Prof. Joel Kaar and his student Eric Nordwald from the University of Colorado in Boulder, in addition to an undergraduate researcher from UW chemical engineering, Katie Hickman. Presently, the manuscript is awaiting approval from the collaborating parties and additional analysis is being considered. The manuscript details molecular simulations of the enzymes CRL and *Bos taurus* α-chymotrypsin in aqueous ionic liquids 1-butyl-3-methylimidazolium chloride and 1-ethyl-3-methylimidazolium ethyl sulfate. These simulations designed to study the change in protein-solvent interactions induced by modification of the enzyme surface charge in direct comparison with two recent experimental studies [31,32]. The enzymes were altered by randomly mutating lysine surface residues to glutamate, effectively decreasing the net surface charge by two for each mutation. These mutations also resemble succinylation of the enzyme by chemical modification detailed in previously mentioned experiments. After establishing that the proteins were stable on the simulated time scales, we focused the analysis on the organization of the ionic liquid

substituents about the protein surface. Calculated solvent charge densities show that for both enzymes and in both solvents that changing positively charged residues to negative charge does indeed increase the charge density of the solvent near the protein surface. The radial distribution of IL constituents with respect to the enzyme reveals decreased interactions with the anion are prevalent in the modified systems when compared to the wild type and this is often accompanied by an increase in cation contact. Finally, the radial dependence of the charge density and ion distribution shows no support for the hypothesis of long range structural ordering of the IL at 20 wt.% concentrations; both metrics show that the IL converges toward bulk-like configurations within the first 1.0 nm from the protein surface.

Chapter 5 is an article I prepared with Prof. Pfaendtner and our collaborator in the Department of Emergency Medicine, Dr. Nathan White. This article was published in *PLOS One* [33] in January, 2014. While this project did not involve ILs, it details successful application of the enhanced sampling techniques that have become a key component of our group's standard methodology. Additionally, as part of an underlying theme of my project – observing conformational change in proteins – we found that even a single atom can create differences in protein structural preference that can be observed using multiscale molecular simulation. The motivation behind this project involved the observation that oxidation of key methionine residues on fibrin leads to altered fibrin polymerization, producing severely altered fibrin gel structure and function. This is important because fibrinogen and its modification by oxidative stress have been implicated as key contributors to both pathological thrombotic and hemorrhagic diseases ranging from cardiovascular thrombosis to the acute coagulopathy of trauma. However, how oxidation leads to altered fibrin polymerization remains poorly understood at the molecular level. We applied a powerful and novel well-tempered ensemble parallel tempering (PT-WTE) technique along with conventional molecular dynamics (MD) simulation to investigate the molecular-level consequences of selective methionine oxidation of fibrinogen. We offer new insights

into molecular mechanisms of oxidation-induced changes in fibrin polymerization, while focusing on the D region knob ‘B’ and hole ‘b’ interaction and αC-domain interactions, both of which are hypothesized to contribute to the lateral aggregation mechanism of fibrin fibrils. Methionine oxidation did not alter the native state or the stability of a bound knob ‘B’ surrogate when interacting with hole ‘b’ in the D region. However, applying PT-WTE simulation to a human homology model of the bovine N-terminal subdomain fragment from the αC-domain revealed that methionine oxidation altered the conformation of the hairpin-linking region to favor open rather than closed hairpin structures. We attribute this alteration to the disruption of the hairpin-linking region’s conformation, with oxidation increasing the radius of gyration for this segment. This result is in agreement with experimental data demonstrating decreased fibrin protofibril lateral aggregation when methionine oxidation is present in the same αC-domain fragment. Therefore, single methionine oxidation within the αC-domain is a likely molecular mechanism.

2 Structural and Dynamic Features of *Candida rugosa* Lipase1 in Water, Octane, Toluene, and Ionic Liquids BMIM-PF6 and BMIM-NO3

2.1 Introduction

Ionic liquids (ILs) are a novel alternative to organic solvents for many areas of chemistry. Specifically, when used as a solvent for biochemical reactions they have the potential to add control over many reaction properties, including selectivity and product solubility. Commonly reported applications of ILs in biocatalysis include the pretreatment and catalytic breakdown of lignocellulosic biomass [6-8] and the enzymatic degradation of food waste oils [9,10]. One exciting concept regarding IL research is the wide variety of cations and anions capable of maintaining liquid phase at near-ambient conditions, providing a large library of possible solvents. It is also feasible to chemically tailor IL ions, fine-tuning atomistic functionality toward desired solvent properties [5]. Another aspect of some ILs is their ability to support enzymatic catalysis at a wide range of temperatures or concentrations of water. Not only do many enzymes retain activity in pure ILs [14], but also exhibit enhanced thermal stability [15,16], selectivity [17] and enantioselectivity [18,19] compared to traditional solvents.

Little is currently known about the molecular level change that ILs induce in an enzyme, especially those changes that lead to altered enzyme activity. The interactions between the IL and the enzyme structure may alter the equilibrium structure or dynamic fluctuations. The link between these interactions and reaction rates [25] and thermal stability remains an area of great interest. For example, *Candida rugosa*'s Lipase 1 enzyme has been shown to retain activity in the IL 1-butyl-3-methylimidazolium hexafluorophosphate (BMIM-PF_6) yet is rendered inactive in BMIM nitrate (BMIM-NO_3) [34]. However, the current literature does little to explain how the NO_3 anion plays a role in deactivating the enzyme [34] or how BMIM-PF_6 can stabilize the enzyme at higher activity and

temperatures than many other ILs [16]. The lack of experimental evidence can mostly be attributed to the inapplicability of many common experimental techniques, including in solution NMR and circular dichroism, in near-pure ionic liquids. In addition to the difficulty of purifying enzymes and the elimination of non-essential water from the solution, these limitations have restricted many studies to focus on discovery of ideal cation-anion pairs and characterizing enzyme fixation without atomistic insight guiding their search.

Molecular dynamics (MD) simulation offers an alternate way to gain molecular scale insight of the interactions between non-aqueous solvents and enzymes that to date has not been widely exploited. MD now offers access to both the dynamics and solvated structure of enzymes up to thousands of amino acids in a wide variety of solvent types at the fs to μs timescale. Not only are these time and length scales below the limit of most experimental techniques, the atomic details from all-atom MD at long times scales may provide a wealth of information about cation-anion [35] and protein-ion interactions [36,37]. Applying MD simulation toward IL research may allow for simulated screening of ILs; virtually sifting through the countless anion-cation combinations toward discovering the limits of enzyme catalyzed reaction conditions and generally optimizing IL enhanced biocatalysis.

Several groups have spearheaded the use of simulation as a tool to investigate IL thermodynamics. Their efforts have included the first stages of developing and testing IL force fields [38,39] and have shown good accuracy when calculating the properties of condensed phases [40,41]. Furthermore, several MD studies investigating the behavior of enzymes, including *Candida antarctica* lipase [42] and the cutinase [43] from *Fusarium solani pisi*, have explored the affect various water dilutions of ILs on ion affects at short time scales. Building from these previous works, we have completed a series of MD simulations for the purpose of investigating the affect of non-aqueous solvents on the behavior of *Candida rugosa* lipase 1 (CRL).

CRL was chosen as a model system for both for its industrial importance—perhaps in the future with ILs—its presence in the IL-biocatalysis literature, and its wide variety of structural motifs compared to most "toy" systems. The enzyme itself consists of 535 amino acids in an alpha-beta hydrolase fold. Features of the protein include a hydrophobic core, a variety of beta sheet and helix secondary structure, and a ~25 residue "lid" (highlighted in Figure 2.1) which covers the active site when "closed" and can spontaneously open to expose the active site's catalytic triad to the solvent. Not only does this opening transition require a proline isomerization at P92, one of the lid's hinge residues, but the lid "bottom" and the covered portion of the active site consist primarily of hydrophobic residues. The opening transition results in a 1,526 $Å^2$ increase in solvent-accessible surface area, of which 892 $Å^2$ is due to exposing hydrophobic residues [44]. These features indicate that solvation likely plays a role in the transition, which coincides with prevailing theory in the literature [45,46] – the lid-opening event is thought to occur at the presence of phase boundaries and hydrophobic environments.

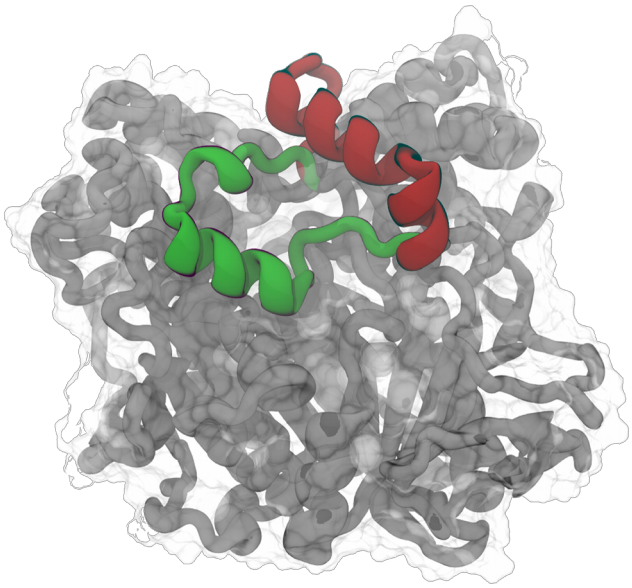


Figure 2.1: Cartoon representation of the two crystal structures used as the CRL initial coordinates when starting from both the open (red) and closed (green) lid conformation. The non-lid region of the structures (grey) was used for alignment.

Specifically, in this study we compare the influence of four non-aqueous solvents, including two ionic liquids: BMIM-PF_6 and BMIM-NO_3; and two non-polar organic solvents: toluene and octane, on CRL at near-ambient conditions. We have also included the simulation of CRL in water for comparison. The ionic liquids BMIM-PF_6 and BMIM-NO_3 were chosen because experiments have demonstrated retained activity for CRL in BMIM-PF_6, yet no activity with the nitrate anion [16,34]. The

specific mechanism of deactivation in BMIM-NO_3 has yet to be demonstrated; although other ionic liquids have been shown with FTIR or fluorescence spectroscopy to cause protein denaturation [47,48], the case of the NO3 anion with BMIM has not been well investigated. It is unreasonable to expect denaturation of this large of an enzyme within typical simulation time scales. For this reason we have included reference simulations at a higher temperature, 375 K, which is well above the unfolding temperature in water [49], but not above the boiling point of our water model. This higher temperature will also serve to enhance sampling of slow protein modes for all solvents. The two non-polar solvents were included to give contrast to both the highly charged as well as the highly viscous nature of the ILs. Lastly, we have completed these ten simulations (five solvents at two temperatures each) starting from both the closed (CRLC) and open (CRLO) crystal structures for a final total of twenty independent MD simulations and 2.8 μs of total simulation time.

2.2 Methods

2.2.1 Model Systems

CRL exhibits the characteristic alpha/beta hydrolase fold of many hydrolase enzymes, including a HIS-SER-GLU catalytic triad and an oxyanion hole. Protein Data Bank [50] entries 1TRH [44] and 1CRL [51] were used as the initial closed and open structures, respectively, for CRL. Comparison between these two structures shows very little (<0.3 Å Cα RMSD) difference in the protein bulk and a large (17 Å Cα RMSD) variation in the lid region, residues 66-92. Aligning the two structures by the lid residues shows that lid closure is not a simple hinge-like rigid body movement and includes both rearrangement of secondary structure and a high-energy barrier in a cis-trans proline isomerization at P92.

Of the two structures, 1CRL is resolved at a higher resolution and contains 310 crystallographic waters and information of glycosylation, in the form of several N-acetyl glucosamine (NAG), at

asparagines N314 and N351. The N351 glycosylation, which contacts and forms a hydrogen bond with the lid in the CRLO crystal structure, may stabilize the lid in the open state. For this reason, the NAG residues were left intact for simulations started from the open structure and added to the closed crystal structure for the other simulations. NAG residues were parameterized with the GLYCAM06 [52] force field. All structural waters were retained from the crystal structure for all simulations, including the non-aqueous simulations, as they have been shown critical toward maintaining enzyme activity [53]. De la Casa et al. have shown that lyophilized Lipase 1 contains between 220 to 522 waters per Lipase molecule [54]. Therefore, to help make connection to experiment, the surface waters from the crystal structure were also maintained, bringing the total number of water molecules in the system to 310. The water simulations were prepared by enclosing the protein in a pre-equilibrated TIP3P [55] water box using Ambertools [56].

IL and organic solvent molecules were parameterized in antechamber [57] using GAFF force field [58] and RESP [59] partial charges after a single point Gaussian 09 [60] calculation at the HF-6/31-g*(d) level of theory. Only the hexafluorophosphate from BMIM-PF_6 could not be parameterized with GAFF and antechamber. Instead, the hexafluorophosphate molecule was treated as rigid at the bond lengths angles determined from a Gaussian 09 geometry optimization at the same level of theory. The force field parameters were tested against known experimental observables such as density and heat capacity (details in the Supporting Materials, Appendix 1). Initial coordinates for pure IL and organic solvent boxes, of similar dimension to the water simulations, were generated in Packmol [61]. These boxes were equilibrated using a 5 ns NVT simulation at 500 K temperature followed by slowly quenching to simulation temperature and a final 10 ns NPT equilibration. The protein and crystallographic waters were then placed in the center of these equilibrated boxes and non-water solvent molecules within 3.0 Å of the protein were deleted. All the simulation boxes were neutralized with

counter-ions by replacing non-crystallographic solvent molecules. For the simulations, the number of anions and cations were equalized by random deletion of whichever were in excess. The number of atoms and the average equilibrated box size for each system is reported in the SI.

2.2.2 Simulation Protocol

All MD simulations and minimizations were performed using NAMD 2.7 [62] and the protein was parameterized with the Amber 99SB force field [63]. First, conjugate gradient minimization was carried out with all atoms except the solvent frozen for 2000 steps; then all atoms except the protein and NAG residues frozen for 1000 steps; and finally the entire system mobile for another 2000 steps. The systems were then gradually heated from 0 to either 310 or 375 K using velocity rescaling NVT simulation to increment the temperature by 10 K every 200 ps. Production simulations were carried out in the NPT ensemble using the Langevin thermostat [64] and Langevin piston barostat [65]. During MD simulations a 2 fs time-step was used; the non-bonded forces were cutoff at 12 Å with a switching function starting at 11 Å; PME was used to reincorporate long-range electrostatics; RATTLE [66] and SHAKE [67] were used to freeze hydrogen bonds. The NPT simulations used a dampening coefficient of 5.0 ps^{-1} for the thermostat and 200 fs for both the barostat's period and decay.

The final simulation time for all non-aqueous simulations were chosen to be 100 ns after the protein relaxed (typically occurring during the first 2-10 ns), as judged using the protein's Cα RMSD. The water simulations were taken to longer time because of the interesting behavior in the lid residues, as detailed below. The water simulations at 310 K starting from CRLO and CRLC crystal structures were taken to 600 ns and 250 ns, respectively; while at 375 K they were taken to 200 ns each. All analysis was completed with either VMD [68] or the MDAnalysis [69] package for python.

2.3 Results & Discussion

2.3.1 Lipase Structure

Cα root mean squared deviation (RMSD) of the entire protein compared to the simulation's starting structure was used both as a measure of protein stability and as a means to monitor significant changes from the crystal structure. The lid-closed position is known to be preferred when CRL is isolated in water. Based on this observation, the water simulation started from the CRLC crystal structure at 310 K is considered our basis for CRL stability. The enzyme relaxes to a RMSD of about 1 Å during the first 10 ns of MD simulation, showing the enzyme backbone remains close to the initial structure upon heating. Similar behavior is seen for the 310 K simulations (Figure 2.2A) in the other four solvents. After 100 ns of NPT simulation at 310 K, the IL and water simulations final RMSD are nearly indistinguishable at about 1 Å, while the two organic solvents are slightly higher, octane being the highest at close to 2 Å.

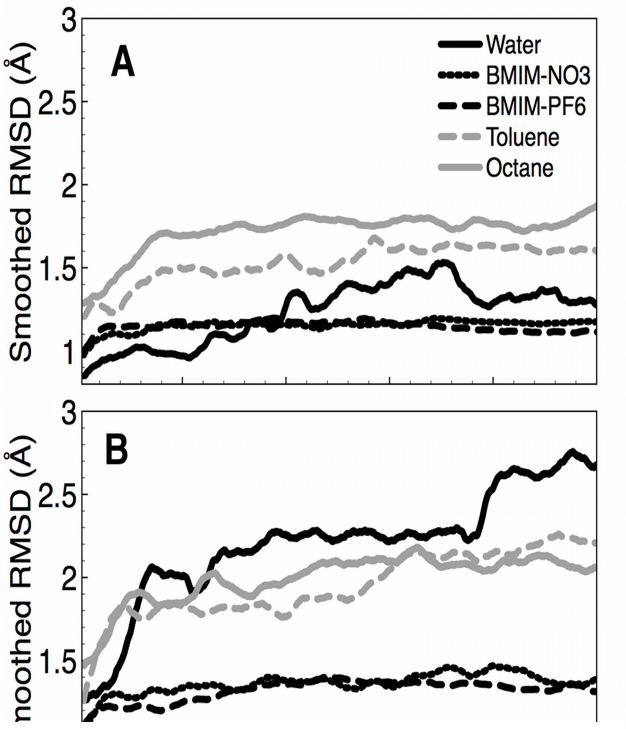


Figure 2.2: Smoothed Cα root mean squared deviation from the closed crystal structure for simulations started from the closed crystal structure at both 310 K (A) and 375 K (B). The raw data is not shown for clarity.

The low RMSDs from simulation of other enzymes in these same ILs has been reported previously at the 10 ns time scale [43] and was concluded to be caused by "kinetic trapping". This describes how the enzyme exhibits little conformational deviation from its starting structure during any of the MD simulations. Although the simulations at high temperature show the same low RMSD (**Error! Reference source not found.**Figure

2.2B), the kinetic trapping previously proposed cannot be excluded as the likely cause. However, in contrast to the suggestions of Micaêlo et al. [43], we conclude that this effect is caused by the transport properties of the ILs rather than the lack of water near the protein surface, since the surface and crystallographic waters were retained in all our simulations. Furthermore, the same behavior is also present in the IL simulations started from CRLO, indicating that the trapping is independent of exposed hydrophobic surface area and preferred native state in IL. Overall, for the chosen IL and enzyme combination we do not observe any significant structural changes on the 100 ns timescale.

In contrast to the IL simulations, the simulations in both organic solvents exhibit relaxed structures further from the initial structure and show increased RMSD at high temperature, much like the simulations in water. There are numerous characteristics that differentiate the organic and ionic liquid solvents. Among these, density, viscosity, charge and hydrogen bond participation are of particular relevance to simulation. Although experiments indicate a large decrease in viscosity for ILs at higher temperature, their viscosity is still several orders of magnitude higher than either water or the organic solvents. It is possible that the high viscosities inhibit movement away from the starting structure at the simulated timescale. This effect could further be exacerbated by the anions and cations of the ionic liquid coordinating with the sidechains at the enzyme surface, which will be discussed below.

2.3.2 Lipase Dynamics

Cα root mean squared fluctuations (RMSF), measured from the corresponding initial structure, were calculated for each trajectory after the simulations equilibrated. The systems were considered equilibrated when the RMSD leveled off after the characteristically sharp increase during the first several nanoseconds of MD simulation. For the non-aqueous simulations, this usually occurred after 2-10 ns; the aqueous simulations were judged as equilibrated after 10-20 ns, as apparent in Figure 2.2. To eliminate sampling as a factor when comparing between simulations, each simulation was extended for

at least 100 ns after equilibration. For consistency, all RMSF values and plots discussed herein are measured over the first 100 ns of simulation after equilibration.

RMSFs for the water simulations (Figure 2.3A,B) reveal select regions of CRL that either underwent some rearrangement or exhibited high flexibility during the equilibrated trajectories. The largest peaks for the water simulations starting from CRLO belong to residues of the lid (residues 66-92). Even in the CRLC water simulations this region shows higher than average fluctuations, which demonstrates the highly dynamic character of the lid. Details of the lid conformational changes and dynamics will be discussed in detail below. Another region of interest for the water simulations are

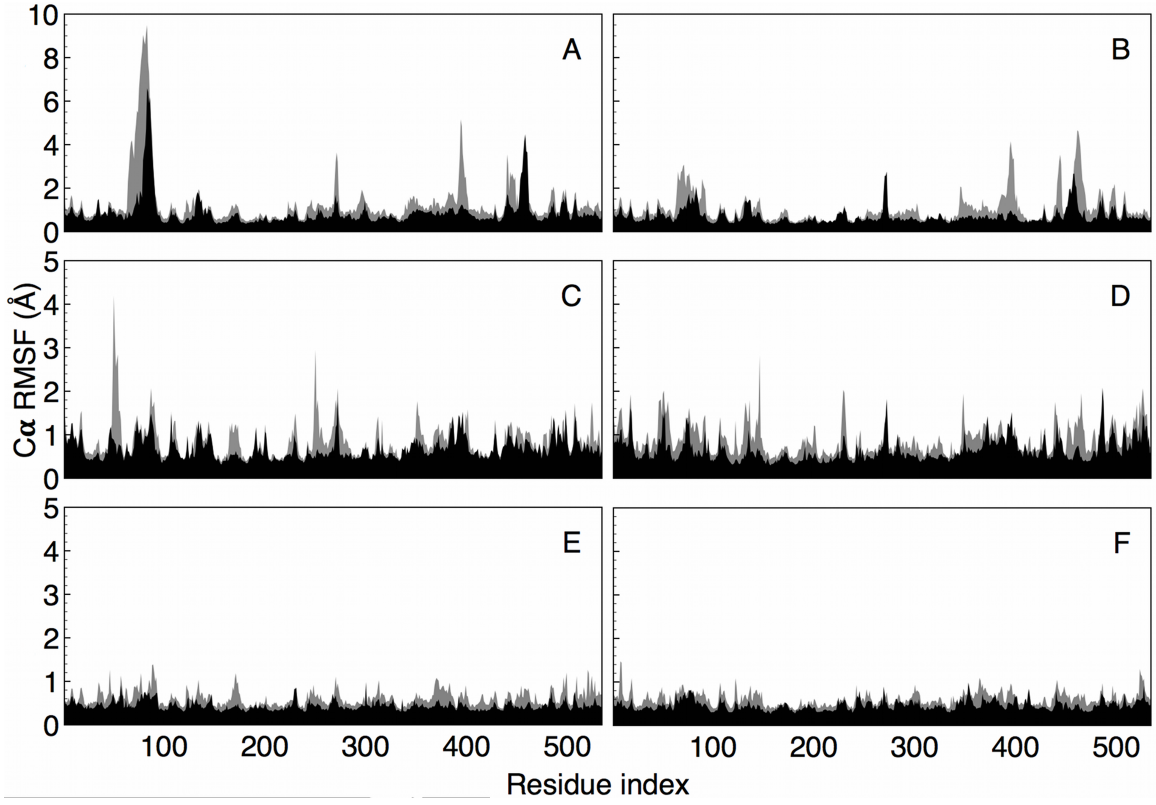


Figure 2.3: Cα root mean squared fluctuations of simulations started from the open (left column) and the closed (right column) crystal structure at both 310 K (black) and 375 K (grey). The water simulations (A,B) show high fluctuations in the lid and the regions of the protein that contact the lid when it is closed. In octane (C,D) these fluctuations are somewhat dampened, but in BMIM-PF6 (E,F) these fluctuations are no longer distinguishable. Note the difference in scale between row 1 and rows 2-3.

residues 430-475, the residues that contact the lid in CRLC and form the rim of the active site pocket. Other regions of high RMSF are the exposed random coils of surface residues 268-273 and 391-399.

In contrast to the water simulations, RMSF from simulations in the ILs (Figure 2.3C-F) exhibit few pronounced features and an overall restrained movement compared to water or organic solvents. This dramatic reduction in RMSF is seen throughout the entire enzyme. Furthermore, in the IL simulations the lid region and the region surrounding the binding site are nearly indistinguishable by RMSF regardless of starting structure or simulation temperature. Although not shown here, the RMSF plots for BMIM-NO_3 are similar to those shown for BMM-PF_6 – dampened and, relative to the organic or water simulations, featureless. Results from the organic solvent simulations manifest this dampening of the lid and other regions but to a lesser degree. The RMSF for these systems also show increased fluctuations in residues 49-54, a heterogeneous patch of residues on the surface of the protein adjacent to the lid. These fluctuations are the highest in the simulation starting from CRLO in octane at 375 K, during which these residues undergo a minor conformational change which places a group of 5 polar residues into close contact and exposes a phenylalanine into the solvent.

While the RMSF is a measure of backbone flexibility, it displays no information of side chain fluctuations. To assess the accessibly of side chain rotamers in the various solvents, we calculated the statistical entropy of the first side chain dihedral (χ_1) as: $S_j = -R \sum_i P_i \ln P_i$. In this equation, S_j is the statistical entropy for residue j and P_i is the probability residue j having a χ value of i [70]. The statistical entropy of set of values (S) is a measure of the heterogeneity of the set; a data set with values spread evenly over a wide range will show a higher value of S compared to a set with all the data contained in a narrow range. Extending this concept to a trajectory of side chain dihedrals results in highly flexible side chains possessing high S values while restricted or confined side chains showing lower S values. In this

case we have calculated the entropy of the χ_1 dihedral angle, which is formed by atoms N-C_α-C_β-C_γ, for residues on the surface of CRLC.

The statistical entropy for the χ_1 dihedral for all the surface residues were calculated over the first 100 ns of equilibrated trajectories for simulations started from CRLC. Using paired z-tests, the sets of entropy values for each simulation were found to be significantly different from those of the water simulation. For example, Figure 2.4 shows that for almost every surface residue the χ_1 entropies are higher in water than BMIM-NO_3. Further comparison between the different simulations showed that at 375 K all data sets are statistically different except when comparing between the two ILs. The trend in mean entropy shows that side chain flexibility decreases as water > octane > toluene > BMIM-PF_6 > BMIM-NO_3. This series was shown to remain unchanged when only considering polar residues (N=147), yet changes to octane > water > toluene > BMIM-NO_3 > BMIM-PF_6 when considering only hydrophobic surface residues (N=81). Considering both the lower RMSF and the decreased statistical entropy, the simulations indicate not only dampened backbone dynamics but also dampened side-chain dynamics. It is unclear whether these dampened dynamics are caused by the IL's high viscosity, coordination between the ions and amino acid residues, or some combination of these effects. The viscosity of the solvents themselves increase as octane < toluene < water < BMIM-NO_3 < BMIM-PF_6. It

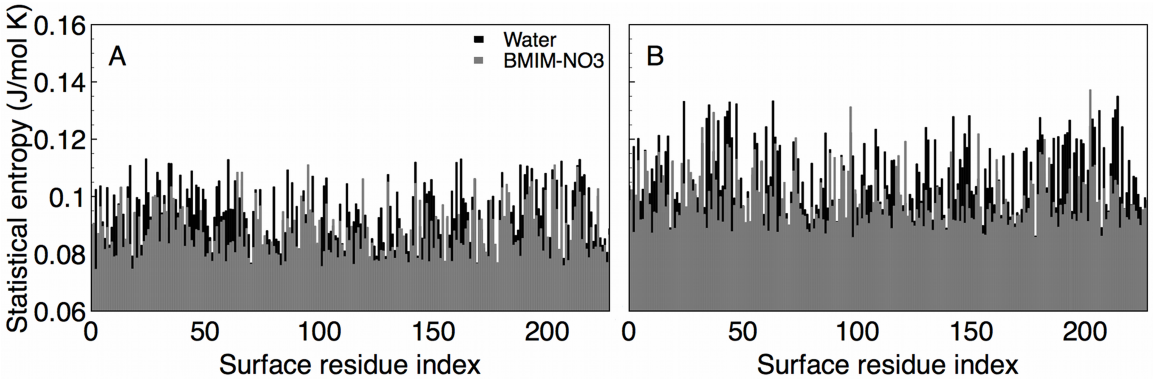


Figure 2.4: χ_1 statistical entropies for the surface residues of the protein are shown for both water and BMIM-NO_3 simulations started from the closed crystal structure at both 310 K (A) and 375 K (B).

is interesting to note that BMIM-NO_3 has a viscosity two order of magnitude higher than either water or the organic solvents, while BMIM-PF_6 has an even high higher viscosity than water or the organic solvents. With this consideration we might make the assumption that the difference in viscosity between water and the polar organic solvents is negligible. In this case we observe that the trend follows solvent viscosity when only considering hydrophobic residues but not when considering all surface residues.

2.3.3 Surface Behavior

With our inclusion of the 310 crystallographic surface/structural waters in the 20 simulations, we have an opportunity to examine the behavior of this water over the 100 ns timescale. Tracking the number of crystallographic waters within 4 Å of any protein atoms (Figure 2.5) shows a clear difference between each solvent. The water simulation shows a loss of approximately half of the initial water after the system heating, prior to NPT simulation. Octane and to a lesser extent toluene have the ability to trap water at the protein surface. Observing the final few nanoseconds of these simulations reveals that not only are both the water and neutralizing sodium ions staying close to the protein surface, but they are partitioned away from non-polar residues and kept in close contact with polar sidechains. This is even more apparent in the CRLO simulations due to the increased amount of exposed nonpolar surface area (Figure 2.6). An interesting aspect of the toluene simulations, at both low and high temperature, is that ~30 waters begin to diffuse away from the protein surface after a 39 ns lag time at 310 K and 17 ns lag time at 375 K. Compared to the octane simulations,

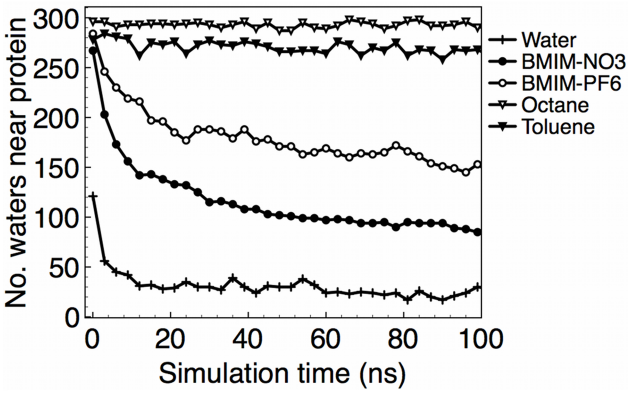


Figure 2.5: The number of crystallographic waters remaining within 4 Å of any protein atoms over the course of 100 ns of NPT simulation is plotted for the simulations starting from the closed crystal structure and at 310 K. This plot for 375 K (not shown) is almost identical.

the water and ions remaining near the protein surface are not as close to the protein surface although are still partitioned away from nonpolar residues.

This water/ion trapping behavior is not present in the IL simulations, including BMIM-PF_6, which is known to be hydrophobic and immiscible with water. As apparent in Figure 2.5, the IL trajectories reveal a slow and steady diffusion of water away from the protein surface. Given the apparent equal or stronger interaction of charged IL solvent with sites occupied by the crystallographic waters, one limitation to this diffusion appears to be the viscosity of ILs instead of protein-water interaction. The trend of final water count follows the trend of viscosity and miscibility for the ILs, with BMIM-NO_3 having a much lower viscosity and greater miscibility than BMIM-PF_6. Further analysis of the IL trajectories shows that the surface waters do not exhibit the same partitioning as seen in the nonpolar solvents.

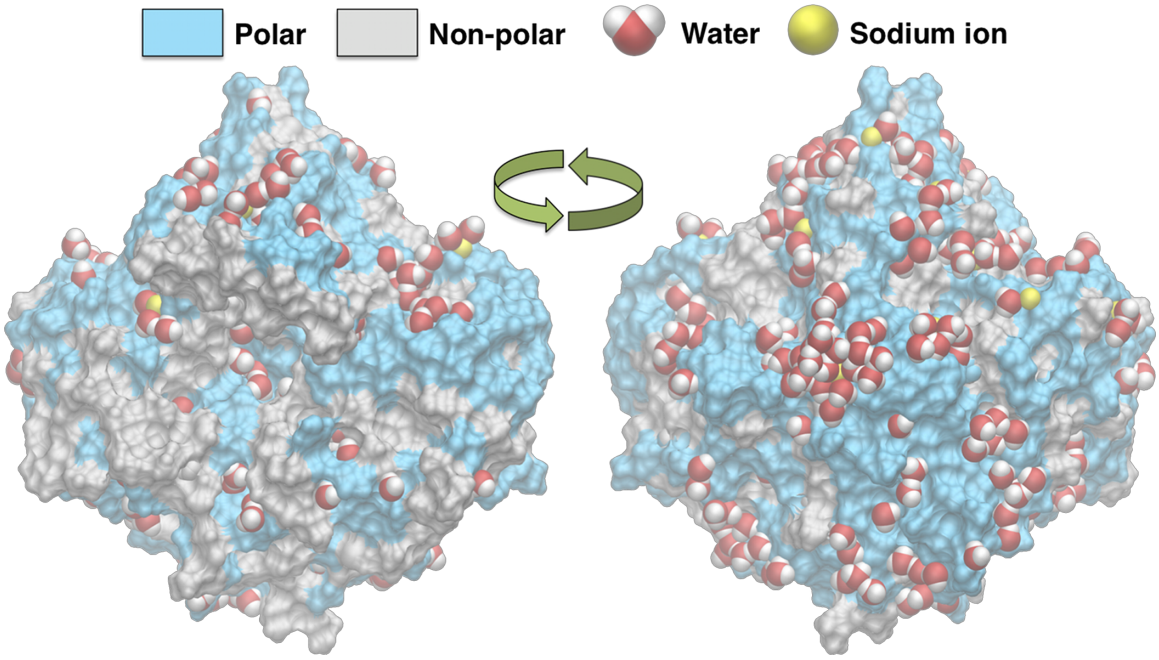


Figure 2.6: Protein representations showing surface solvation partitioning in Toluene. Snapshots of the final coordinates from the simulation started from the open crystal structure at 310 K in toluene shows that both water and sodium ions are partitioned away from hydrophobic residues. The two snapshots show the open lid face (left) and the opposite face (right) of the protein.

We next calculated the trajectory of contacts between protein surface residues and solvent molecules for the set of simulations at 310 K starting from CRLO. This data was sorted by residue type and residue classification (i.e., polar or nonpolar) and inspected for apparent trends. Confirming the previous analysis and the trapping effect, the data for the protein in octane shows overwhelming contact with water across all residue types. In fact, 79% of closest-contacts made by the protein surface over the entire simulation are with water. When limiting this consideration by residue classification, the proportion changes to 73% for nonpolar and 83% for polar residues. Polar residues make up the majority of contacts for octane molecules, while sodium ions are heavily clustered around the negatively charged glutamic and aspartic acid. Comparison of the data sets for toluene and octane show that toluene can penetrate to the protein surface much more effectively than octane. Specifically, the proportion of total contacts made by water drops to 64% and the proportion of contacts made between nonpolar residues and the organic solvent molecules is 54%, compared to 24% in octane.

Applying this same analysis to the IL simulations reveals more interesting differences. For the BMIM-NO_3 system, water is replaced by the nitrate anion in making the majority of contacts with the protein surface. The most numerous contact for the nitrate anions are atoms from polar residues. Surprisingly, nitrate is able to outcompete every other solvent molecule for all contact types, including the nonpolar residues, yet not including negatively charged residues. The like-like repulsion between nitrate and negatively charged residues is enough to reduce the number of nitrate contacts to 5% for these residues. Comparison between the BMIM-PF_6 and BMIM-NO_3 simulations show few differences. The most significant change in proportion of contacts occurs in the negatively charged residues of which the hexafluorophosphate makes up 16% of the contacts as compared to the 5% for nitrate. The BMIM-PF_6 simulations also show a slight decrease in the number of contacts between BMIM and protein surface. Since there is no apparent increase in the number of anion-protein contacts compared to nitrate,

it is difficult to conclude that the size of the anion is helping to exclude BMIM. The dominant coverage of the protein surface by anions may indicate that the rigidity and size of the BMIM cation restricts its ability to fill the small spaces near the side chain residues. Furthermore, anion preference at the surface may lead to stratification of the IL in order to avoid buildup of charge gradient in the solvent.

2.3.4 Lid Behavior

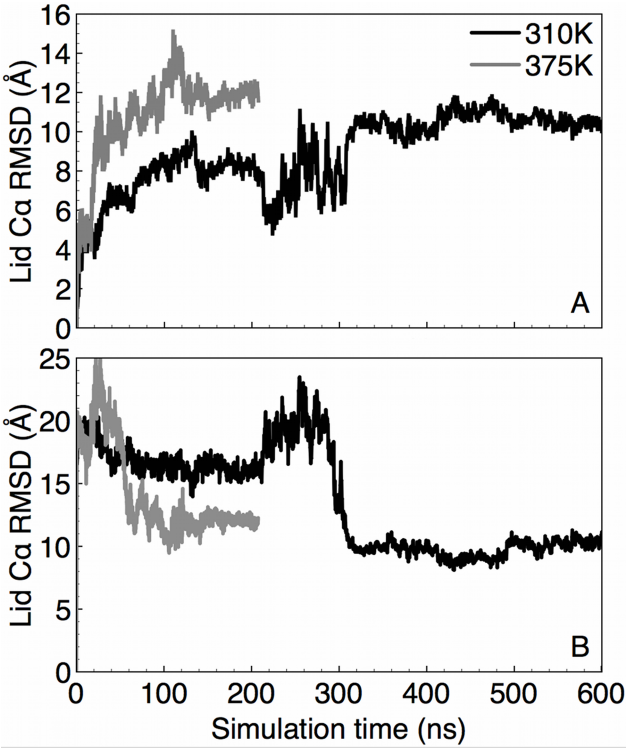


Figure 2.7 Cα root mean squared deviation of the lid residues compared to the open (A) and closed (B) crystal structure after aligning to the non-lid residues are shown here for simulations stared from both the open crystal structure. The high temperature simulation shows similar motions on a significantly shorter time scale.

The increased RMSD of CRLO in water at both temperatures (not shown), and CRLC at 375 K (Figure 2.2B) are mostly attributable to the lid region. The lid region in all water simulations shows the most dynamic fluctuations, as seen by the RMSF in Figure 2.3; both simulations starting from CRLO in water exhibit the highest RMSF and RMSD of any of the 20 simulations. The RMSD of the lid residues calculated from the closed crystal structure can be used to measure closing movement for simulations started from the open crystal structure. Likewise we can calculate the lid RMSD from the open crystal structure to measure lid opening for simulations started from CRLC. Based on the lid RMSD (Figure 2.7) and visualization of the trajectory, the CRLO in water simulation at 310 K shows the lid initially opening further and making contact with another region of the enzyme bulk then undergoing some secondary structure rearrangement before closing down over the active site after ~300 ns of total simulation time.

After witnessing the drastic change in RMSD, we extended this simulation to a total of 600 ns to check the stability of this new structure. Figure 2.7B, which shows the lid RMSD compared to CRLC for these simulations indicates that after the large closing movement around 300 ns, the lid remains stable about 10 Å RMSD from CRLC. Viewing the final 100 ns of trajectory for this new closed state shows that there are many missing hydrophobic contacts between the lid and active site, and that while the lid makes a larger footprint on the enzyme bulk, the new closed structure only achieves partial coverage of the active site. Indeed, visualization of the trajectories allowed us to witness free transport of water molecules into and out of the active site after the lid reached this new state. Examining the trajectory of the P92 ω dihedral also reveals that the isomerization did not occur during this transition; P92 remained trans in this new closed state. These results may show that the isomerization is not a requirement for the lid to cover the binding site in water, but it is still unclear whether it is required to achieve complete active site occlusion. In addition to this 310 K closing event, another closing event occurred after only 100 ns in the 375 K simulation of CRLO in water. The final state for the higher temperature simulation is different from both CRLC and the final structure from the lower temperature. The P92 isomerization also did not occur for this high-temperature closing event. Renderings of example frames for these closed structures have been included in an appendix (Appendix 1, Figure A1.1) for reference.

We stress that in addition to the bulk hinge-like movement of the lid there is a concomitant change in the lid's secondary structure, as can be seen in Figure 2.1, required for complete lid closure. Furthermore, the P92 trans to cis transition to move between the open and closed structure, respectively, is a major obstacle for molecular simulation due to the high energy barrier for proline isomerization (~20 kcal/mol) [71]. It is has also been shown previously that the preference for either the cis or trans configuration changes with the properties of the solvent [71]. Furthermore, comparing the CRLC and

CRLO structures shows an increase of 1,526 $Å^2$ in solvent-accessible surface area upon opening, of which 892 $Å^2$ is due to exposing hydrophobic residues [44]. Based on the literature and these factors, we might have also expected an opening event in the non-polar organic solvents. However, there was effectively no deviation in the lid configuration in any of the organic or IL simulations.

In contrast, the high temperature CRLC in water simulation did exhibit some interesting behavior (c.f., Figure 2.2B). Visualizing the trajectory and the final structure revealed that the lid underwent some structural rearrangement and moved along the surface of the protein without opening (example structure render in Figure A1.1). Comparing the final structure from this simulation to the other closed structures revealed it to be ~4 Å RMSD from the final structure of the CRLO in water at 375 K simulation. Visual comparison of the two structures shows remarkable similarity, with the biggest difference being in the secondary structure of a small portion and the backbone curvature near the P92 hinge, most likely caused by the difference in the ω dihedral. Considering the similarity between these structures, we have sampled an almost complete transition between the two crystal structures at 375 K. The missing components of this transition include proper placement of several hydrophobic residues and the isomerization of P92.

2.4 Conclusions

With MD we studied the affect of non-aqueous solvents on lipase 1 of *Candida rugosa* at both 310 and 375 K starting from both the open and closed crystal structures. We analyzed the change in protein structure, fluctuations, the enzyme solvation, and the lid conformation. This work provides a perspective on the differences between the five solvents at the 100 ns time scale and probes aqueous conformational changes on up to the 600 ns time scale. While we can see clear differences between the enzyme's behavior when comparing between water and the ILs, at the 100 ns time scale there is little demonstrable difference between the two ILs. This observation stands despite strong experimental

evidence of differences in CRL activity within these solvents and the validity of the force fields used. For the pair of ILs, it is clear that the characteristic fluctuations in the protein structure and the protein's ability to relax from its initial structure are severely reduced. There is also no evidence at this time scale of BMIM-NO3 causing the protein to denature. These observations lead us to conclude that the study of these ILs requires significantly longer sampling, which we hypothesize is due to 1) the intrinsic transport properties of the solvent and 2) the strong interactions between the protein surface and IL anions. For this reason, future MD simulation work of proteins in highly concentrated ILs would be advised to use longer simulation times or to employ some enhanced sampling methods, such as parallel tempering. Denaturing is only one of four likely mechanisms by which ILs might deactivate enzymes, the other three being: inhibition by solvent molecules, covalent interactions between the IL and enzyme, and enzyme aggregation. While the latter two would be difficult to predict using MD, the former could be investigated using a substrate docking study with biased simulation such as umbrella sampling or metadynamics.

Our analysis of solvent-protein contacts demonstrates how even between two similar solvents, such as BMIM-NO3/PF6 and octane/toluene, there can be large differences in the solvation of the protein surface. The hydration (or lack thereof) of biomolecules in nonaqueous media continues to be an important point of discussion in the literature, which many researchers use to propose detailed mechanistic insights (e.g., the role of an "ice like layer of water on proteins"). Our simulations are able to discriminate subtle differences between similar solvents, and MD simulation of biomolecules in solvents using general force fields appears to be a viable way to gain qualitative insights. However, creating classical IL force fields that faithfully reproduce all structural and dynamic properties continues to be a challenge and future studies will need to couple detailed experimental observations of

IL/biomolecules with simulations (e.g., coupling MD with small angle neutron scattering) to ensure quantitative accuracy.

While it has been described in literature previously [72,73], we have sampled several large conformational changes in the CRL lid over the course of multiple simulations. In simulations starting from the open crystal structure in water the lid closing down over the active site characterizes these conformational changes. Though these trajectories contain the lid swinging closed to the protein surface, they fail to reach the same conformation as the closed crystal structure also do not exhibit the P92 cis-trans isomerization evident in the crystal structures. However, we can use our nearly complete lid closing trajectories to generate an initial lid-closing pathway for using path collective variable metadynamics. Results from such simulations may provide key details in the lid closing mechanism and reveal the preferred structure of the lipase lid in non-aqueous solvents.

3 Comparison of Three Ionic Liquid Tolerant Cellulases by Molecular Dynamics

3.1 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 [74]. 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 [75,76]. 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. [77]. 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 [78,79]. 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.

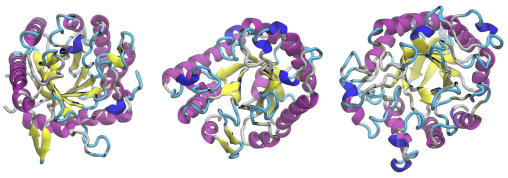
Many researchers have demonstrated that ionic liquids (ILs), organic salts that have melting temperatures near 100 °C, can completely solubilize cellulosic biomass – separating the hemicellulose and cellulose from lignin while overcoming the recalcitrance of crystalline cellulose [6,80,81]. In

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 [7,82-85].

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 improves our ability to control many properties of the reaction, especially substrate and product solubility. Not only do many enzymes retain activity in pure ILs [14], but also exhibit enhanced thermal stability [15,16], selectivity [17] and enantioselectivity [18,19] compared to traditional solvents. The potential to dissolve cellulosic biomass while maintaining hydrolase activity therefore 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.% [EMIM][OAc] in water [86]. 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 [20]. 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 [21].



Organism	T. viride (VIR)	T. maritima (TMA)	P. horikoshii (PHO)
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: Renderings of the initial configurations for each enzyme along with details highlighting their differences. The "optimum" pH and temperature, taken from Datta et al. [21] are conditions for maximum activity in water. The surface charge corresponds to the optimum pH and details of determining net surface charge can be found in the Methods section.

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] [21]. The three enzymes are: a cellulase from *Trichoderma viride*, cellulase 5A from *Thermotoga maritima*, and endoglucanase from *Pyrococcus horikoshii*. Hereafter referred to as VIR, TMA and PHO, respectively, structures and relevant details of these enzymes are provided in Figure 3.1. While 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.

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 [87]. 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 [88-92]. Our previous work in this area analyzed the stability and fluctuations of two different hydrolases', *Trichoderma longibrachiatum* GH11 and *Candida rugosa* lipase, structure in IL containing solvents [30,93]. 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 [94]. Herein, we will test some of these previous hypotheses and develop new hypotheses with different systems 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

While previous simulation studies have given some insight into the relationships between IL containing solvents and some specific enzymes' structures and dynamics, a comparative study to investigate IL effects across a family of enzymes has not yet, to the best of our knowledge, been conducted. In order 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 [21]. Specifically, we vary the enzyme species, temperature, and IL concentration in order to see how each affects the structure and dynamics of the protein.

3.2 Methods

All simulations were performed using AMBER compatible force fields implemented in GROMACS 4.6 [95] using the TIP3P water model [96]. The enzyme forces were calculated with AMBER99SB-ILDN force field [97], and the ionic liquid force field was adopted from our previous studies [30] and utilizes the general AMBER force field [98] and restrained electrostatic potential fitting for point charges [99] from quantum calculations using Gaussian 09 [60] at the HF/6-31g* level of theory. We scaled the charges by 0.8 in order to more accurately represent dynamic properties of the solvent, as has been demonstrated in the literature [100]. Cubic boxes of the solvent and protein were

generated using Packmol [101]. 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 freezing the bonds between hydrogens and heavy atoms using the LINCS algorithm [102]. 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 [103], and pressure was held at 1 bar using a Berendsen barostat [104].

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. The IL chosen for these simulations was [EMIM][OAc]. Other GHs have been shown to tolerate this IL [20,86,105]. We chose three different concentrations of IL that corresponded with those found by Datta et al. [21] 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. The density of [EMIM][OAc] is close to water, so the concentrations are quite similar. These three concentrations are important to illustrate the differences in the enzymes. First, 0% was chosen as a control. 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 [106]. VIR was modeled using an x-ray structure from *Trichoderma reesei* (PDB: 3QR3) [107]. GLU53 was mutated to ASP in order to match a known sequence for *T. viride* endoglucanase [108]. TMA and PHO were modeled using x-ray structures (PDB: 3AMD, 3AXX) [109,110] without mutation. Propka [111,112] 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. The surface charge of each enzyme noted in Figure 3.1 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 [113,114], then counting the number of positive and negatively charged surface residues at the pHs mentioned above. Root-mean-square deviation (RMSD), RMSF, and secondary structure analysis used g_rms, g_rmsf, and do_dssp from GROMACS 4.6, respectively.

3.3 Results & Discussion

3.3.1 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.

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. The structures of VIR and PHO remain close to their respective crystallographic structures. Only a few unstructured loops are displaced in the presence of ionic liquid. TMA, on the other hand, deviates noticeably from its

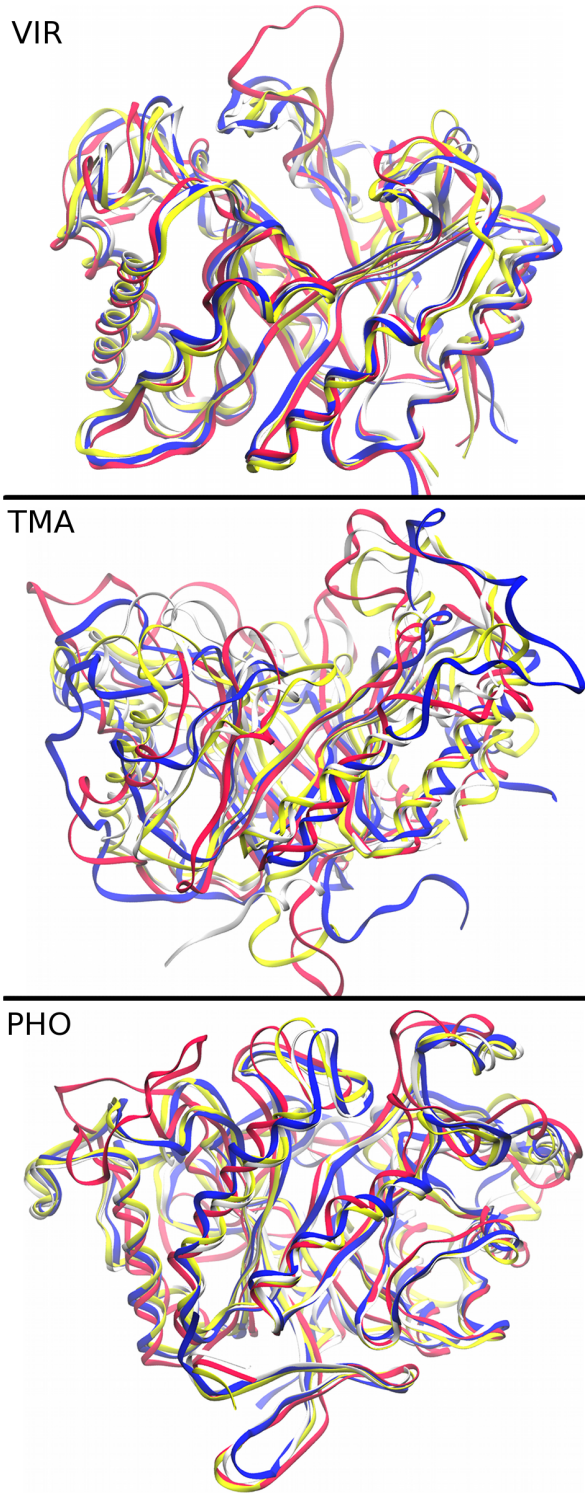


Figure 3.2: Structure of each enzyme after 500 ns of NPT simulation compared to the crystallographic structure after aligning all Cα atoms. Crystallographic – white. Water – yellow. 15% IL – Red. 50% IL – Blue.

crystallographic structure. 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.

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. One concern we noted in our previous studies was that the limited simulation time might have incorrectly indicated the degree to which the proteins were stabilized and we noted our expectation that increasing the sampling of our simulations would allow us to observe destabilization [29,30]. Table 3.1 contains key statistics summarizing the RMSD data, but does not explain some important observations we discuss below. Therefore we have included detailed plots exhibiting the raw RMSD data for each of the 54 simulations in Appendix 2 (Figures A2.1 and A2.2). For example, the large changes we observe in RMSD

are frequently characterized by "ballistic events", i.e., large structural changes over relatively short time periods, occurring after significant elapsed simulation time (hundreds of nanoseconds) for multiple trajectories. Based on the observations of the MD trajectories it is clear that these events are indicative of least partial protein unfolding. 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).

Table 3.1: Maximum, average, and standard deviation of three RMSDs for each of the simulated systems for all frames of three replicas combined.

T. viride	310 K			353 K		
(VIR)	Max (Å)	Avg (Å)	σ (Å)	Max (Å)	Avg (Å)	σ (Å)
Water	3.46	2.22	0.54	6.83	4.17	1.14
15% IL	4.04	2.41	0.57	6.55	3.44	1.31
50% IL	1.96	1.38	0.16	4.81	2.27	0.78
T. maritima	310 K			353 K		
(TMA)	Max (Å)	Avg (Å)	σ (Å)	Max (Å)	Avg (Å)	σ (Å)
Water	4.55	2.26	0.65	6.75	3.35	1.38
15% IL	4.84	2.56	0.95	9.63	4.69	1.25
50% IL	2.97	1.33	0.13	8.61	2.88	1.81
P. horikoshii	310 K			353 K		
(PHO)	Max (Å)	Avg (Å)	σ (Å)	Max (Å)	Avg (Å)	σ (Å)
Water	1.43	0.85	0.09	3.10	1.35	0.39
15% IL	1.36	0.88	0.09	6.16	1.61	1.32
50% IL	1.47	0.92	0.10	2.81	1.24	0.36

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 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 A2.2). 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.

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 while the largest departures 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. 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. Generally, enzymes 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 [29]. 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 [115,116].

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 ballistic unfolding events (c.f., Figures A2.1 and A2.2). Visualizing the trajectories shows that the ballistic events are preceded by the breaking

of key salt bridges within the protein and exposure of hydrophobic residues to the solvent. We hypothesize that the IL mediates this unfolding by stabilizing the exposure of sidechains previously involved in salt bridges internal to the protein.

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 3.3. 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 outlier in a 15% IL simulation in which an unstructured loop 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

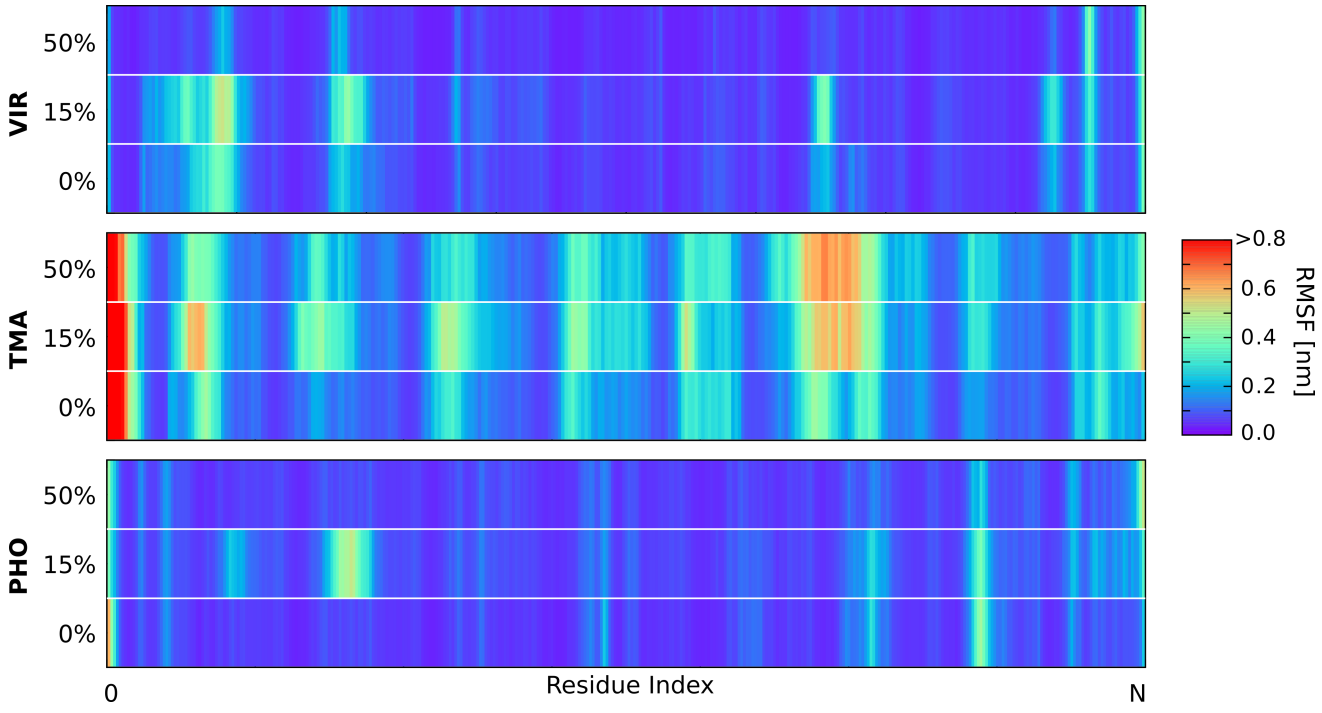


Figure 3.3: Root-mean-square fluctuations (RMSF) averaged over the triplicate simulations at 310 K for VIR and 353 K for TMA and PHO**.** N = number of amino acids in each enzyme.

substrates, and since the catalytic function of the enzyme depends highly upon the precise arrangement of the catalytic region, it makes sense 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 [21].

TMA and VIR, in comparison to PHO, exhibit diverse features in their RMSFs 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 and subsequent destabilization of an unstructured loop near the entrance to the enzyme's binding site, both of which are connected linearly in primary sequence from the N-terminus. Similar behavior is also observed in the 50% IL simulations.

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 while 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.

While RMSD and RMSF are useful gauges of the global average behavior of the three enzymes, 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.

3.3.2 Enzyme Specific IL Induced Changes

VIR was deactivated significantly by even the lowest experimentally tested concentration of IL [21]. 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 (while 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, ARG108 and ASP151 establish a salt bridge. ASP19 is nearby this salt bridge, but it does not participate. Within the first 200 ns of each 15% IL trajectory, ASP19 along with the nearby charged residue LYS26 move farther away from the bulk of the protein into more solvated states. This helps expose the loop that contains ASP19. Interestingly, the hydrophobic VAL23 follows its two neighbors and becomes exposed to the surrounding solvent. Late in the trajectory, ARG108 breaks contact with ASP151 and forms a new salt bridge with ASP19. 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. While some salt bridges break and reform in the water solvated trajectories, such dramatic and predictable rearrangement of contacts is not observed.

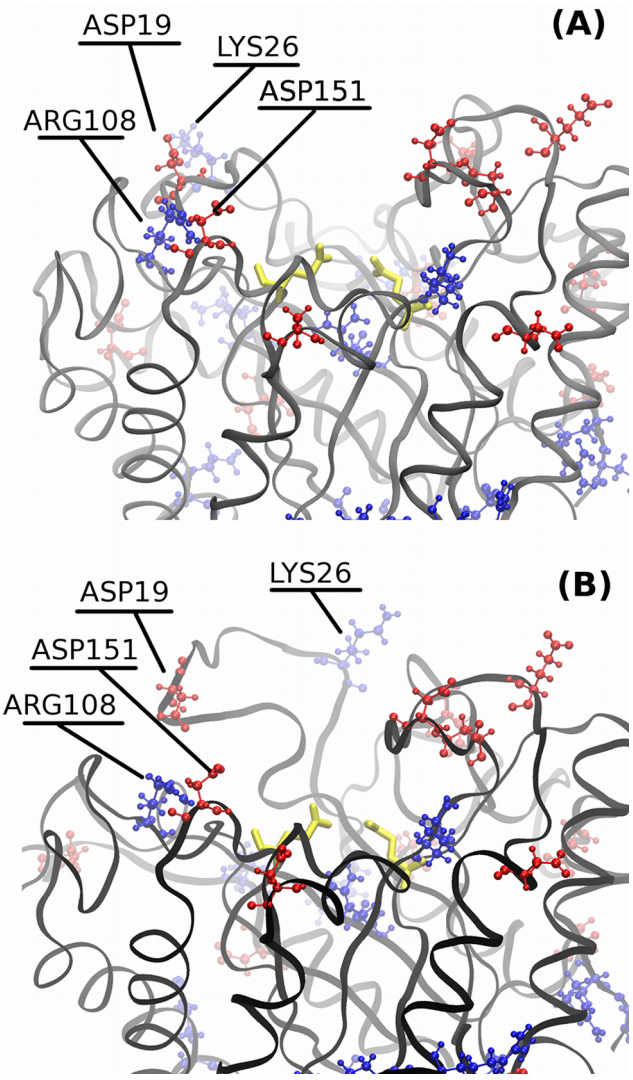


Figure 3.4: A view of VIR's binding cleft in 15% IL. (A) Structure after minimization. (B) Structure near the midpoint of the trajectory. Positive residues in blue. Negative residues in red. Catalytic residue heavy atoms in yellow.

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. Assuming that this rate has been optimized by natural selection, by changing the configuration of the binding pocket, the rate of binding would be decreased relative to the rate of unbinding. A more occluded binding pocket is less accessible to proper binding of the substrate. This would manifest itself in a higher Michaelis-Menten constant for the enzyme in IL, which we predict for this system.

Compared to VIR, TMA was deactivated less severely at low concentrations of IL [21]. 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 bridges and hydrogen bonding [20,31]. 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, while 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 [31].

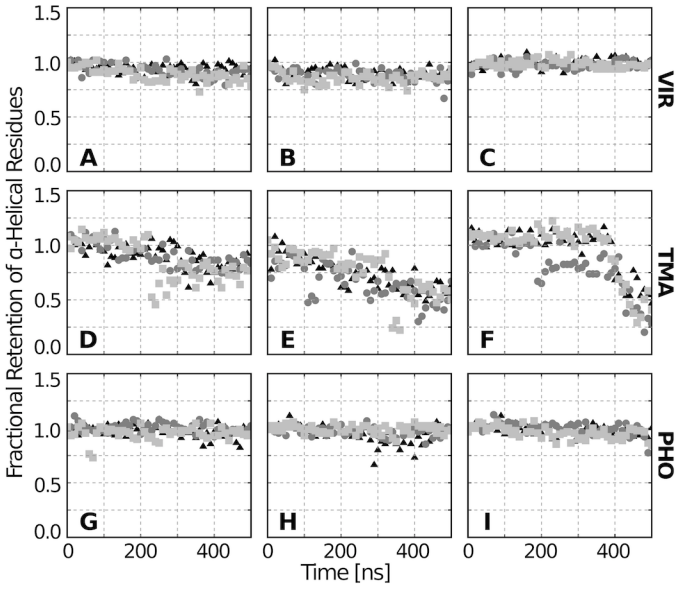


Figure 3.5: Retention of alpha helical structure of each enzyme over the full trajectories. 0% IL – A, D, G. 15% IL – B E, H. 50% IL – C, F, I. Three different replicate simulations are shown in each subfigure. Secondary structure contents were calculated using DSSP and normalized by the secondary structure calculated from the crystal structures.

Compared to the other enzymes, TMA experiences much higher loss of secondary structure. Figure 3.5 compares the alpha helical content, as determined by DSSP [117], 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. Similar calculations were also completed for the enzymes' beta sheets and showed that this loss of secondary structure is specific to helices. 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". While 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.

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. However, 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. The IL might act as a competitive inhibitor. The more viscous medium might lead to lower overturn rates. Yet, each of the aforementioned reasons would be expected to manifest itself with a gradual decrease in activity with increasing IL concentration. This is not observed experimentally. Instead 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. Based on our previous computational and experimental characterizations of GH11, we hypothesize that the deactivation of PHO's in high concentrations of IL influenced by enzyme aggregation. To be more specific, at some critical concentration of IL, the protein salts out of solution, and that concentration in

this case is somewhere between 20% and 50% v/v IL. This could be tested by simple dynamic light scattering measurements.

3.4 Conclusions

3.4.1 Unfolding vs Aggregation

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 [21]. Assuming that unfolding is dominated by the exposure of each protein's hydrophobic core, a hydrophobic cosolvent would be expected to stabilize the unfolded states and partially unfolded intermediates, thus reducing the unfolding temperature.

Datta and coworkers did not discuss aggregation in their treatment of the modes of deactivation, but we believe that it is an important consideration, especially for concentrations of IL that are similar to concentrations of inorganic salts that cause the protein to salt out. The experimentalists observed a large difference in the unfolding temperature when increasing the IL concentration from 20% to 50% v/v for TMA and PHO. TMA and PHO's unfolding temperatures decrease from 87.2 °C to 67.0 °C and from 91.8 °C to 66.1 °C across the aforementioned range, respectively. This indicates that there is a substantial difference between these two states. For the purposes of Datta's work, this change in unfolding temperatures is important, as it is enough to dismiss the enzyme as a candidate for use in industrial processes. What is not considered is whether the protein has aggregated before it is heated in the experiment. If the protein exists as a multimer, the local environment of each monomer is substantially changed. We believe that this is an important distinction, and we are curious to know if these enzymes aggregate in binary mixtures of IL and water.

If our aggregation hypothesis is true, our simulations of the protein in 50% w/w IL at 353 K do not reflect the experimental environment but an idealized environment instead. 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. For example, we observe no unfolding in the simulations of PHO at 50% w/w IL and 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.

3.4.2 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 hypotheses explaining the loss of activity of the enzymes in IL containing solvents, and we made predictions on how our hypotheses could be experimentally confirmed.

We hypothesize the following mechanisms for deactivation for the three enzymes and ways that they might be tested experimentally to confirm or dismiss them:

1. VIR is deactivated by deformation in the binding pocket that lead to the substrate to be unable to bind efficiently. This could be demonstrated by measurement of the Michaelis-Menten constant for this reaction at each concentration of IL [94].
2. TMA is deactivated by extensive loss of secondary structure in IL. X-ray scattering or small angle neutron scattering would be able to provide high level

detail about the changes in structure without being hindered by the unorthodox solvent.

3. PHO shows no signs of deactivation in IL. Aggregation is a likely cause of the deactivation observed at 50% IL, as we have observed experimentally for another glycoside hydrolase [30]. Dynamic light scattering would be able to detect large aggregates of this protein and determine the concentration at which it salts out.

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.

4 The Role of Enzyme Surface Charge on Mediating Ion-Specific Interactions in Ionic Liquids

4.1 Introduction

Interest in ionic liquids (ILs), organic salts that are liquid near room temperature, as solvents for biocatalysis has broadened in recent years. Their low vapor pressure, low flammability, high recoverability, and the library of potential cation and anion choices have presented considerable opportunity to improve enzyme stability and enantioselectivity compared to traditional organic solvents [118]. However, a major obstacle to widespread application of ILs toward enzymatic catalysis lies in the inability to predict the tolerance of a chosen enzyme to any particular ionic liquid [119]. This inability stems from a poor understanding of the atomic-scale interactions between the ions and enzyme that are likely responsible for the destabilization or inactivation of enzymes observed in many experiments [24]. Rectifying this issue will directly lead to an overall improvement in the selection and/or modification of enzymes as well as the rational design of ILs for IL-based biocatalysis.

Experimental studies into the loss of activity amongst many enzymes in ILs have indicated that detrimental interactions are strongly anion-dependent [26,34]. Several researchers have proposed Hofmeister series to rank-order the likelihood of anions to disrupt enzyme activity [22-24,120]. It has been hypothesized in other studies that chaotropic anions cause proteins to at least partially denature by disrupting crucial salt bridging and other tertiary structure elements and decreasing the entropic penalty to exposure of hydrophobic residues, either through interaction of partially folded intermediates or direct interaction with the folded protein [26,121]. Enzyme deactivation through induced aggregation or competitive inhibition has also been indicated as possible mechanisms [30,48,94].

Furthermore, the composition of the enzyme surface is thought to influence enzyme tolerance to increasing concentrations of aqueous IL mixtures. This is indicated through the observation that structurally similar family 5 cellulases from multiple organisms show disparate tolerance toward the ionic liquid 1-ethyl-3-methylimidazolium acetate. Cellulase from the halophilic organism *Halorhabdus utahensis* showed remarkable activity in IL, which may be attributable to its high proportion of negatively charged surface residues [20]. Comparison between a group of five cellulases also showed the highest retained activity from the cellulase with the most negatively charged surface [21]. Although currently not supported by any direct experimental evidence, this aspect of enzyme surface composition may be directly related to the pH induced resilience toward non-IL aqueous salt mixtures observed by investigations into the specific-ion effect in the past [122].

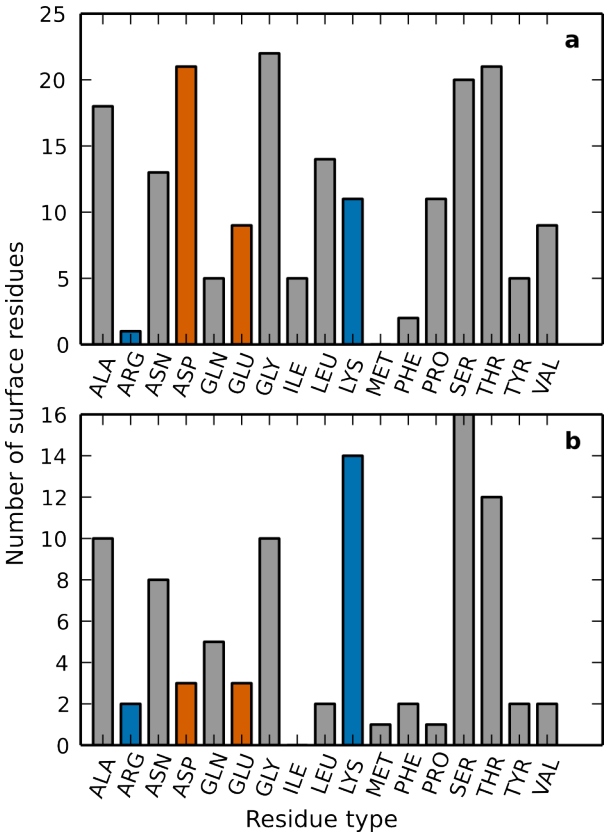


Figure 4.1: Surface composition of the enzymes CRL (a) and CT (b), with positively charged amino acids (blue) and negatively charged amino acids (orange) highlighted. Determination of which residues are on the surface is described in the methods section.

While investigating the role of enzyme surface charge on IL tolerance, recent studies have indicated that chemical modification of the enzyme toward decreasing the ratio of positive to negatively charged amino acids improves several enzymes' activities in increasing concentrations of the ILs 1-butyl-3-methylimidazolium chloride ([BMIM][Cl]) or 1-ethyl-3-methylimidazolium ethyl sulfate ([EMIM][$EtSO_4$]) in water [31]. Extending this work with the inclusion of fluorescence assays has demonstrated that decreasing

this charge ratio, for the enzymes *Candida rugosa* lipase (CRL) and *Bos taurus* α-chymotrypsin (CT), will increase enzyme performance and stability in ILs and increase the binding of [BMIM]+ cations to the enzyme surface [32]. These observations have led to a methodology for mediating IL interference of enzyme activity—which was confirmed by the application of surface charged modification to improve the activity of *Trichoderma reesei* cellulase in [BMIM][Cl] [123]. The advent of this new enzyme preparation to increase IL tolerance while quite powerful, naturally suggests that detailed understanding of the molecular scale solvent/enzyme interactions will provide new insights as to how new enzyme design rules such as site specific mutation could be used to even further bolster the effects of surface charge modification.

In support of these recent studies, we have completed multiple, all-atom molecular dynamics (MD) simulations to investigate with atomistic detail the influence of charged surface residues in mediating the contact of IL substituents with the protein surface and how these interactions change after sweeping modifications of the enzyme surface charge. Specifically, this study includes simulations of both CRL and CT in an aqueous mixture of either 20 wt.% [BMIM][Cl] or 20 wt.% [EMIM][$EtSO_4$] with random replacement of surface lysine residues with glutamate. This replacement strategy is designed to mimic to the effect of succinylation used in previous experiments– replacing amide groups with carboxylate groups and decreasing the ratio of positive-to-negatively charged surface residues [31]. CRL and CT have a few key differences that that add an extra level of complexity to this study. Not only do the two enzymes differ in size, with CRL and CT possessing a total of 534 and 240 residues, respectively, but the composition of their surfaces are drastically different (Figure 4.1). In their native configurations, CRL has a relatively few positively charged amino acids on its surface while CT has not only more positively charged residues but also relatively few negatively charged residues. This diversity

in surface characteristics allows us to explore a larger variety of positive-to-negative charge ratios when studying both enzymes in tandem.

By modeling our computational methodology after the experiments described in previous publications [31,32], the hypotheses and experimental observations can be matched to atomistic descriptions. Specific hypotheses of interest include that decreasing positive-to-negative charge ratio: increases the structural stability of the enzyme in IL; increases the number of bound cations to the enzymes surface, as indicated by experimentally measured Stern-Volmer quenching constants; and that anions are preferentially excluded from the enzyme surface with increased number of negatively charged surface residues. To our knowledge, the role of enzyme surface charge in changing biomolecule/solvent interactions has not yet been investigated with MD simulations. Herein we focus our attention on understanding whether changes in the interface between the enzyme and solvent lead to significant structural or dynamic changes in the enzyme or only lead to changes in the solvent environment around the enzyme. By comparing differently mutated enzymes with identical levels of charge modification, an additional goal is to identify specific features of the modified enzyme surface that lead to beneficial solvent reorganization, potentially suggesting new biomolecular design rules to be tested in future studies.

4.2 Methods

4.2.1 System preparation

X-ray crystal structures, obtained from the protein databank [106], for CRL (1CRL) [51] and CT (1YPH), serve as the initial structures before any subsequent modification. The coordinates for all water molecules from the experimental structures were left intact for all systems. Surface glycosylation sites from the CRL structure were removed due to the uncertainty in the length of the attached sugar chains and the types of sugars involved.

This study used replacement of surface lysine residues by glutamate residues to alter the ratio of positive to negatively charged amino acids. Only surface residues were chosen to avoid disrupting the protein stability, which would complicate analysis of the solvent structure about the enzymes. Additionally, the experimental conditions we are attempting to mimic use chemical modifications of the protein side chains that presumably affect only solvent exposed residues. Residues with solvent-accessible surface area (SASA) greater than at least 30% of the maximum possible SASA for that residue type were counted as being part of the protein surface [113,114]. The maximum SASA is a reference values from a tri-peptide where the residue in question is between two glycine residues. From the subset of the amino acids identified as part of the enzyme surface, randomly selected lysine residues were changed to glutamate by deleting the side-chain amine group from the structure file and using Ambertools 11 [56] leap environment to guess coordinates for the carboxylate heavy atoms.

Two sets of 3 modified systems were prepared in this way – three with approximately 50% of the surface lysines altered and three with approximately 75% of the lysines altered. Repeating the randomized mutation process in triplicate is used to ensure no single residue is responsible for any general observations about changes in solvent interaction. It should be noted that the proteins with 75% lysine modification were built by randomly modifying additional lysine residues from the 50% modified proteins. This was an intentional aspect of our study to allow for tracking of beneficial modifications by comparing amongst the modified systems. The 50% and 75% total modification portions were chosen because they fall within the range of primary amine to acid ratios observed by experiment [32]. Specifically, for CRL and CT: 50% of lysines modified corresponds to primary amine / acids ratios of 0.15 and 0.91, respectively, and 75% modification corresponds to ratios of 0.08 and 0.27, respectively. The 75% modified systems are similar to the succinylation chemical modification conditions.

For each CRL and CT, the 6 modified proteins in addition to the unmodified protein were each solvated in water, 20 wt.% [BMIM][Cl], or 20 wt.% [EMIM][$EtSO_4$]; 14 proteins in 3 solvents for each CRL and CT resulted in a total of 42 simulations. Packmol was used to solvate each protein inside a cubic box with a side length chosen to ensure a 1.25 nm minimum distance between any protein atom and the box boundary [101]. Ambertool's leap environment was used to randomly replace water molecules with counter-ions until each system had no net charge. For all systems except unmodified CT, the neutralizing ion was sodium; the unmodified CT system included chlorine counter-ions. The total number of counter-ions added is different for each system, but significantly less than the number of IL ions included in the solvent. IL ions were not used as counter-ions to avoid an uneven number of cations and anions complicating our analysis.

4.2.2 Molecular dynamics

GROMACS 4.5 was used as the MD engine for all simulations [124]. The Amber99SB-ildn force field [125] was used for all protein atoms. IL molecules were parameterized using the general Amber force field [58] with RESP partial charges [59] fit to an electrostatic potential field calculated from a Gaussian 09 [60] single-point energy calculation at the HF/6-31g* level of theory. The RESP partial charges were scaled by a constant value of 0.8 for all atoms in order to improve the model's approximation of dynamic properties for the IL molecules [100]. Water molecules were approximated by the TIP3P model [96]. Short range electrostatic and Van der Waals interactions were cut off at 1.1 nm with a switching function starting at 1.0 nm and long-range electrostatics were reintroduced using PME. All MD simulations attained 2 fs time steps by constraining all bonds involving hydrogen atoms using LINCS [102]. Heating and NPT simulations used a stochastic velocity-rescaling algorithm to maintain target average temperatures [126].

Each system was relaxed in the solvent using 10,000 steps of steepest-descent minimization. The minimized coordinates were subsequently heated from 0 to 300 K over the course of 20 ps using the simulated annealing protocol in Gromacs 4.5. The heated systems were used as the initial coordinates for 50 ns NPT simulations using the Berendsen barostat [104]. All coordinates for this final simulation were saved every 500 ps and the protein coordinates were saved every 250 ps. Analysis of trajectories accumulated during the final NPT simulation is completed by a combination of Gromacs 4.5 tools in addition to VMD [68] and custom python scripts, which made extensive use of the MDAnalysis package [69].

4.3 Results & Discussion

4.3.1 Enzyme stability

Cα root-mean-square deviation (RMSD) of the proteins compared to the corresponding experimental structures was used to track structural stability over the course of the production simulations. Ensuring that no large-scale loss of protein structure or significant domain movement has occurred simplifies the comparison of solvent structuring between each modification and solvent. CRL possesses a highly dynamic "lid" domain that has been noted in a previous simulation study to exhibit significant motion during MD simulations [29]. The dynamics of the lid are intrinsically related to the mechanism of the enzyme and aside from any secondary structure loss, which is not observed upon visualizing the trajectories, does not reflect structural instability for this enzyme. Therefore, when measuring the RMSD for the CRL systems the lid residues (66-92) were excluded from the alignment and calculation. Losses in secondary structure, tertiary structure, or significant domain movement generally results in RMSD values greater than 0.4 nm for proteins the size of CRL or CT.

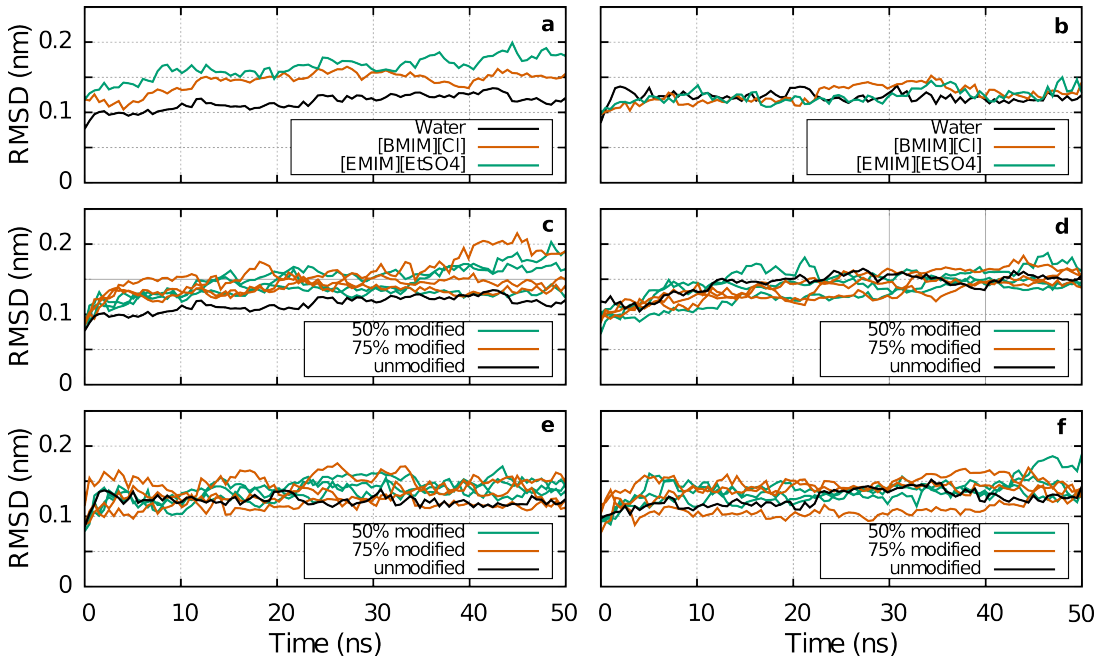


Figure 4.2: Bezier curves fit to the Cα RMSD after alignment to and calculated from the experimental structure for CRL (a,c,e) and CT (b,d,f). Curve fitting is used to ease comparison and visualization. The unmodified enzymes in the various solvents (a,b) can be used as a baseline for stability. The behavior of the modified and unmodified enzymes in water (c,d) and the 20% [BMIM][Cl] (e,f) are also shown.

Figure 4.2a,b reveals the RMSD of the unmodified enzymes in all three solvents. Both CT and CRL have remarkable stability in both ILs during the simulations with CT showing virtually no difference in deviation from the reference structure when comparing between the three solvents. Visualizing the trajectories to confirm the RMSD results shows that indeed the unmodified enzymes are relaxing to the solvent without any observed changes in secondary structure or domain movement aside from the mobile lid of CRL. A previous simulation study using similar time scales and concentrations of aqueous IL also showed similar behavior of a xylanase enzyme when comparing between increasing aqueous IL concentrations and pure water [30]. We note that the time scale of our simulation (50 ns each) is expected to be sufficient to allow for solvent equilibration but far shorter than that required for lead to large scale unfolding or denaturation of the enzyme.

Comparing the behavior of each modification in water to the unmodified enzyme in water details the influence of the modifications themselves on the overall stability of the enzymes. Modifying lysine

to glutamate not only creates a new negatively charged group but also eliminates the positively charged amide. This swing in local electrostatic interactions has the potential to not only interrupt salt bridges but to also alter the local solvation shell, possibly leading to changes in the enzyme stability. However, Figure 4.2c,d shows that each modification reaches the same final value of RMSD as that of the unmodified enzyme within the observed fluctuations. On the simulated time scale the presented data does not reflect any major instabilities in the enzyme created by the modification of surface lysine residues for either enzyme in water. Furthermore, a similar analysis of the modifications in IL shows that the root mean squared fluctuations (RMSF) of the enzyme in 20% [BMIM][Cl] (Figure 4.2e,f) and 20% [EMIM][$EtSO_4$] (not shown) are identical within the observed fluctuations to their behavior in water. In summary, for both proteins and all modifications and all solvents, the final RMSD values were less than 2.0 nm. The RMSD data demonstrates that all modifications of both CRL and CT are stable on the 50 ns time scale and that the ILs have similar and minimal influence on the stability of any of the tested modifications.

One hypothesis found in literature about the role of IL in the disruption of enzyme activity is that the ILs interfere with salt bridges, potentially disrupting important enthalpic contributions to tertiary structure. Furthermore, if local salt bridges are disrupted by the surface modifications there could be a synergistic effect created by the previous salt-bridged residues now being more open to solvent interaction. In other words, a single mutation may effectively increasing the possible number of strong electrostatic interactions with the solvent instead of simply changing which ionic species is more preferentially bound by swapping a single charge. Using the salt bridge plugin in VMD, we have enumerated the number of salt bridges observed in each simulation and tracked the distance between the donor/acceptor nitrogen and oxygen pairs. Using a switching function with a smooth cutoff at 0.4 nm to identify “active” salt bridges [127], the disruption of salt bridges can be monitored throughout each

trajectory. The average number of active salt bridges for unmodified CRL in water is 16±2. The number of salt bridges identified on unmodified CT was 2±1; remarkably few salt bridges are observed within the CT structure during the simulations. Interestingly, the CT experimental structure contains no active salt bridges compared to CRL's 18 salt bridges according to the plugin's search initial search criteria of 0.32 nm N/O interatomic distance.

Comparing solely amongst solvent type for each degree of modification, there is no statistically significant difference in the number of active salt bridges in ILs on the time scales we have observed. However, when comparing between the different random modifications of CRL, we noted that three of the modified systems show an overall decrease in active salt bridges to 13±2. These three systems are the only modifications to show a significant difference compared to the unmodified enzyme, and all three share four lysine-to-glutamate modifications in common, including a K363E modification that is

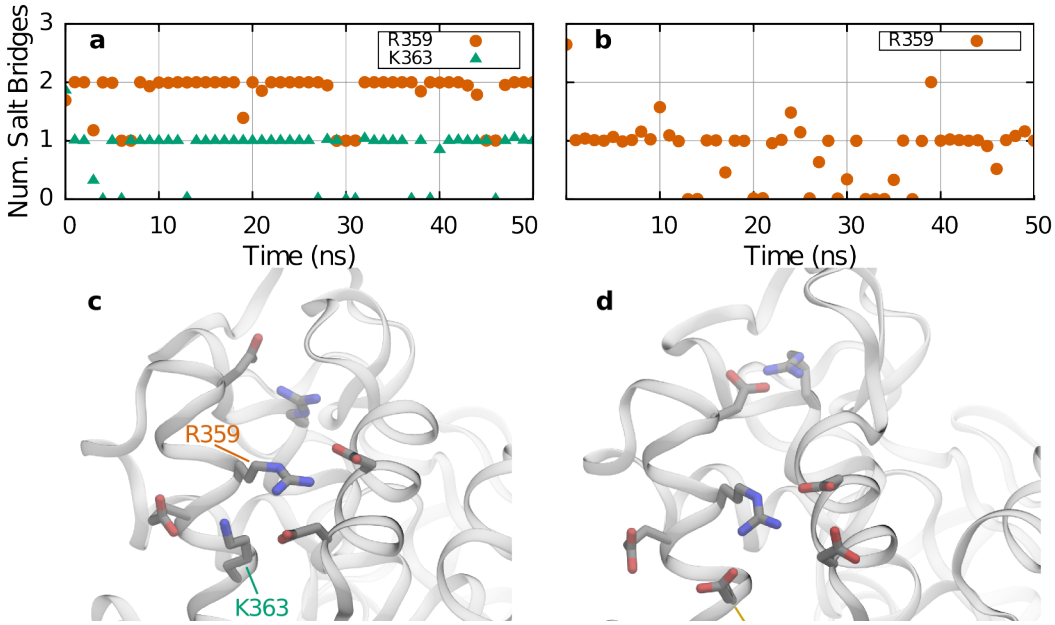


Figure 4.3: The total number of active salt bridges involving R359 and K363 near the unmodified (a) and modified (b) simulations of 1CRL in water. Since the only observed salt bridge for K363E in the modified simulation is with R363, panel b only contains a single data series to avoid redundancy. Snapshots representative of the organization of charged residue interactions near the unmodified (c) and modified (d) K363 residue are shown.

unique to these three. K363 resides within a patch of charged residues that form salt bridges across two short helical segments (Figure 4.3c). K363 and the adjacent R359 together establish a network of salt bridges with three nearby negatively charged residues throughout trajectories of the unmodified system (Figure 4.3a). The three modified systems we noted above show that the K363E residue causes R359 to reorient into the space formerly occupied by lysine as it forms interactions with the new glutamate (Figure 4.3d). This altered orientation increases the number of residues R359 interacts with during the simulation but decreases the lifetime and simultaneous number of salt bridges it forms with the nearby negatively charged residues (Figure 4.3b). The negatively charged residues themselves sporadically participate in interacting with R359 as they continuously fluctuate between bonding with R359 and stretching out into the solvent. In summary, the K363E modification leads to R359 increasing the variety of salt bridges formed but decreasing their longevity and average number. The negatively charged residues formerly participating in the strong salt bridges with K363 and R359, as observed for the unmodified system, spend significant periods of the simulation with no intramolecular interaction. When not interacting with a positively charged residue the negatively charged side chains extend into the solvent and interact with solvent molecules, demonstrating how positive-to-negative charge mutations can have secondary effects leading to even more potential solvent interactions.

4.3.2 Solvent organization

Our previous study involving simulation of CRL in near-pure ILs [BMIM][NO_3] and [BMIM][PF_6] revealed that IL anions can outcompete all other solvent molecules for contact with the enzyme surface except for negatively charged residues [29]. We attributed this to the relatively large size of the cations and the lack of flexibility in the imadazolium ring. We also hypothesized that the IL would organize itself in concentric spherical layers that alternate charge due to the anion-dominated surface layer. To test whether this hypothesis was true for surface-charge modified enzymes, the net

charge per unit volume (charge density) as a function of radial distance from the protein was found to be a useful metric. Moreover, it is one that can be calculated from simulation trajectories in a facile way.

We calculated the charge density using only solvent atoms (protein atoms were excluded from the charge summation) in concentric spherical shells about the enzyme center of mass. The time-averaged values using frames from the last 25 ns of the trajectories and distance bin sizes of 0.3 nm are displayed in Figure 4.4. Similar results are also observed when calculating this metric using the first 25ns of the simulation or when averaged over the entire simulation. The calculated bounding sphere radius of all backbone atoms of CRL and CT are 3.63 and 2.48 nm, respectively, using the experimental structures. The distance between the protein center of mass and the closest observed IL molecule is used as the inner shell radius of the first bin. Given that the first bin for each system is less than the radius of the enzyme-bounding sphere, the first bin is representative of solvent either within the enzyme binding site or other depressions in the protein surface. The

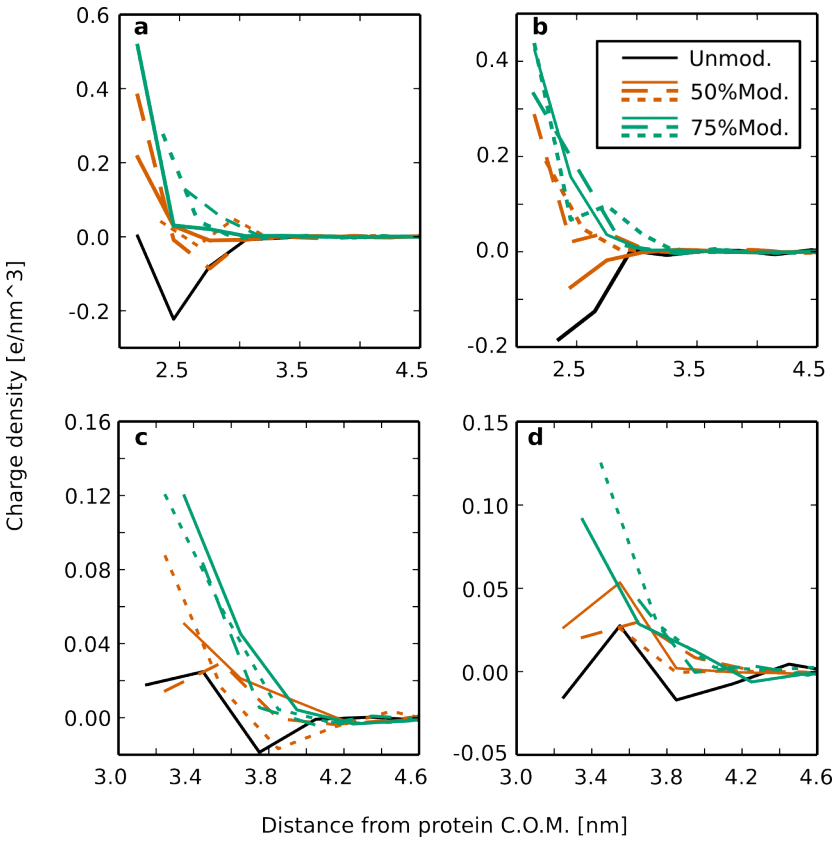


Figure 4.4: Charge density of only solvent molecules (water, ILs, and counter ions) in spherical shells centered on the protein center of mass. The charge density for all simulations, including both CT (a,b) and CRL (c,d), in both 20 wt.% [BMIM][Cl] (a,c) and 20 wt.% [EMIM][$EtSO_4$] (b,d), are shown. The spherical shells used to calculate charge density have a radius of 0.3 nm.

calculated asphericity of CRL and CT are approximately 0.07 and 0.03, respectively [128]; both enzymes are highly spherical in shape.

The time-averaged charge density data shows that the unmodified enzymes have a net negative or neutral close-range solvation shell. The modified systems have net positive solvation shells that increase in strength with the increasing number of modifications. Interestingly, for both enzymes in either ionic liquid the initial solvation shells show that the solvent charge density universally increases with increasing number of modifications (decreasing positive-to-negative charge ratio). Considering the change from negative net charge to positive, the data indicates that increasing the number of negatively charged residues shifts the population of IL substituents from anion to cation-dominant. When taking into account the bounding sphere radii of CRL and CT (3.63 and 2.48 nm, respectively), the charge density plots show that the solvent achieves an equal proportion of cations and anions within the first 1.0 nm from the protein surface. This suggests that long-range organization of the IL on this length scale is not occurring. However, several of the systems, especially with the smaller chlorine cation of [BMIM][Cl], there is a noticeable decrease in the charge density, sometimes dipping into net negative values, before the solvent becomes net-neutral. This may indicate a second solvation shell about the enzyme with some net-organization of anions.

The solvation shell around CRL is comparatively more neutral than that observed for CT. Since the metric we are employing does not depend on the volume of the spherical shell (i.e., charge/volume), the relative enzyme sizes should not influence comparing the values between the two systems. Therefore, the most likely difference between the two enzymes leading to this difference in charge density values may be created by the compositions of the surface (c.f., Figure 4.1a,b). CRL possesses a much larger portion of negatively charged residues and the relative abundance of negative charges on the surface only increases with the number of modifications. CRL also possesses a total of 74 hydrophobic surface

residues compared to CT's 31. As a percentage of unmodified CRL's total number of surface residues: 40, 6, and 16 % of surface residues side chains can be characterized as hydrophobic, positively charged, and negatively charges respectively. In comparison, these values for CT are 33, 17, and 6 %, respectively. Considering that the lowest positive-to-negative surface charge ratio of these systems is that of CRL with 75% lysine modification, we expect that 75% modified CRL systems have the largest solvent charge density at the surface. Furthermore, the higher proportion of exposed hydrophobic residues on CRL would hypothetically lead to an increase in bound cations, further increasing the charge density. This hypothesis is based on the fact that BMIM and EMIM cations are the most hydrophobic residues present in the system, and the large number of Van der Waals interactions provided by their long alkyl chains would increase their probability to strongly interact with hydrophobic residues.

To gain a more concrete picture of how the proportions of IL substituents near the enzyme surface are changing with enzyme modification, we have also calculated a radial distribution function (RDF) about the protein for each type of ion. RDFs are a standard metric employed in simulation studies to characterize the organization of one species with respect to another (often inter-atomic organization). The values of the RDFs presented here are ratios of the ion density compared to the bulk ion density within increasing radii of spherical shells about the protein. In contrast to the charge density metric described above, the RDF was calculated for distances away from the nearest protein atom instead of the distance from the protein center of mass. This allowed us to employ distance matrices to significantly speed up this computationally expensive metric. After calculation of the distance matrix between all ion atoms and all protein atoms, the minimum distance between each ion from the protein using the minimum image convention was recorded. Calculation of the volume element for each shell required assuming that the minimum distance from the protein is equal to the minimum distance from the

bounding sphere of all protein backbone atoms. Figure 4.5 shows the RDF profiles for each ion type for both enzymes.

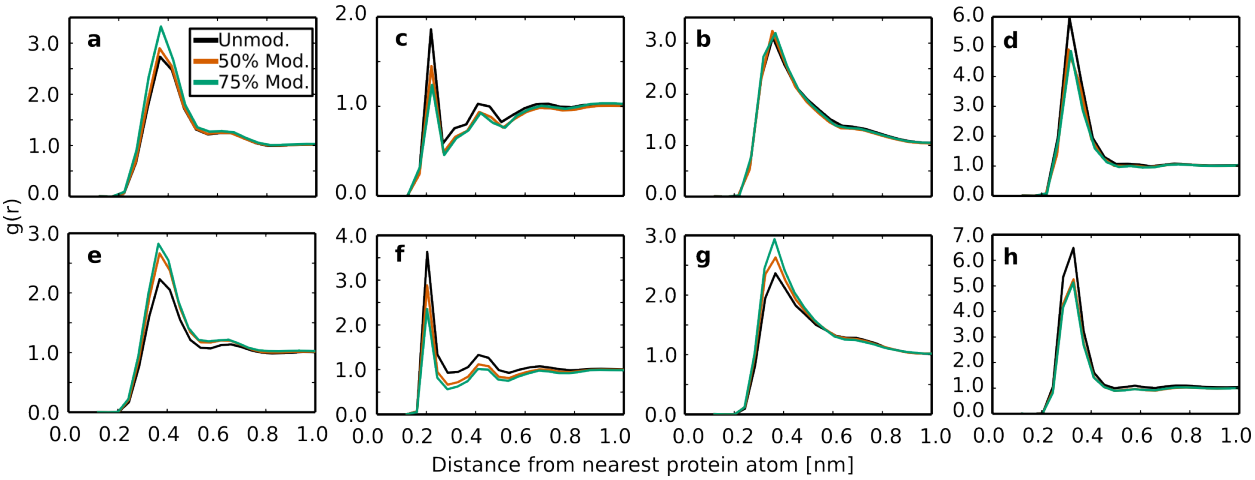


Figure 4.5: The radial distribution of IL constituents about the enzymes vs the minimum distance between each solvent molecule and any protein atom for CRL (a-d) and CT (e-h). From left to right, each panel describes the distribution of [BMIM]+, [Cl]-, [EMIM]+, [$EtSO_4$]- about the protein. Note the difference in y scale for each panel. Data for the modified systems are an average of three simulations. Averaging of the replicate simulations was used to increase clarity.

Overall, the cation RDF profiles show a broad first peak slightly under 0.4 nm from the protein. The BMIM cation also shows a slightly increased density near 0.6 nm from the protein. On the other hand, the two anions, whose size and structure differ considerably, reveal dissimilar density profiles. The chlorine anion RDF reveals that chlorine's small size and concentrated charge allows it to penetrate closely to the protein surface. The first chlorine peak of 0.2 nm less than the first BMIM peak, and the approximately 0.2 nm difference between their first peaks may indicate a charge stratification effect near the enzyme surface. Additionally, chlorine RDFs also reveal a second peak slightly above 0.4nm. This secondary peak may be caused by interaction of chlorine atoms with the BMIM cations within the first BMIM solvation shell. Between the two peaks of chlorine, regions of anion density that dips below the chlorine bulk density are apparent. This may be attributable to repulsive electrostatic interactions between chlorine atoms within the high-density shells. In contrast, the larger and more charge-

distributing $EtSO_4$ anion shows similar behavior to that of the cations, with a primary peak slightly less than 0.4 nm from the protein surface with no apparent secondary peak. Regarding the literature discussion of the role of anions in protein deactivation, a remarkable hypothesis is that chaotropic anions, which typically have less distributed charges and fewer number of atoms, can penetrate more closely to the protein surface or even penetrate within the protein surface leading to an increased ability for these anions to disrupt protein structure [26,42]. The RDF profiles presented here for the chlorine atom do confirm that they can indeed penetrate closer to the protein surface compared to the more kosmotropic $EtSO_4$. Finally, it may also be significant that the density of all ion types for both enzymes approach the bulk value after 0.8 nm distance from the protein surface, further confirming that long range organization is not apparent for 20 wt.% concentrations of these ILs.

Another interesting feature of the RDF profiles is the separation of the primary peaks for each system according to the degree of enzyme modification. For the BMIM cation near CRL and CT and the EMIM cation near CT, the density profiles describe an increase in cations relative to the bulk with decreasing positive-to-negative surface charge ratio. It is not apparent why the EMIM cation near CRL does not show a similar trend. However, considering the decrease in $EtSO_4$ density calculated from the same simulations, we can predict that there must be an increase in the number of waters near the protein surface as a result of the displaced anions. For both anions and both enzymes, the RDF profiles indicate a decrease in anion density near the surface upon protein modification, confirming the hypothesis presented in previous experiments that decreasing the positive-to-negative charge ratio helps to exclude anions from interacting with the protein [32].

4.4 Conclusions

Molecular simulations were employed to probe the effect of decreasing the positive-to-negative surface charge ratio of the enzymes CRL and CT in 20 wt.% [BMIM][Cl] and [EMIM][$EtSO_4$]. The analyses of these simulations included RMSD and counting of salt bridges to indicate the structural stability of the modified enzymes and a combination of charge density and radial distribution functions to gain insight into the structuring and organization of IL substituents. Our results indicate that on the 50 ns time scale that there is no evidence of structural destabilization or denaturing for either the modified or unmodified enzymes. It should however be noted that the time scale of these simulations are likely too short to capture the large-scale movements of denaturation.

The design of this computational study was inspired by experiments using chemical means to modify the surface of CRL and CT. Results from these experiments were hypothesized to indicate: an increase of stability for these enzymes, an increase in interaction of cations with the protein surface, and increased ability for the enzyme to exclude anion binding with decreasing positive-to-negative charge ratio. Although our results do not indicate either increased or decreased stability of the modified enzymes, we have demonstrated with two independent metrics that changing the surface characteristics of the protein can alter the close-range ion solvation shells. The changes are characterized by an increase in local concentration of anions and increase in bound cations or water molecules due to the gain of negatively charged residues. We believe that the solvent reorganization, which occurs on a very short timescale, is a critical component in protecting the enzyme from the smaller anions, which can penetrate the enzyme surface and disrupt structure in a more facile way.

We have also demonstrated that there is a possibility for synergistic effects created by modifying residues that participate in hydrogen bonding or salt bridge interactions. Finally, comparing the metrics we have employed to probe solvent structure, there is a strong indication that at the 20 wt.%

concentrations there is no observable long range structuring of the IL molecules about the enzyme. Within the first 0.8-1.0 nm from the protein surface the charge density approaches 0 and the density of all ion types approaches that of the bulk solvent.

Both of the solvent structure analyses we employed required treating the protein surface as spherical to calculate volume elements. Although supported by the calculated asphericity of the proteins, the nonspherical features of the protein such as the binding site or surface depressions may affect our ability to interpret the physical meaning of the metrics calculated using the spherical assumption. However, visual inspection of a surface representation of the protein does not suggest that any such feature sizes are significant enough to change the conclusions drawn from the RDF and charge density analyses. As we have shown previously, the protein structures are stable and similar to the experimental configurations on these times scales. The large-scale domain motions of denaturing are not indicated, allowing us to assume that any affects caused by the asphericity of the protein will be consistent across simulations of the same enzyme.

5 Structural Effects of Methionine Oxidation on Isolated Subdomains of Human Fibrin D and αC regions

5.1 Introduction

During blood clot formation, the plasma protein fibrinogen is normally converted by thrombin to fibrin monomers that self-polymerize both linearly – forming long fibrils – and laterally, increasing fibrin fiber diameter and forming networks of fibers. These branching fibers make up the 3-D scaffold of blood clots and confer strength, flexibility and structure to clots so they can adhere to wounds and stop bleeding. Increasing fibrinogen concentration is typically associated with a condensed clot structure, thinner fibrin fibers, and increased clot stiffness [129]. This structure-concentration relationship may mediate the long recognized positive association between fibrinogen concentration and the risk of thrombotic cardiovascular disease, including acute myocardial infarction [130] and microvascular disease in diabetes mellitus [131].

There is also evidence that post-translational modification of fibrinogen may contribute to cardiovascular disease. Glycosylation of fibrinogen in diabetics has been hypothesized as a possible mechanism of decreased fibrin gel permeation despite normal fibrinogen concentration [132]. Fibrinogen is also among the most sensitive plasma proteins to post-translational modification by oxidative stress [133]. This susceptibility is important because there is a direct link between fibrinogen oxidation and/or nitration and fibrin polymerization [134,135]. Depending on the type and extent of modification, these changes can induce a prothrombotic or antithrombotic clot phenotype [136,137]. Fibrinogen oxidation has been associated with increased cardiovascular events in chronic kidney disease patients [138,139] and fibrinogen nitration has been associated with myocardial infarction, possibly due to its induction of a prothrombotic fibrin clot phenotype [140]. This evidence highlights the potential for oxidative modification of fibrinogen to contribute to the mechanism of thrombotic cardiovascular

diseases. However, little is known of how the molecular events associated with oxidation can alter fibrin polymerization.

Recent evidence uncovered by Weigandt et al. [141], shows that selective oxidation of methionine on fibrinogen – forming methionine sulfoxide – can inhibit lateral aggregation of fibrin protofibrils after activation by thrombin and produces an ultra-thin fibrin fiber network with weakened mechanical properties and increased resistance to fibrinolysis. They found that oxidation of methionine primarily occurs at two positions – residue 476 of the α chain (αMet476) and 367 of the β chain (βMet367). Understanding the mechanism of fibrin lateral aggregation is an area of active research and many studies have indicated these specific sites are important, with αMet476 belonging to the structured portion of the αC region [142] and βMet367 being part of hole 'b' of the fibrin D region [143-145]. Their finding of selective methionine oxidation in association with altered fibrin polymerization and altered gel structure offers an opportunity to examine more closely the effect of oxidation on fibrin polymerization via a novel pathway in addition to the basic mechanism of fibrin polymerization. An important first goal would be to discover the mechanism by which oxidation influences the behavior of fibrin at the molecular level.

Despite the wealth of mesoscale information about the influence of oxidation on fibrin gel structure, little experimental evidence is available at the molecular scale. With the current understanding of the multiple proposed mechanisms of fibrin lateral aggregation along with the crystal and NMR structures of fibrin, molecular simulation provides an avenue to further understanding of these mechanisms and clarifying which portion of fibrin is responsible for the observed gel characteristics. In this study we utilize classical molecular dynamics (MD) in addition to parallel tempering MD (PT) with the recently developed well-tempered ensemble (PT-WTE) to independently examine the effect of βMet367 and αMet476 oxidation on the human fibrin D region and the N-terminal subdomain of the

human αC-domain, respectively. Specifically, MD simulation of a fibrin monomer fragment is used to examine the effect oxidation has on the structure of hole ‘b’ and the stability of a bound knob ‘B’ surrogate in the D region. Enhanced sampling PT-WTE simulations are used for the first time without additional metadynamics bias on the protein degrees of freedom in a fully solvated representation of an αC-subdomain in order to explore the conformational ensemble of this protein fragment and characterize the differences created by methionine oxidation.

5.2 Methods

5.2.1 Molecular Structures

The D region structure from human fibrin was retrieved from the Protein Data Bank (PDB) entry 2Z4E [146]. The structure contains only a portion of the coiled-coil E region but contains the entirety of the D region (see Figure 5.1A), including both hole ‘a’ and hole ‘b’ with bound GHRP and GHRPY peptides, respectively. These peptides represent surrogates of knob ‘B’ of fibrin, which has been shown to associate with fibrin holes ‘a’ and ‘b’ [145]. The structure was chosen over the more complete

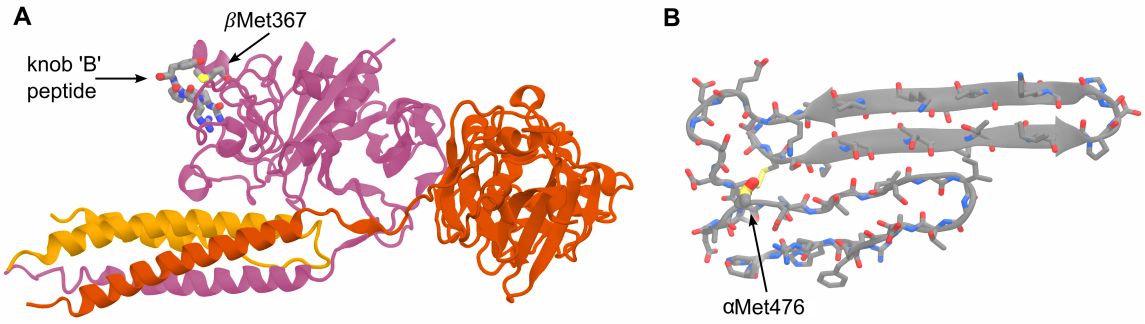


Figure 5.1: Molecular representations of the initial structures of (A) the D region corresponding to PDB 2Z4E and (B) the αC-subdomain homology model. The methionine of interest in each system is in a location critical to the hypothesized lateral aggregation mechanisms. The D region βMet367 is in direct contact with the bound knob ‘B’ peptide and is part of hole ‘b’. αMet476 is part of the “hairpin-linking region”. The highly helical (coiled-coil) segment of the D region system is part of the fibrin E region.

structure, found as PDB entry 3GHG [147], in order to limit system the size to a single D region without

having to arbitrarily choose a location to truncate the coiled-coil region. The Cα root mean square deviation (RMSD) between the matching residues of 2Z4E and 3GHG (662 residues) is 0.10 nm, showing that the protein backbone coordinates for these structures are consistent. To avoid confusion, we follow the recommended nomenclature for fibrin and fibrinogen in this paper [148].

In order to examine the effects of methionine oxidation on hole 'b' influenced lateral aggregation – that methionine sulfoxide either disrupts the native conformation or the stability of the bound knob 'B'– a total of four systems were built. These systems include all combinations of the fibrin fragment with or without the knob 'B' surrogates and with or without oxidation of βMet367. The first system built was that with methionine not oxidized and with bound knob 'B' surrogates. This initial system was then used to create the other three systems, either by addition of the methionine sulfoxide's oxygen, and/or deletion of the knob 'B' surrogates. The GHRPY fragment found in hole 'b' of the crystal structure was modified to GHRPL by replacing the tyrosine side chain with that of leucine in order to replicate the N-terminal sequence of the human fibrin β chain, knob 'B'. The protein structure for each system was solvated in Amber's [56] leap with an equilibrated TIP3P [96] water box to a maximum of 2.5 nm from any protein atoms. This large padding allows for some rotation of the protein before an undesirable interaction with its own periodic image. Using leap, chlorine ions were added by replacing water molecules to neutralize the system charge. The initial box dimensions for each system was approximately 8.8 x 8.8 x 15.0 nm with approximately 110,000 atoms. All of the D region systems retained the glycosylation and bound calcium ions present in the crystal structure.

The structured portion of the human αC region, termed the αC-domain, has yet to be resolved by X-ray crystallography or NMR experiments. To date, conjecture about the N-terminal subdomain of the αC-domain's lateral aggregation mechanism and estimates of its structure comes from homology models built from recombinant bovine αC-domain fragments, which has been resolved by solution NMR [149].

The homology model used for the initial protein coordinates in this study was obtained from the Medved group and is discussed in their recent publication [150]. The model consists of 55 amino acids (441-496α) and possesses two beta hairpins arranged into a beta-sheet like interaction, a disulfide bonded cysteine pair between one hairpin and the connecting segment, and a single methionine near the connecting segment on the larger hairpin (see Figure 5.1B). Additionally, two homology models were built using the primary sequence of the human α chain [151] using the homology model program Modeller [152]. Of the 10,000 homology models generated, the best models as determined by Rosetta [153] and DOPE score were tested alongside the "Medved model" for stability during preliminary 300, 350 and 400 K MD simulations in the NVT ensemble. Of the three models, the Medved model maintained the lowest Cα root mean squared deviation (RMSD) after 25 ns and was subsequently chosen for the simulations described herein.

Two PT-WTE simulations using the homology model of the N-terminal subdomain of the αC-domain were used in this study to examine the effects of oxidation. They correspond to the αC-subdomain with and without αMet476 oxidation. Similar to the D region simulations discussed above, the methionine sulfoxide residue was built by adding the methionine sulfoxide oxygen in a position corresponding to the geometry optimization of methionine sulfoxide in vacuum. Using leap, the initial coordinates of each model were solvated with TIP3P water molecules to a box dimension of 6.0x6.0x10.0 nm and solvent molecules were replaced with chlorine ions to neutralize the systems.

5.2.2 Molecular Dynamics Simulations

Simulations described in this study were completed in GROMACS 4.5 [124] with the proteins parameterized using the Amber 99SB-ildn force field [125]. The glycosylation and associated asparagine residues were parameterized with the GLYCAM06 [52] force field. Methionine sulfoxide residues, which represent the oxidized state of methionine, were parameterized in Amber's antechamber

using the GAFF force field in addition to RESP charges calculated from a Gaussian 09 structure optimization at the HF-6/31-g*(d) level of theory.

The systems were first subjected to 1000 steps of conjugate gradient minimization with all protein positions restrained followed by an additional 4000 steps with no restraints. The minimized system was then gradually heated from 0 to 310 K in increments of 10 K for 2000, 2 fs time steps of NVT simulation at each temperature. The αC-subdomain coordinates were further heated to 470 K, again by increments of 10 K, with the final coordinates at each 10 K increment saved, to generate the initial coordinates for each parallel tempering replica (more details below). For each system the long range Van der Waals forces were smoothly reduced to zero, between the range of 0.9 to 1.0 nm, by shifting the Lennard-Jones potential. The Coulombic interactions were cut off at 1.0 nm and long-range electrostatic interactions were reincorporated using PME. The neighbor list in all simulations were updated every 10 simulation steps with a neighbor list cutoff of 1.1 nm. All MD and PT-WTE simulations used a 2 fs time step and LINCS [102] to constrain bonds to their equilibrium values. Center of mass translation is removed every 1000 simulation time steps.

The final frames from the heating simulations were used as the initial coordinates for the production simulations for each system. The D region simulations were run in the NPT ensemble with the Berendsen [154] barostat and a global stochastic thermostat [126] keeping the average pressure and temperature at 1 bar and 310 K, respectively. Replicas in the PT-WTE simulations also used this global stochastic thermostat. NPT simulations of the four D region systems were run for a total of 350 ns each and an additional two replicate simulations were run for each system for 200 ns after repeating the initial heating simulations with randomized initial velocities. These repeat simulations were used to verify the observations made in the longer production simulations. The PT-WTE simulations (more details follow) of the αC-subdomain were run for a total of 300 ns per replica for both the oxidized and non-oxidized

systems. This resulted in aggregate simulation times of 8.4 and 3.0 μs for the αC-subdomain and D region, respectively.

5.2.3 Enhanced sampling with PT-WTE

Detailed descriptions including underlying assumptions and the simulation protocol for well-tempered ensemble and PT-WTE are available elsewhere [155,156]. Here we present a brief summary of this enhanced sampling method, our application of it, and distinguish the methodology applied in contrast to our recent publications [156,157].

The metadynamics algorithm [158] is typically used as an enhanced sampling method for overcoming large energy barriers by biasing the collective variables (CVs) that govern important modes of motion and forcing the system to explore new areas of phase space with respect to these CVs. Collective variables are functions of the atomic positions of the system that can be calculated at any of the simulation's time steps. A simple example of a CV is the distance between two atoms and a more complex example is the system's potential energy (which is evaluated at every simulation time step). The bias implemented in metadynamics takes the form of small Gaussian-shaped hills, centered at the current value of the CV, and is added to the system's Hamiltonian every τ time steps. These hills gradually fill the local free-energy surface in the form of a history dependent bias. The well-tempered extension to metadynamics [159] limits the amount of bias that can deposited in any region of the CV-space by allowing the hill height to decay exponentially with the amount of bias deposited previously at the given CV value. Subsequently the bias potential for well-tempered metadynamics with a single collective variable (denoted *s*) takes the form of:

$$v(s(r),t) = \sum_{n=1}^{t'=n\tau<t} \omega \bullet \exp\left[\frac{-V(s(r),t')}{k_B \Delta T}\right] \bullet \exp\left[\frac{-(s(r)-s(r(t')))}{2\sigma^2}\right] \quad (5.1)$$

The initial hill height – in units of energy/mole – is denoted by ω and is modified by the exponential term containing the current value of the bias potential. The final exponential term within the summation describes the Gaussian hill of σ width that is added every τ time steps. ΔT represents a user-defined virtual temperature that in practice is implemented as the bias factor:

$$\gamma = \frac{T + \Delta T}{T} \quad (5.2)$$

The PT method utilizes multiple non-interacting replicate simulations that are spaced over a distribution of temperatures. These replicas periodically exchange coordinates according to the Metropolis criterion [160] in a way that preserves detailed balance and canonical sampling at each temperature. The method is useful for enhanced sampling of slow degrees of freedom without prior distinction or specific bias of the important modes of motion. However, the PT method is burdened by high computation cost due to energy fluctuations scaling proportionally to $N^{1/2}$, whereas the average potential energy of a system scales on the order of N. This leads to the requirement of large numbers of replicas to cover a small temperature range and maintain satisfactory exchange frequency in larger systems. It is well-known that combining PT and metadynamics (PTMetaD) can accelerate sampling and convergence [161], however this method does not address the severe limitation imposed by large system sizes.

Recent exploration of the use of the systems potential energy as a biased CV with well-tempered metadynamics has shown that energy fluctuations in the system can be increased while preserving the canonical average [155] (illustrated in Figure 5.2). However, as systems under the influence of this metadynamics bias and augmented energy fluctuations do not conserve the canonical distribution, they have been dubbed to be in the well-tempered ensemble (WTE) and the scale of the fluctuations increases monotonically with the well-tempered metadynamics bias factor (γ). This advance has subsequently

been applied toward amplifying the energy fluctuations during PT simulations and increasing the probability of replicas exchanging configurations. We have shown that PT-WTE can reduce the computational cost of PT simulations with negligible effects on canonically averaged observables for all-atom systems [156]. Another recent study by Sutto and Gervasio applied this method to explore the conformational free-energy surface of a receptor tyrosine kinase enzyme and examined the effects of single and double mutations [162]. In contrast to these other studies, here we present the first use of PT-WTE in an all-atom system without additional metadynamics bias on CVs related to our system configuration or modes of motion.

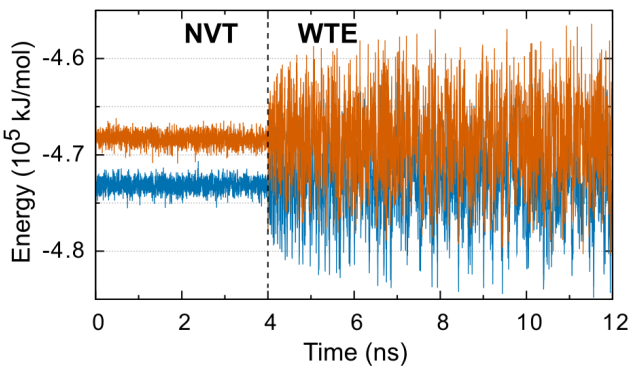


Figure 5.2: Energy fluctuations are increased by WTE bias: comparing the lowest two temperature replicas (310 K, blue and 319 K, orange) before (left) and after (right) the addition of WTE bias. Data was collected during the initial simulations used to determine the appropriate temperature spacing. The NVT and WTE data are taken from independent simulations, without and with the addition of PE bias, respectively, and do not represent a continuous simulation.

Prior to production PT-WTE simulations several parameters and system properties must first be determined. For this study we used a bias factor of 30, a temperature range of 310 to 450 K, a hill deposition rate of 200 simulation steps, an initial hill height of 4.0 kJ/mol, and a PT exchange attempt every 100 simulation steps. The σ value (Gaussian hill width) of 665 kJ/mol for the energy CV was determined from the standard deviation of the non-oxidized system's potential energy (PE) during a 2 ns NVT simulation at 310 K starting from the heated coordinates. The initial coordinates of the non-oxidized system are used to heuristically determine the two lowest temperatures for which PT-WTE simulation with a bias factor of 30 gives an average exchange probability of 30% after 10 ns. Determining the lowest two adjacent temperatures allows calculation of the number of replicas and

correct spacing of replica temperatures to give constant exchange probability over the desired temperature range of 310 to 450 K [163], which was then used to space 14 replicas over this span.

PT-WTE simulations were implemented with the PT framework of GROMACS 4.5 along with the metadynamics plugin, PLUMED [164]. Each replica used initial coordinates corresponding to the final frame from the nearest temperature as generated during the initial system heating. In contrast to our previous publications, the current implementation of PT-WTE actively accumulated bias throughout the entire simulation and no preliminary simulation is used to create a static bias potential in the PE space. Due to the large system size and gap in PE between adjacent replicas there is initially no exchange between replicas during the early parts of the simulation (Figure 5.2). We stress that without the PE bias and the WTE methodology, neighboring replicas with this temperature spacing will not exchange configurations. The exchange probability converges to 30% within the first 5 ns and remains near this probability during the remainder of the simulation for all replicas.

Additional constraints were used to restrict the conformational explorations to structures with the individual hairpins intact. These restraints were implemented as two harmonic potentials on the Cα RMSD from the initial model for each hairpin individually. Therefore, the system is restricted to exploring configurations of the peptide with both hairpins intact yet not confined to the initial anti-parallel beta sheet structure. The harmonic potential was set to exert bias when the RMSD becomes higher than one standard deviation from the model structure (0.30 nm and 0.22 nm for the large and smaller hairpins, respectively), which was determined from the preliminary NVT simulations used to test the model stability, as described above. The spring constant for the potential was set to reach a value of 100 kJ/mol when the RMSD was two standard deviations from the initial model.

5.2.4 Analysis

All analyses presented in this study were completed using the tools within GROMACS 4.5, the driver utility of PLUMED, and VMD 1.8 [68] in addition to custom scripts. Specifically, calculations involving the CV and PMF analysis used the driver utility of PLUMED. The functional form of the PMF, an estimate of the free-energy surface along a CV, is described elsewhere [165] and the CVs are described in the PLUMED documentation. Trajectory visualization and rendering of molecular representations was completed in VMD.

The clustering analysis used the g_rms and g_cluster tools from GROMACS. The g_rms tool was used to calculate an initial RMSD matrix from the lowest temperature trajectories using frames from every 20 ps after the first 10 ns of simulation, for a total of 14500 frames for each simulation. This matrix was used with g_cluster to cluster the frames by the Jarvis-Patrick method [166]. Jarvis-Patrick distinguishes clusters by comparing the intra-frame RMSD of frames within a cutoff and assigning cluster members if they share at least P frames within a cutoff of other cluster members. The cutoff was determined by examining the RMSD matrix to identify the typical distance between neighboring frames and was chosen to be 0.4 nm. Selection of such a large cutoff was necessary due to the flexibility of the protein when in its most extended conformations. The P value for the Jarvis-Patrick method, the number matching neighbors required to form a cluster, was chosen to be 10 by increasing the number from an initial 2 until the top 10 clusters represented clearly distinguishable conformations.

Bootstrapping was used to produce a confidence interval [167] for each bin in the PMF calculation. Specifically, we used every tenth frame (10%) from the full data set to form a reduced data set that was subsequently randomly sampled with replacement to recalculate the PMF estimates. This resampling and recalculation procedure was repeated until the standard deviation of each of the bin estimates converged within 10^{-3} kJ/mol. The bootstrap confidence intervals are displayed in the PMF plots as

shaded regions surrounding the PMF calculated from the reduced data set without random sampling. We used a reduced data set to avoid oversampling and also repeated the bootstrapping procedure, further reducing the data set to every hundredth frame (1%), without any major changes in our results.

5.3 Results

5.3.1 D region

The fibrin D region has been shown to possess a feature critical in the ability of fibrin to aggregate and form proto-fibrils – hole 'a'. The D region has also been shown to contribute in the lateral aggregation of the fibrils, but may not be critical to this process [145]. Its role in this process is hypothesized to involve hole 'b', an indented binding site located on the globular portion of the β chain of Fibrin, interacting with knob 'B', the N-terminal portion of the fibrin β chain after thrombin cleavage of the B-peptide. Figure 5.1A shows a representative snapshot of the modeled protein, with bound GHRP peptides and with methionine sulfoxide, and exemplifies the key location of βMet367 in relation to the 'B:b' knob-hole interaction. We have hypothesized multiple mechanisms by which methionine oxidized to methionine sulfoxide may interrupt the proposed mechanism that are amenable to investigation by MD simulation: methionine oxidation disrupts the conformation of the globular portion of the D region, disrupts the conformation of hole 'b', and/or creates an unfavorable interaction with knob 'B' and destabilizes its bound state. To test these hypotheses we performed 12 simulations of four different systems: the D region either with or without methionine oxidized to methionine sulfoxide and with or without bound knob 'B' surrogates. Analysis of the effects of oxidation on the protein structure and characteristic fluctuations were tracked by the Cα root-mean-square deviation (RMSD) and fluctuation (RMSF) over the course of the 350 ns NPT simulations in addition to the 200 ns replicate simulations.

Measuring the RMSD throughout an MD trajectory allows the overall deviation from the initial structure of the protein to be tracked. Comparing the Cα RMSD, calculated for and aligned to just the D region (residues β202-459 and γ143-393), between the systems with or without oxidation (Figure 5.3) shows that, after sharp increase typical in MD simulation of proteins, the D region equilibrates to a similar deviation from the initial structure regardless of the state of oxidation of βMet367. The final RMSD for the systems without the bound knob 'B' surrogates are also independent of oxidations albeit larger than the systems with the peptides docked. Although the final RMSD values change to a small degree, the replicate simulations verify that indeed the overall structure of the globular portion of the D region are similar between the oxidized and non-oxidized systems on the 350ns time scale and that relatively few changes from the x-ray structure in the D region are observed.

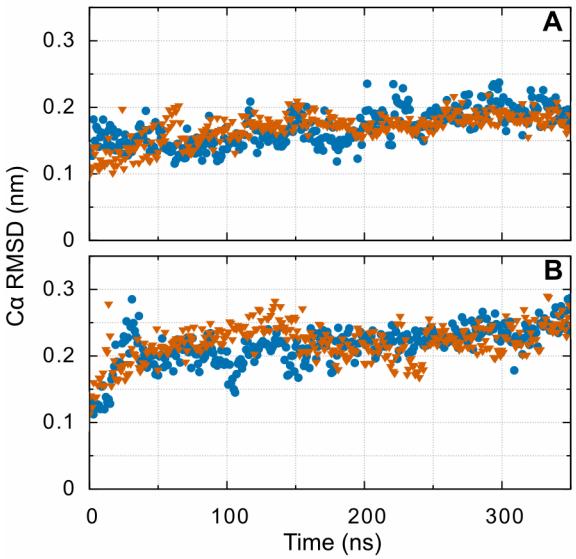


Figure 5.3: D region deviations from the initial structure highlighting the difference between oxidized (orange) and non-oxidized simulations (blue) for (A) the systems with bound GHRP peptides and (B) without bound GHRP peptides. The D region simulations show no augmented deviation created by *β*Met367 oxidation for either group. These features are confirmed in the replicate simulations, although these simulations show some variability in the final RMSD value.

In addition to the structural changes observed for the entire D region we used the trajectory after aligning to the D region to measure the RMSD of just hole 'b' and monitor the structural deviations related to hypothesis surrounding altered confirmations of hole 'b'. This portion of the experimental structure [146] corresponds to residues β352-370 and β381-447. These residues were determined by their proximity to the knob 'B' peptide in the experimental structure. Similar to the results for the whole

D region, hole ‘b’ exhibits a similar final value for both the oxidized and non-oxidized systems (Figure 5.4A). Additionally, the RMSD of these regions without bound GHRP peptides show a larger deviation from the initial structure that is also independent of oxidation. This feature is also seen when performing a similar calculation for hole ‘a’ (residues 289-307and 320-379 γ). These observations, when combined with the RMSF analysis described below, indicate that the increased RMSD in the whole globular domain can be at least partially ascribed to the increased fluctuations of holes ‘b’ and ‘a’ in the absence of a bound knob ‘B’ surrogates.

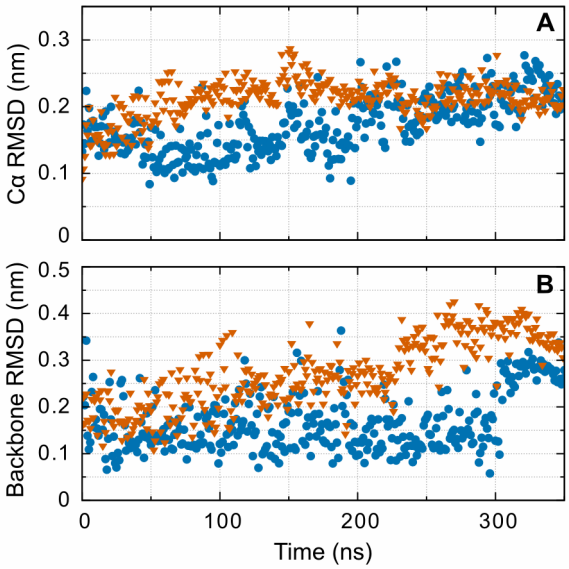


Figure 5.4: RMSD from the initial structure highlighting the difference between oxidized (orange) and non-oxidized simulations (blue) for (A) hole ‘b’ corresponding to residues *β*352-370 and *β*381-447 and (B) the GHRPL fragment bound to hole ‘b’. Note the difference in the vertical axis scale. Like Figure 3, the final RMSD after oxidation is similar to the non-oxidized systems. The sharp increase in backbone RMSD for the GHRPL peptide (knob ‘B’ surrogate peptide) is caused by the solvated proline and leucine backbones reorienting toward increased interaction with hole ‘b’.

The backbone RMSD of the knob ‘B’ fragment itself is used to measure displacement and binding rearrangement of the GHRPL peptide bound to the fibrin hole ‘b’ when calculated from the trajectory aligned to the D region. Backbone RMSD is used here to increase the number of atoms used when calculating the displacement, reducing the noise observed in the Cα RMSD. Examining the results for this calculation (Figure 5.4B) reveals that knob ‘B’ peptide is remarkably stable when bound to hole ‘b’ of fibrin, with only a slight increase in final RMSD in the oxidized system. Repeating this calculation for the knob ‘B’ peptide bound to hole ‘a’, a location physically removed from the oxidized βMet367, along with the results from the replicate simulations reveal that this increase is negligible and

likely unrelated to methionine oxidation. However, analysis of the trajectory does reveal differences in how βMet367 interacts with surrounding residues upon oxidation. Viewing the trajectory reveals that, in the non-oxidized system, βMet367 interacts and stays in close contact with the histidine and proline of the GHRP peptide. In the oxidized system, this changes from a primarily hydrophobic interaction to hydrogen-bonding-like electrostatic interaction between the methionine sulfoxide oxygen and either the backbone nitrogen between histidine and arginine during the first half of the simulation or to the histidine side chain later in the simulation. The sharp, nearly 0.1 nm increase in RMSD for both of these systems, near the end of the simulations, is due to the backbone of residues proline and leucine, of the knob 'B' peptide, reorienting toward increased interaction with hole 'b', as opposed to being primarily solvent exposed. In the oxidized system this leads to hydrogen bonding between the knob 'B' C-terminal carboxylate and the leucine side chain with βLys392 and βLeu386, respectively. In the non-oxidized system this leads to closer contact between the βMet367 side chain and the histidine and proline of knob 'B'.

Like the structural deviations described by RMSD, the characteristic fluctuations of the protein can be described by the Cα RMSF calculated over the equilibrated portion of the trajectories. Simulations were deemed equilibrated after the relatively sharp increase in RMSD observed during the first several nanoseconds. Taking the equilibration period to be the first 10 ns of NPT simulation ensured that each measurement contained the same number of structures. To aid in the description of changes in RMSF between the systems, per-residue difference in the RMSF, rather than the magnitude of the individual RMSF calculations, was plotted by subtracting the results of the non-oxidized from the oxidized systems (Figure 5.5). Therefore, positive values indicate larger fluctuations in the oxidized system while negative values indicate larger fluctuations in the non-oxidized system.

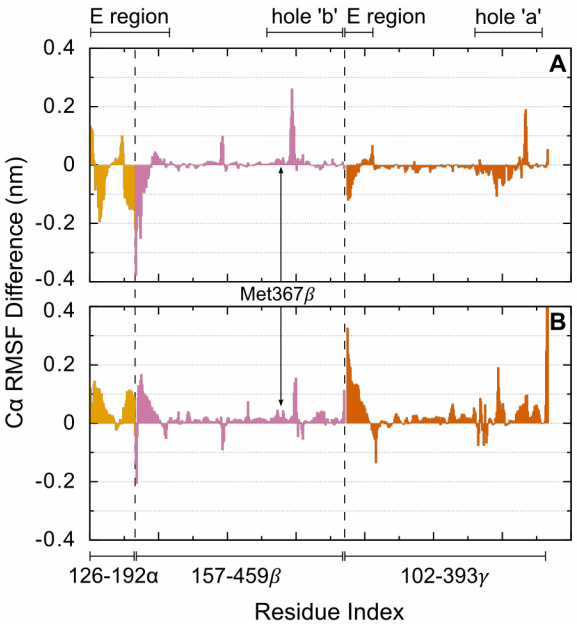


Figure 5.5: Cα fluctuation differences between the oxidized and non-oxidized simulations for systems (A) with bound GHRP peptides and (B) without bound GHRP peptides. Colors differentiate the multiple protein chains that comprise the modeled system and correspond to the colors used in Figure 5.1A. Dashed lines indicate the division between each chain. Positive and negative values indicate increased fluctuations in the oxidized and non-oxidized systems, respectively.

Figure 5.5 highlights several regions of the enzyme with increased fluctuations or large differences in fluctuations upon oxidation. The coiled-coil E region of the protein shows the largest RMSD and RMSF amongst all the simulations particularly in residues near the point of truncation in the crystal structure. This section of the peptide consists of 4 helical coils, which are contributed from each of the α, β, and γ chains, and is stabilized by multiple, inter-chain disulfide bonds. However, these disulfide bonds do not prevent translation of the helix bundle relative to the D region nor create stability near the point of truncation. These domain motions in the E region lead to relatively large deviations in the protein structure that are likely stochastic and unrelated to methionine oxidation. The other remarkable regions, in terms of structural fluctuations, are holes 'a' and 'b' whether with or without bound knob 'B' surrogate peptides. In all simulations these residues exhibit largest fluctuations of all the D region and exhibit some structural reorganization from the crystal structure toward solvating several hydrophilic amino acids. This is particularly evident in the 0.24 nm increase in RMSF at aspartic acid β383 for the oxidized system with bound knob 'B' over its non-oxidize counterpart. There is a similar increased RMSF and increased RMSF difference for this residue in the system without knob 'B' peptides but the difference seems comparatively insignificant. Viewing the trajectory of the oxidized

system with bound GHRPL shows that this aspartic acid residue moves from being partially buried near the surface of the protein in a network of side-chain-to-backbone hydrogen-bonding interactions toward becoming fully solvent exposed and forming multiple short-lived interactions with nearby hydrophilic amino acids. This aspartic acid and its sequence neighbors form interactions with the knob 'B' peptide toward the end of this simulation but are beyond the non-bonded cutoffs from βMet367. Overall, comparing the Cα fluctuations reveals little to no evidence that oxidation increases the fluctuations of any region of the peptide at the 340 ns time scale.

5.3.2 αC-subdomain

The αC-subdomain displays a dynamic character that is thought to be critical toward its role in lateral aggregation of fibrin [168,169]. The dynamic character has been hypothesized to include an "opening" of the double hairpin structure, characterized by a loss of contacts between the two hairpins and a hinge-like transition toward a linear and extended state. This feature is hypothesized to allow formation of large beta sheet structures with other fibrin molecules [168]. Opening and closing of the hairpins is explored in this study through the enhanced sampling added by the PT-WTE method. Due to the nature of the large structural variations upon opening and the frequency of exchanges with neighboring replicas, the trajectory for a single temperature contains a large and discontinuous variety of structures. When combined with the increased number of accessible configurations after the hairpins lose contact, this feature makes analysis by RMSD and RMSF less effective in describing the effects created by methionine oxidation. Therefore, in contrast to the analysis presented for the D region, the PT-WTE simulations were analyzed using RMSD clustering and supported by the potential of mean force (PMF) of multiple CVs. These analyses allow a description of the most preferred structures of the αC-subdomain and require the enhanced sampling of peptide conformations provided by PT.

Jarvis-Patrick clustering analysis was applied to the backbone RMSD between frames after aligning to the larger of the two hairpins, corresponding to residues 443 to 465 α. This alignment creates increased RMSDs for protein structures at different degrees of extendedness and can still distinguish structures where the hairpins are still in contact (closed) yet in a different configuration compared to the original model. The heterogeneity of observed structures in the fully extended (open) conformation required a high clustering cutoff. Clusters represent the most likely conformations of the peptide according to the conformations explored at a given temperature. Indeed the top ten most populated clusters for the oxidized and non-oxidized systems contain 44% and 48% of the frames used in the clustering calculation, respectively. Viewing the structures included in each cluster along with the cluster representative, which correspond to the "center-most" member of the cluster, reveals that of the top ten most populated structures of each simulation show a discrepancy in preference for the open versus closed conformation between the two simulations. The most populated cluster for both the oxidized and non-oxidized systems at 310 K are similar in configuration to the original model. Figure 5.6 shows that

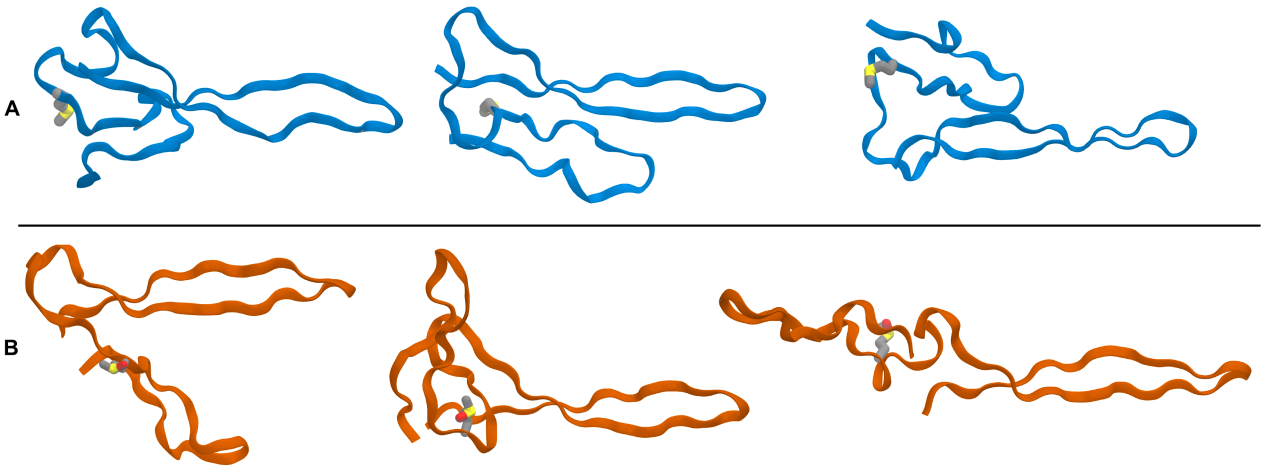


Figure 5.6: Ribbon representations of the cluster center from top three most populated clusters of the (A, blue) non-oxidized and (B, orange) oxidized αC-subdomain simulations. Clustering was calculated from the inter-frame Cα RMSD of the 310 K replicas. The ranking of the structures decreases from left to right; the leftmost structure is most populated cluster for each system. The side-chain of αMet476 is shown in each representation. The flexibility of the "hairpin-connecting region" affords a large variety of associations between the two hairpins and permits a hinge-like transition toward an extended "open" conformation (similar to the third ranking cluster of the oxidized system).

the first and third highest populated clusters in the oxidized system correspond to a partially open and fully extended configuration, respectively. In contrast, the non-oxidized system the top 4 most populated clusters indicate multiple closed states are most prevalent in the trajectory. The fifth most populous cluster, representative of 7% of the clustered frames, exhibits the first open structure in the non-oxidized system. Of the top ten clusters in each system, the oxidized system has 4 clusters representing open or partially open structures (23 % of the clustered frames) while the non-oxidized system has 2 (11 % of the clustered frames). These data indicate that either methionine oxidation disturbs the stability of the closed configurations or stabilizes particular opened structures allowing them to contribute clusters with relatively higher population.

Several CVs describing aspects of the protein structure were calculated over the trajectories at each temperature. The potential of mean force (PMF) were in turn calculated from the histogram of CV values. These estimates of the free-energy surface along the CV coordinates entail characteristics of the preferred configurations for the αC-subdomain over the 300 ns PT-WTE simulations and are rescaled to set the lowest energy feature to zero. PMF calculations were recalculated at 80-100% of the trajectory in increments of 5% to check for convergence of the estimated free-energy surface features. We used the backbone radius of gyration of the peptide to examine the stability toward the hypothesized native structure, which has a low radius of gyration. Figure 5.7 compares the calculated PMF at 310K between the oxidized and non-oxidized peptide. The PMF for the non-oxidized system describes a deeper well near a radius of gyration typical in closed configurations when compared to the oxidized system. The corresponding well for oxidized system is much broader in the radius of gyration, showing that a variety of structures – both slightly lower and higher in radius of gyration – contributes to the closed configuration in the oxidized system. Furthermore, there is a distinct well in the PMF for the oxidized system at large radius of gyration. This well is not as distinguished in the non-oxidized system. The

frames corresponding to this value of radius of gyration are similar to the fully extended configuration seen in the third most populous cluster from the oxidized system, as shown in Figure 5.6. These features of the free-energy surface along the radius of gyration indicate that methionine oxidation has caused a

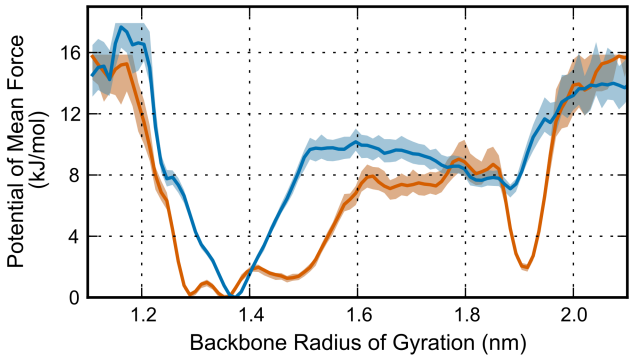


Figure 5.7: Comparing the PMF of the backbone radius of gyration of the entire α C-subdomain between the oxidized (orange) and non-oxidized (blue) systems. PMFs were calculated from trajectories of the 310 K replicas using every tenth frame to avoid oversampling. The well near 1.9 nm radius of gyration corresponds to fully extended structures similar to the third ranking cluster of the oxidized system, which can be seen in Figure 5.6. Shaded regions represent a 99% confidence interval calculated using bootstrapping.

relative destabilization of the most preferred radius of gyration in the oxidized system and exhibits a higher proportion of open structures, particularly those similar to the third most populous cluster from the oxidized system. When combined with the comparison between the clusters calculated for each system, this analysis reinforces the observation that oxidation of methionine contributes to a more open configuration or possibly destabilizes the structure corresponding to the original model.

Viewing the trajectory and the cluster centers shows that for both systems the hairpin-hairpin interaction seen in the closed structures is not restricted to the anti-parallel beta sheet like conformation seen in the original model. The structures in the closed conformation reveal that a variety of hairpin-hairpin interactions occur with multiple orientations of the hairpins, including “face-to-face” hairpin stacking and side-by-side beta sheet on the opposite side compared to the starting model. The multiple orientations allowed by the hairpins are supported through the high proportion of polar or charged amino acids comprising the hairpin regions of this peptide. Of the 56 amino acids comprising the modeled system, 31 have side chains capable of forming hydrogen bonds. These amino acids associate through hydrogen bonds with other side-chains, forming salt-bridges, and with the backbone oxygen and nitrogen atoms. This aspect of the composition of this subdomain allows for stabilizing interactions from

multiple orientations of the two hairpins beyond the backbone hydrogen-bonding characteristic in beta-sheets.

The sequence of amino acids forming the link between the hairpins, which includes αMet476, confers much of the flexibility allowing the wide variety of structures seen in the closed state and serves as the hinge in the opening transition. Visual comparison between the cluster centers in both simulations shows that different orientations, especially when comparing between open and closed structures, are supported by extension and reorientation of the structure in this hairpin-linking region. We calculated the PMFs for the backbone RMSD, calculated from the initial model, and the backbone radius of gyration CVs for the hairpin-linking region. Figure 5.8 shows that, when comparing the estimate of the free-energy surface for these collective variables between the two systems, there are differences created by methionine oxidation. Specifically, the RMSD of the linking region in the oxidized system readily accesses a much broader distribution of structures compared to the linking region. This is interpreted from the PMF by a lower energy, relative to the minimum, up to 3.2 nm radius of gyration when

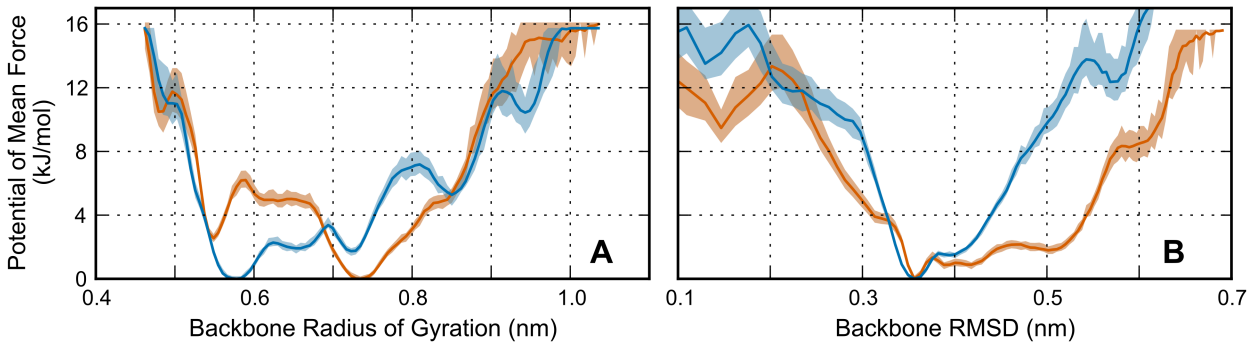


Figure 5.8: Comparing the PMF of the backbone (A) radius of gyration and (B) RMSD of the hairpin-linking region between the oxidized (orange) and non-oxidized (blue) systems. PMFs were calculated from trajectories of the 310 K replicas using every tenth frame to avoid oversampling. Both of these plots exhibit features indicative of a more extended linking region in the oxidized system. Shaded regions represent a 99% confidence interval calculated using bootstrapping.

compared to the PMF of the non-oxidized system. This feature of the PMF indicates that a larger variety of structures, particularly those that deviate more from the initial model, are relatively more stable in the

oxidized system. The PMF calculated from the backbone radius of gyration of the linking region supports this observation and reveals that the most visited configurations during the oxidized simulation have over a 0.15 nm increase in radius of gyration over the most visited conformations of the non-oxidized system. This amount may appear small at first glance but given the relatively small size of the connecting region, a 0.15 nm corresponds to significantly different structures. In summary, the features of the PMFs for these two CVs indicate that the oxidation of methionine alters the conformations of the connecting region toward increased deviation from the original model and an increased extendedness.

5.4 Discussion

On the 350 ns time scale, the MD simulations of the fibrin D region provide no indication of an alteration in the D region structure or characteristic fluctuations – either as a whole or when confining our analysis to hole 'b'. Over the same time scale these simulations do not support the hypothesis that methionine oxidation disrupts the stability of the bound knob 'B' peptide. However, the enhanced sampling simulations indicate clear differences in the preferred configurations of the αC-subdomain indicative of oxidation influencing the structure's preference away from the assumed native conformation. We ascribe the disruption of the αC-subdomain to observed changes in the conformation of the hairpin-linking region, the segment of the subdomain that contains αMet476.

Experimental investigations detailed in the current literature have established that the αC-domain plays a critical role in the lateral aggregation of fibrin [142,170]. Additionally, the experiments conducted by Weigandt et al. [141] show that oxidized fibrin gels exhibit similar properties to fibrin gels without functional αC-domains [142]. This work contributes to a growing body of evidence that the oxidation of the αC-domain is one likely contributor to the mechanism of disrupted lateral aggregation of oxidized fibrin. However, the molecular mechanism of αC-domain influenced aggregation has yet to be established. One hypothesis is that the double hairpin structure of the N-terminal subdomain forms

extended beta-sheet structures through its interaction with the same subdomain of nearby fibrin molecules [168]. The disruption of the preferred conformation of this subdomain could, thus, influence its ability to interact favorably in this interaction. However, until the mechanism of αC-domain influenced lateral aggregation is known, the importance of the native structure in this mechanism cannot be established.

Reactive oxygen species (ROS) and ROS-generating oxidases are known to play important roles as signaling molecules in the vasculature. Many cellular components including platelets, endothelial cells, and polymophonuclear cells release ROS in response to tissue injury and ischemia. Lipoxygenases and cyclooxygenases in platelets can make superoxide via NADPH oxidases, nitric oxide from nitric oxide synthases, and ultimately peroxynitrite via interaction of superoxide with nitric oxide [143]. Specific to our study, polymorphonuclear cells (neutrophils) produce hypochlorite by the interaction of myeloperoxide and hydrogen peroxide, which is also key host-defense mechanism mediating bacterial killing by the formation of methionine sulfoxide in bacterial membrane proteins [171]. Therefore, our results identifying a change in fibrin αC-domain structure as a result of selective methionine sulfoxide formation perhaps suggest that neutrophil-specific oxidation can also regulate fibrin clot structure. Heffron et al found further evidence for this process taking place in vivo, after inducing experimental endotoxemia in healthy humans. The authors found that endotoxemia-induced fibrinogen nitration was preceded by a stable increase in myeloperoxidase concentration in plasma [172]. Interestingly, neutrophil-mediated methionine oxidation by hypochlorite can also dose-dependently inhibit other important coagulation proteins including activated protein C [173], and thrombomodulin. In the case of thrombomodulin, inhibition has been mapped to oxidation of a single methionine residue (Met388) [174]. Taken together, the influence of methionine sulfoxide formation on the key regulators of protein

C, thrombomodulin, and fibrinogen also suggests a possible wider role for oxidative stress as a mechanism of regulation of coagulation function by neutrophil inflammation.

Several assumptions were required to conduct targeted simulation studies of oxidation's effect on fibrin. In particular, our investigation into the αC-subdomain hypotheses requires the assumption that the human homology model of the bovine NMR structure is an appropriate representation of this subdomain. Furthermore, we have restricted our conformational exploration of this peptide by restraining the secondary structure of each hairpin near its initial configuration. This was done to avoid exploration of the complete conformational ensemble of this peptide – an endeavor that would require unreasonably long simulation times. Therefore, our simulations cannot investigate an alternative hypothesis: that the oxidation of αMet476 completely disrupts the preferred fold of the αC-subdomain through its interactions with nearby residues. Additionally, the D region simulations are not sufficient to investigate another alternative hypothesis: that oxidation of βMet367 interferes with knob 'B's initial docking with hole 'b'. Another consideration is that oxidation within the D region occurs in an area of high stability with well-defined secondary structure, whereas the oxidation in the aC subdomain is in a region comprised of a flexible loop, which is expected to have greater mobility.

Our results and untested hypotheses suggest several opportunities for future simulation work to further investigate methionine oxidation of fibrin. In particular, the hypothesized aggregation mechanism involving multiple αC-domains forming extended structures suggest that methionine oxidation might interrupt lateral aggregation by disruption of this interaction. Additionally, the altered conformational preference described in this study after methionine oxidation could influence the interaction between two αC-subdomains. Docking studies or metadynamics simulations could be used to test this hypothesis by studying the interaction of pairs of αC N-terminal subdomains with and without methionine oxidation. Furthermore, the 'B:b' knob-hole interaction could be further examined by a

metadynamics docking study to calculate the free energy surface of a knob ‘B’ surrogate docking to an oxidized hole ‘b’. Additionally, future computational work using GPUs or millisecond-time MD simulation machines [175] could contribute in both of these areas.

It may also be useful to seek to identify similar oxidative modifications in diseases where oxidative stress and coagulopathy are known to co-exist. Hemorrhagic shock from blood loss in trauma is a prime example of such a condition. Tissue hypoxia from blood loss paradoxically increases oxidation both from mitochondria [176] and from activation of NADPH oxidases [177]. Coagulopathy is also almost immediately present in almost 25% of severely injured trauma patients, requires a high degree of cardiovascular shock with tissue hypoperfusion, and results in a 4-6 fold increased mortality [178-182]. Moreover, functional fibrinogen concentration and fibrinolysis are both strongly associated with mortality in trauma patients [183,184] and both can be strongly influenced by methionine oxidation. However, further study is required to delineate any potential role for fibrinogen oxidation in the pathophysiology of traumatic coagulopathy.

6 Conclusion

6.1 Significance and Contributions of this Research

While we are not the first group to publish results from simulation studies of proteins in ILs, our rigorous experimental design, thorough analysis, and high standards for publishable results (e.g., in terms of proper equilibration of the MD simulations) will set the standard for future contributions throughout this small but growing field. When starting our first study into the interaction between ILs/enzymes, described in Chapter 2, the state of published theoretical investigations in this area – to our knowledge – were insufficient compared to the current capabilities of MD simulations. Within the few peer-reviewed articles we have identified, there were no studies before ours to publish analysis of equilibrated simulation trajectories spanning more than tens of nanoseconds. Our initial study in near-pure ILs and the subsequent study by my colleague Vance Jaeger – using more dilute simulations of ILs in water – span the hundreds of nanosecond timescales that are standard in current literature. Our work clearly establishes that significantly longer simulations times are of critical importance to fully equilibrate the solutes within these viscous solvents and to examine changes in domain motions in large proteins. The subsequent collaborative study with Vance (c.f., Chapter 3) utilized the longest simulations of IL/enzyme systems to date, with individual simulations using 500 ns of sampling and the sum total simulation time of 27 μs. Lastly, our group will likely be the first to apply state-of-the-art enhanced sampling techniques such as PT-WTE – which is novel within this field and generally – to drive or enhance our investigations of these important interactions. As described in Section 6.2, this capability will enable us to probe aspects of the IL/protein interactions that are yet unresolved at any level of detail.

Using simulation to predict experimental results is a highly sought goal in computational research. We hope to achieve this by providing suggested experiments toward confirming observations made based on our simulations. This is a prominent feature of the discussion and conclusions of the manuscript provided in Chapter 3. However, our primary focus is the golden standard of molecular simulation – to compliment established experimental results with atomistic details. We accomplish this in each project described in Chapters 2-5, either through collaboration, like the work detailed in Chapters 4 and 5, or using exemplary publications within the field to guide our investigations, like the work described in Chapters 1 and 3.

Through our support by and comparison to experiment in these projects, we have established a methodological basis for future exploratory simulations. While my individual projects have well-identified research objectives, these studies also establish the methodology and meaningful analysis that will facilitate more complex investigations. This will help future researchers avoid the heuristics required in these early studies and waste of computational resources, both within our group and potentially throughout the field. Although it is a long-term goal of our group and not a specific goal for my project, these pioneering simulation studies into IL/enzyme interactions will contribute significantly in establishing a high-throughput methodology for predicting IL tolerance toward the ultimate goal of virtually screening for beneficial ILs.

Historically, it has been extremely challenging to quantify the changes single point mutations cause in enzyme structure using molecular simulation. This is possibly due to the lack of sensitivity in the force fields or a lack of sufficient sampling in simulations investigating these mutations (e.g., short simulation times or not using applicable enhanced sampling methods). Through the work described in Chapter 5, we have shown that parallel tempering provides sufficient sampling to observe changes caused by the addition of a single atom in the double hairpin structure of the human fibrin αC

subdomain. Additionally, we accomplished sufficient sampling through the application of methods that are computationally more efficient than traditional parallel tempering; based upon previous work within our group [156], we estimate an order of magnitude fewer replicas are required to cover the same temperature range when using the PT-WTE methodology. Although the PT-WTE method remains expensive for large (> 20,000 atom) systems, proper application of this technique could enable computational investigations across many fields of interest.

6.2 Future Work

Following the work presented here, our immediate focus is on the analysis of a series of enhanced sampling simulations, enabled by the PT-WTE methodology described in Chapter 5, involving the changes in the free energy landscape of peptide folding caused by increasing concentrations of ionic liquids. Specifically, we have completed a series of all-atom PT-WTE simulations with the addition of metadynamics to bias the folding of a tryptophan cage peptide in four concentrations of three different ionic liquids. From these simulations we will calculate the free energy difference between the folded and unfolded states and construct folding "phase diagrams" as a function of IL concentration and temperature. The hypothesis that denaturing ionic liquids stabilize the unfolded state of the protein by "salting in" hydrophobic residues [26] will be examined in the context of the phase diagrams. Reweighted clusters for the trajectories will enumerate the observed partially folded intermediates. Examining the free energy difference of these intermediates compared to the folded peptide will also allow us to examine the hypothesis that deactivating ILs stabilize partially folded intermediates. This work will likely be followed by extension to more complex peptides, such as the fast-folding NTL9, which possesses more secondary structure elements and a more pronounced hydrophobic core compared to tryptophan cage.

As a follow up to the solvent organization study presented in Chapter 4, the analyses we have developed to examine the structuring and interaction of IL substituents with the protein surface can be applied to simulation trajectories from our previous studies. Specifically, calculation of the charge density and radial distribution functions about the enzyme CRL in near-pure ILs, from the trajectories described in Chapter 2, will allow us to determine if high concentration ILs exhibit structural organization on length-scales beyond the first solvation shell. Analysis of the trajectories from Chapter 2 will also enable us to examine organization of non-denaturing ILs, which was not done in the study described in Chapter 4. Additionally, the trajectories of three cellulase enzymes in increasing concentrations of IL, described in Chapter 3, will allow us to examine the affect of increased temperature on IL organization and also to determine if the critical concentration required for long-range IL structuring is within the range we have examined (up to 50 wt.% IL). These trajectories will allow us to further examine the affects of surface composition, since each cellulase has a different number of charged surface atoms. Finally, since the cellulase simulations span 500 ns, the trajectories can also be used to examine the time dependence of IL organization and test our assumption (Chapter 4) that the solvent is well equilibrated during our 50 ns simulations.

There is also opportunity in future simulation studies to explore the interaction strength of IL substituents with amino acid side chains. Free-energy methods such as the recently developed "funnel metadynamics" [185] can be applied to calculate the binding free energy profiles of IL ions with the binding site of enzymes. Pairing the simulation results with experiments measuring the Michaelis-Menten constant of said enzymes will clarify which ions are responsible for the observed competitive inhibition, with potential to propose design rules to protect the enzyme active site from ion penetration. Additionally, metadynamics or umbrella sampling simulations of "GXG" tri-peptides, where X represents any of the 20 standard amino acids, can be used to calculate the free energy of IL interactions

with amino acid side chains. Knowledge about IL/protein interactions gained from such a study will enhance our ability to predict how enzyme surface characteristics mediate IL tolerance.

Bibliography

(1) Welton T (1999) Room-Temperature Ionic Liquids. Solvents for Synthesis and Catalysis. Chem Rev (Washington, DC, U S) 99: 2071-2084.

(2) Seddon KR (1997) Ionic Liquids for Clean Technology. J Chem Technol Biotechnol 68: 351-356.

(3) Hallett JP, Welton T (2011) Room-Temperature Ionic Liquids: Solvents for Synthesis and Catalysis. 2. Chem Rev (Washington, DC, U S) 111: 3508-3576.

(4) Seddon KR (2003) Ionic Liquids: A Taste of the Future. Nat Mater 2: 363-365.

(5) Carlisle TK, Bara JE, Gabriel CJ, Noble RD, Gin DL (2008) Interpretation of Co2solubility and Selectivity in Nitrile-Functionalized Room-Temperature Ionic Liquids Using a Group Contribution Approach. Ind Eng Chem Res 47: 7005-7012.

(6) Fort DA, Remsing RC, Swatloski RP, Moyna P, Moyna G, et al. (2007) Can Ionic Liquids Dissolve Wood? Processing and Analysis of Lignocellulosic Materials with 1-N-Butyl-3-Methylimidazolium Chloride. Green Chem 9: 63.

(7) Swatloski RP, Spear SK, Holbrey JD, Rogers RD (2002) Dissolution of Cellose with Ionic Liquids. J Am Chem Soc 124: 4974-4975.

(8) Binder JB, Raines RT (2009) Simple Chemical Transformation of Lignocellulosic Biomass into Furans for Fuels and Chemicals. J Am Chem Soc 131: 1979-1985.

(9) Ha SH, Lan MN, Lee SH, Hwang SM, Koo Y-M (2007) Lipase-Catalyzed Biodiesel Production from Soybean Oil in Ionic Liquids. Enzyme Microb Technol 41: 480-483.

(10) Guo Z, Xu X (2006) Lipase-Catalyzed Glycerolysis of Fats and Oils in Ionic Liquids: A Further Study on the Reaction System. Green Chem 8: 54.

(11) Xin J-Y, Zhao Y-J, Shi Y-G, Xia C-G, Li S-B (2005) Lipase-Catalyzed Naproxen Methyl Ester Hydrolysis in Water-Saturated Ionic Liquid: Significantly Enhanced Enantioselectivity and Stability. World J Microbiol Biotechnol 21: 193-199.

(12) Roberts NJ, Seago A, Carey JS, Freer R, Preston C, et al. (2004) Lipase Catalysed Resolution of the Lotrafiban Intermediate 2,3,4,5-Tetrahydro-4-Methyl-3-Oxo-1h-1,4-Benzodiazepine-2-Acetic Acid Methyl Ester in Ionic Liquids: Comparison to the Industrial T-Butanol Process. Green Chem 6: 475.

(13) Cerqueira Pereira S, Bussamara R, Marin G, Lima Camargo Giordano R, Dupont J, et al. (2012) Enzymatic Synthesis of Amoxicillin by Penicillin G Acylase in the Presence of Ionic Liquids. Green Chem 14: 3146.

(14) Lozano P, de Diego T, Guegan J-P, Vaultier M, Iborra JL (2001) Stabilization of A-Chymotrypsin by Ionic Liquids in Transesterification Reactions. Biotechnol Bioeng 75: 563-569.

(15) Lozano P, De Diego T, Carrié D, Vaultier M, Iborra JL (2003) Enzymatic Ester Synthesis in Ionic Liquids. J Mol Catal B: Enzym 21: 9-13.

(16) Ulbert O, Belafi-Bako K, Tonova K, Gubicza L (2005) Thermal Stability Enhancement of Candida Rugosa Lipase Using Ionic Liquids. Biocatal Biotransform 23: 177-183.

(17) Lozano P, De Diego T, Carrié D, Vaultier M, Iborra JL (2001) Over-Stabilization of Candida Antarctica Lipase B by Ionic Liquids in Ester Synthesis. Biotechnol Lett 23: 1529-1533.

(18) Park S, Kazlauskas RJ (2001) Improved Preparation and Use of Room-Temperature Ionic Liquids in Lipase-Catalyzed Enantio- and Regioselective Acylations. J Org Chem 66: 8395-8401.

(19) Ulbert O, Fráter T, Bélafi-Bakó K, Gubicza L (2004) Enhanced Enantioselectivity of Candida Rugosa Lipase in Ionic Liquids as Compared to Organic Solvents. J Mol Catal B: Enzym 31: 39-45.

(20) Zhang T, Datta S, Eichler J, Ivanova N, Axen SD, et al. (2011) Identification of a Haloalkaliphilic and Thermostable Cellulase with Improved Ionic Liquid Tolerance. Green Chem 13: 2083-2090.

(21) Datta S, Holmes B, Park JI, Chen Z, Dibble DC, et al. (2010) Ionic Liquid Tolerant Hyperthermophilic Cellulases for Biomass Pretreatment and Hydrolysis. Green Chem 12: 338-345.

(22) Zhao H, Campbell SM, Jackson L, Song Z, Olubajo O (2006) Hofmeister Series of Ionic Liquids: Kosmotropic Effect of Ionic Liquids on the Enzymatic Hydrolysis of Enantiomeric Phenylalanine Methyl Ester. Tetrahedron Asymmetry 17: 377-383.

(23) Constantinescu D, Weingärtner H, Herrmann C (2007) Protein Denaturation by Ionic Liquids and the Hofmeister Series: A Case Study of Aqueous Solutions of Ribonuclease A. Angew Chem, Int Ed 46: 8887-8889.

(24) Yang Z (2009) Hofmeister Effects: An Explanation for the Impact of Ionic Liquids on Biocatalysis. J Biotechnol 144: 12-22.

(25) Kamerlin SCL, Warshel A (2010) At the Dawn of the 21st Century: Is Dynamics the Missing Link for Understanding Enzyme Catalysis? Proteins 78: 1339-1375.

(26) van Rantwijk F, Sheldon RA (2007) Biocatalysis in Ionic Liquids. Chem Rev (Washington, DC, U S) 107: 2757-2785.

(27) Zhao H (2010) Methods for Stabilizing and Activating Enzymes in Ionic Liquids-a Review. J Chem Technol Biotechnol 85: 891-907.

(28) Toral AR, de los Ríos AP, Hernández FJ, Janssen MHA, Schoevaart R, et al. (2007) Cross-Linked Candida Antarctica Lipase B Is Active in Denaturing Ionic Liquids. Enzyme Microb Technol 40: 1095-1099.

(29) Burney PR, Pfaendtner J (2013) Structural and Dynamic Features of Candida Rugosa Lipase 1 in Water, Octane, Toluene, and Ionic Liquids Bmim-Pf6 and Bmim-No3. J Phys Chem B 117: 2662-2670.

(30) Jaeger VW, Pfaendtner J (2013) Structure, Dynamics, and Activity of Xylanase Solvated in Binary Mixtures of Ionic Liquid and Water. Acs Chemical Biology 8: 1179-1186.

(31) Nordwald EM, Kaar JL (2013) Stabilization of Enzymes in Ionic Liquids Via Modification of Enzyme Charge. Biotechnol Bioeng 110: 2352-2360.

(32) Nordwald EM, Kaar JL (2013) Mediating Electrostatic Binding of 1-Butyl-3-Methylimidazolium Chloride to Enzyme Surfaces Improves Conformational Stability. J Phys Chem B 117: 8977-8986.

(33) Burney PR, White N, Pfaendtner J (2014) Structural Effects of Methionine Oxidation on Isolated Subdomains of Human Fibrin D and Ac Regions. PLoS One 9: e86981.

(34) Kaar JL, Jesionowski AM, Berberich JA, Moulton R, Russell AJ (2003) Impact of Ionic Liquid Physical Properties on Lipase Activity and Stability. J Am Chem Soc 125: 4125-4131.

(35) Canongia Lopes JNA, Pádua AAH (2006) Nanostructural Organization in Ionic Liquids. J Phys Chem B 110: 3330-3335.

(36) Curtis RA, Ulrich J, Montaser A, Prausnitz JM, Blanch HW (2002) Protein-Protein Interactions in Concentrated Electrolyte Solutions. Biotechnol Bioeng 79: 367-380.

(37) Ghosh T, Kalra A, Garde S (2005) On the Salt-Induced Stabilization of Pair and Many-Body Hydrophobic Interactions. J Phys Chem B 109: 642-651.

(38) Micaêlo NM, Baptista AM, Soares CM (2006) Parametrization of 1-Butyl-3-Methylimidazolium Hexafluorophosphate/Nitrate Ionic Liquid for the Gromos Force Field. J Phys Chem B 110: 14444-14451.

(39) Morrow TI, Maginn EJ (2002) Molecular Dynamics Study of the Ionic Liquid 1-N-Butyl-3-Methylimidazolium Hexafluorophosphate. J Phys Chem B 106: 12807-12813.

(40) Urukova I, Vorholz J, Maurer G (2005) Solubility of Co2, Co, and H2in the Ionic Liquid [Bmim][Pf6] from Monte Carlo Simulations. J Phys Chem B 109: 12154-12159.

(41) Maginn EJ (2009) Molecular Simulation of Ionic Liquids: Current Status and Future Opportunities. J Phys: Condens Matter 21: 373101.

(42) Klähn M, Lim GS, Seduraman A, Wu P (2011) On the Different Roles of Anions and Cations in the Solvation of Enzymes in Ionic Liquids. Phys Chem Chem Phys 13: 1649.

(43) Micaêlo NM, Soares CM (2008) Protein Structure and Dynamics in Ionic Liquids. Insights from Molecular Dynamics Simulation Studies. J Phys Chem B 112: 2566-2572.

(44) Grochulski P, Li Y, Schrag JD, Cygler M (1994) Two Conformational States of Candida Rugosa Lipase. Protein Sci 3: 82-91.

(45) Dave R, Madamwar D (2010) Preparations for the Use of Candida Rugosa Lipase in Non-Conventional Solvents. Biocatal Biotransform 28: 157-166.

(46) Domínguez de María P, Sánchez-Montero JM, Sinisterra JV, Alcántara AR (2006) Understanding Candida Rugosa Lipases: An Overview. Biotechnol Adv 24: 180-196.

(47) Madeira Lau R, Sorgedrager MJ, Carrea G, van Rantwijk F, Secundo F, et al. (2004) Dissolution of Candida Antarctica Lipase B in Ionic Liquids: Effects on Structure and Activity. Green Chem 6: 483.

(48) Turner MB, Spear SK, Huddleston JG, Holbrey JD, Rogers RD (2003) Ionic Liquid Salt-Induced Inactivation and Unfolding of Cellulase from Trichoderma Reesei. Green Chem 5: 443.

(49) Fuciños González JP, Bassani G, Farruggia B, Picó GA, Pastrana Castro L, et al. (2011) Conformational Flexibility of Lipase Lip1 from Candida Rugosa Studied by Electronic Spectroscopies and Thermodynamic Approaches. Protein J 30: 77-83.

(50) Berman H, Henrick K, Nakamura H (2003) Announcing the Worldwide Protein Data Bank. Nat Struct Biol 10: 980.

(51) Grochulski P, Li Y, Schrag JD, Bouthillier F, Smith P, et al. (1993) Insights into Interfacial Activation from an Open Structure of Candida Rugosa Lipase. J Biol Chem 268: 12843-12847.

(52) Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, et al. (2008) Glycam06: A Generalizable Biomolecular Force Field. Carbohydrates. J Comput Chem 29: 622-655.

(53) Levy Y, Onuchic JN (2006) Water Mediation in Protein Folding and Molecular Recognition. Annu Rev Biophys Biomol Struct 35: 389-415.

(54) de la Casa RM, Sánchez-Montero JM, Rojas R, Sinisterra JV (1998) Simple Determination of the Water Content in Lyophilized Isoenzymes of Candida Rugosa Lipase. Biotechnol Tech 12: 823-827.

(55) Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulation Liquid Water. J Chem Phys 79: 926-935.

(56) Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, et al. (2005) The Amber Biomolecular Simulation Programs. J Comput Chem 26: 1668-1688.

(57) Wang J, Wang W, Kollman PA, Case DA (2006) Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. J Mol Graphics Modell 25: 247-260.

(58) Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and Testing of a General Amber Force Field. J Comput Chem 25: 1157-1174.

(59) Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. J Phys Chem 97: 10269-10280.

(60) Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, et al. (2009) Gaussian 09, Revision B.01. Wallingford CT.

(61) Martínez L, Andrade R, Birgin EG, Martínez JM (2009) Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations. J Comput Chem 30: 2157-2164.

(62) Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable Molecular Dynamics with Namd. J Comput Chem 26: 1781-1802.

(63) Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, et al. (2006) Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins 65: 712-725.

(64) Schneider T, Stoll E (1978) Molecular-Dynamics Study of a 3-Dimensional One-Component Model for Distortive Phase-Transitions. Physical Review B 17: 1302-1322.

(65) Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. J Chem Phys 103: 4613-4621.

(66) Andersen HC (1983) Rattle - a Velocity Version of the Shake Algorithm for Molecular-Dynamics Caluclations. J Comput Phys 52: 24-34.

(67) Krautler V, van Gunsteren WF, Hunenberger PH (2001) A Fast Shake Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. J Comput Chem 22: 501-508.

(68) Humphrey W, Dalke A, Schulten K (1996) Vmd: Visual Molecular Dynamics. J Mol Graph 14: 33-38.

(69) Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) Mdanalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. J Comput Chem 32: 2319-2327.

(70) Pickett SD, Sternberg MJE (1993) Empirical Scale of Side-Chain Conformational Entropy in Protein-Folding. J Mol Biol 231: 825-839.

(71) Wedemeyer WJ, Welker E, Scheraga HA (2002) Proline Cis–Trans Isomerization and Protein Folding†. Biochemistry 41: 14637-14644.

(72) James JJ, Lakshmi BS, Seshasayee ASN, Gautam P (2007) Activation of Candida Rugosa Lipase at Alkane–Aqueous Interfaces: A Molecular Dynamics Study. FEBS Lett 581: 4377-4383.

(73) Rehm S, Trodler P, Pleiss J (2010) Solvent-Induced Lid Opening in Lipases: A Molecular Dynamics Study. Protein Sci 19: 2122-2130.

(74) da Costa Sousa L, Chundawat SPS, Balan V, Dale BE (2009) 'Cradle-to-Grave' Assessment of Existing Lignocellulose Pretreatment Technologies. Curr Opin Biotechnol 20: 339-347.

(75) Várnai A, Siika-aho M, Viikari L (2010) Restriction of the Enzymatic Hydrolysis of Steam-Pretreated Spruce by Lignin and Hemicellulose. Enzyme Microb Technol 46: 185-193.

(76) Nakagame S, Chandra RP, Saddler JN (2010) The Effect of Isolated Lignins, Obtained from a Range of Pretreated Lignocellulosic Substrates, on Enzymatic Hydrolysis. Biotechnol Bioeng 105: 871-879.

(77) Agbor VB, Cicek N, Sparling R, Berlin A, Levin DB (2011) Biomass Pretreatment: Fundamentals toward Application. Biotechnol Adv 29: 675-685.

(78) Mok WSL, Antal MJ, Varhegyi G (1992) Productive and Parasitic Pathways in Dilute Acid-Catalyzed Hydrolysis of Cellulose. Ind Eng Chem Res 31: 94-100.

(79) Ramos LP (2003) The Chemistry Involved in the Steam Treatment of Lignocellulosic Materials. Quim Nova 26: 863-871.

(80) Kilpeläinen I, Xie H, King A, Granstrom M, Heikkinen S, et al. (2007) Dissolution of Wood in Ionic Liquids. J Agric Food Chem 55: 9142-9148.

(81) Zhang H, Wu J, Zhang J, He J (2005) 1-Allyl-3-Methylimidazolium Chloride Room Temperature Ionic Liquid: A New and Powerful Nonderivatizing Solvent for Cellulose. Macromolecules 38: 8272-8277.

(82) Shill K, Padmanabhan S, Xin Q, Prausnitz JM, Clark DS, et al. (2011) Ionic Liquid Pretreatment of Cellulosic Biomass: Enzymatic Hydrolysis and Ionic Liquid Recycle. Biotechnol Bioeng 108: 511-520.

(83) Engel P, Mladenov R, Wulfhorst H, Jäger G, Spiess AC (2010) Point by Point Analysis: How Ionic Liquid Affects the Enzymatic Hydrolysis of Native and Modified Cellulose. Green Chem 12: 1959.

(84) Dadi AP, Schall CA, Varanasi S (2007) Mitigation of Cellulose Recalcitrance to Enzymatic Hydrolysis by Ionic Liquid Pretreatment. Appl Biochem Biotechnol 137: 407-421.

(85) Mansfield SD, Mooney C, Saddler JN (1999) Substrate and Enzyme Characteristics That Limit Cellulose Hydrolysis. Biotechnol Prog 15: 804-816.

(86) Wang Y, Radosevich M, Hayes D, Labbé N (2011) Compatible Ionic Liquid-Cellulases System for Hydrolysis of Lignocellulosic Biomass. Biotechnology and Bioengineering 108: 1042–1048.

(87) Vieille C, Zeikus GJ (2001) Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. Microbiol Mol Biol Rev 65: 1-43.

(88) Shao Q (2013) On the Influence of Hydrated Imidazolium-Based Ionic Liquid on Protein Structure Stability: A Molecular Dynamics Simulation Study. J Chem Phys 139: 115102.

(89) Micaelo NM, Soares CM (2008) Protein Structure and Dynamics in Ionic Liquids. Insights from Molecular Dynamics Simulation Studies. J Phys Chem B 112: 2566–2572.

(90) Kim HS, Ha SH, Sethaphong L, Koo Y-M, Yingling YG (2014) The Relationship between Enhanced Enzyme Activity and Structural Dynamics in Ionic Liquids: A Combined Computational and Experimental Study. Phys Chem Chem Phys 16: 2944-2953.

(91) Figueiredo AM, Sardinha J, Moore GR, Cabrita EJ (2013) Protein Destabilisation in Ionic Liquids: The Role of Preferential Interactions in Denaturation. Physical Chemistry Chemical Physics 15: 19632-19643.

(92) Haberler M, Schroder C, Steinhauser O (2011) Solvation Studies of a Zinc Finger Protein in Hydrated Ionic Liquids. Physical Chemistry Chemical Physics 13: 6924-6938.

(93) Burney PR, Pfaendtner J (2013) 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 117: 2662-2670.

(94) Li H, Kankaanpää A, Xiong H, Hummel M, Sixta H, et al. (2013) 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 Microb Technol 53: 414-419.

(95) Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) Gromacs 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4: 435–447.

(96) Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. Journal of Chemical Physics 79: 926-935.

(97) Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, et al. (2010) Improved Side-Chain Torsion Potentials for the Amber Ff99sb Protein Force Field. Proteins-Structure Function and Bioinformatics 78: 1950-1958.

(98) Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and Testing of a General Amber Force Field. J Comput Chem 25: 1157–1174.

(99) Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. Journal of Physical Chemistry 97: 10269–10280.

(100) Zhang Y, Maginn EJ (2012) A Simple Aimd Approach to Derive Atomic Charges for Condensed Phase Simulation of Ionic Liquids. J Phys Chem B 116: 10036-10048.

(101) Martinez L, Andrade R, Birgin EG, Martinez JM (2009) Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations. J Comput Chem 30: 2157-2164.

(102) Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) Lincs: A Linear Constraint Solver for Molecular Simulations. J Comput Chem 18: 1463-1472.

(103) Bussi G, Donadio D, Parrinello M (2007) Canonical Sampling through Velocity Rescaling. Journal of Chemical Physics 126.

(104) Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) Molecular-Dynamics with Coupling to an External Bath. J Chem Phys 81: 3684-3690.

(105) Trivedi N, Gupta V, Reddy CRK, Jha B (2013) Detection of Ionic Liquid Stable Cellulase Produced by the Marine Bacterium Pseudoalteromonas Sp. Isolated from Brown Alga Sargassum Polycystum C. Agardh. Bioresource Technology 132: 313-319.

(106) Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235-242.

(107) Lee TM, Farrow MF, Arnold FH, Mayo SL (2011) A Structural Study of Hypocrea Jecorina Cel5a. Protein Sci 20: 1935-1940.

(108) Li XH, Zhang P, Wang MX, Zhou F, Malik FA, et al. (2011) Expression of Trichoderma Viride Endoglucanase Iii in the Larvae of Silkworm, Bombyx Mori L. And Characteristic Analysis of the Recombinant Protein. Mol Biol Rep 38: 3897-3902.

(109) Wu TH, Huang CH, Ko TP, Lai HL, Ma YH, et al. (2011) Diverse Substrate Recognition Mechanism Revealed by Thermotoga Maritima Cel5a Structures in Complex with Cellotetraose, Cellobiose and Mannotriose. Biochim Biophys Acta 1814: 1832-1840.

(110) Kim HW, Ishikawa K (2011) Functional Analysis of Hyperthermophilic Endocellulase from Pyrococcus Horikoshii by Crystallographic Snapshots. Biochem J 437: 223-230.

(111) Olsson MHM, Sondergaard CR, Rostkowski M, Jensen JH (2011) Propka3: Consistent Treatment of Internal and Surface Residues in Empirical Pk(a) Predictions. J Chem Theory Comput 7: 525-537.

(112) Li H, Robertson AD, Jensen JH (2005) Very Fast Empirical Prediction and Rationalization of Protein Pk(a) Values. Proteins 61: 704-721.

(113) White AD, Nowinski AK, Huang W, Keefe AJ, Sun F, et al. (2012) Decoding Nonspecific Interactions from Nature. Chem Sci 3: 3488-3494.

(114) Voss NR, Gerstein M (2005) Calculation of Standard Atomic Volumes for Rna and Comparison with Proteins: Rna Is Packed More Tightly. J Mol Biol 346: 477-492.

(115) Kaar JL, Jesionowski AM, Berberich JA, Moulton R, Russell AJ (2003) Impact of Ionic Liquid Physical Properties on Lipase Activity and Stability. Journal of the American Chemical Society 125: 4125–4131.

(116) Erbeldinger M, Mesiano AJ, Russell AJ (2000) Enzymatic Catalysis of Formation of Z-Aspartame in Ionic Liquid - an Alternative to Enzymatic Catalysis in Organic Solvents. Biotechnol Prog 16: 1129-1131.

(117) Joosten RP, te Beek TAH, Krieger E, Hekkelman ML, Hooft RWW, et al. (2010) A Series of Pdb Related Databases for Everyday Needs. Nucleic Acids Res 39: D411-D419.

(118) Moniruzzaman M, Nakashima K, Kamiya N, Goto M (2010) Recent Advances of Enzymatic Reactions in Ionic Liquids. Biochem Eng J 48: 295-314.

(119) Kumar A, Venkatesu P (2014) Does the Stability of Proteins in Ionic Liquids Obey the Hofmeister Series? Int J Biol Macromol 63: 244-253.

(120) Lai J-Q, Li Z, Lü Y-H, Yang Z (2011) Specific Ion Effects of Ionic Liquids on Enzyme Activity and Stability. Green Chem 13: 1860.

(121) Baldwin RL (1996) How Hofmeister Ion Interactions Affect Protein Stability. Biophys J 71: 2056-2063.

(122) Boström M, Tavares FW, Finet S, Skouri-Panet F, Tardieu A, et al. (2005) Why Forces between Proteins Follow Different Hofmeister Series for Ph above and Below Pi. Biophys Chem 117: 217-224.

(123) Nordwald EM, Brunecky R, Himmel ME, Beckham GT, Kaar JL (2014) Charge Engineering of Cellulases Improves Ionic Liquid Tolerance and Reduces Lignin Inhibition. Biotechnol Bioeng.

(124) Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) Gromacs 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4: 435-447.

(125) Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, et al. (2010) Improved Side-Chain Torsion Potentials for the Amber Ff99sb Protein Force Field. Proteins 78: 1950-1958.

(126) Bussi G, Donadio D, Parrinello M (2007) Canonical Sampling through Velocity Rescaling. J Chem Phys 126: 014101.

(127) Kumar S, Nussinov R (2002) Close-Range Electrostatic Interactions in Proteins. ChemBioChem 3: 604-617.

(128) Dima RI, Thirumalai D (2004) Asymmetry in the Shapes of Folded and Denatured States of Proteins†. J Phys Chem B 108: 6564-6570.

(129) Carr ME, Jr., Carr SL (1995) Fibrin Structure and Concentration Alter Clot Elastic Modulus but Do Not Alter Platelet Mediated Force Development. Blood Coagul Fibrinolysis 6: 79-86.

(130) Ma J, Hennekens CH, Ridker PM, Stampfer MJ (1999) A Prospective Study of Fibrinogen and Risk of Myocardial Infarction in the Physicians' Health Study. J Am Coll Cardiol 33: 1347-1352.

(131) Le DS, Miles R, Savage PJ, Cornell E, Tracy RP, et al. (2008) The Association of Plasma Fibrinogen Concentration with Diabetic Microvascular Complications in Young Adults with Early-Onset of Type 2 Diabetes. Diabetes Res Clin Pract 82: 317-323.

(132) Jorneskog G, Egberg N, Fagrell B, Fatah K, Hessel B, et al. (1996) Altered Properties of the Fibrin Gel Structure in Patients with Iddm. Diabetologia 39: 1519-1523.

(133) Shacter E, Williams JA, Lim M, Levine RL (1994) Differential Susceptibility of Plasma Proteins to Oxidative Modification: Examination by Western Blot Immunoassay. Free Radical Biol Med 17: 429-437.

(134) Suontaka AM, Blomback M, Chapman J (2003) Changes in Functional Activities of Plasma Fibrinogen after Treatment with Methylene Blue and Red Light. Transfusion 43: 568-575.

(135) Carr ME, Jr., Powers PL (1989) Methylene Blue Enhances Lateral Association of Fibrin Resulting in Rapid Gelation and Thick Fiber Formation. Thromb Haemostasis 61: 230-233.

(136) Parastatidis I, Thomson L, Burke A, Chernysh I, Nagaswami C, et al. (2008) Fibrinogen Beta-Chain Tyrosine Nitration Is a Prothrombotic Risk Factor. J Biol Chem 283: 33846-33853.

(137) Nowak P, Zbikowska HM, Ponczek M, Kolodziejczyk J, Wachowicz B (2007) Different Vulnerability of Fibrinogen Subunits to Oxidative/Nitrative Modifications Induced by Peroxynitrite: Functional Consequences. Thromb Res 121: 163-174.

(138) Selmeci L, Szekely M, Soos P, Seres L, Klinga N, et al. (2006) Human Blood Plasma Advanced Oxidation Protein Products (Aopp) Correlates with Fibrinogen Levels. Free Radical Res 40: 952-958.

(139) Descamps-Latscha B, Witko-Sarsat V, Nguyen-Khoa T, Nguyen AT, Gausson V, et al. (2005) Advanced Oxidation Protein Products as Risk Factors for Atherosclerotic Cardiovascular Events in Nondiabetic Predialysis Patients. Am J Kidney Dis 45: 39-47.

(140) Vadseth C, Souza JM, Thomson L, Seagraves A, Nagaswami C, et al. (2004) Pro-Thrombotic State Induced by Post-Translational Modification of Fibrinogen by Reactive Nitrogen Species. J Biol Chem 279: 8820-8826.

(141) Weigandt Katie M, White N, Chung D, Ellingson E, Wang Y, et al. (2012) Fibrin Clot Structure and Mechanics Associated with Specific Oxidation Of methionine Residues in Fibrinogen. Biophys J 103: 2399-2407.

(142) Collet JP (2005) The αc Domains of Fibrinogen Affect the Structure of the Fibrin Clot, Its Physical Properties, and Its Susceptibility to Fibrinolysis. Blood 106: 3824-3830.

(143) Weisel JW, Litvinov RI (2013) Mechanisms of Fibrin Polymerization and Clinical Implications. Blood 121: 1712-1719.

(144) Mosesson MW, DiOrio JP, Muller MF, Shainoff JR, Siebenlist KR, et al. (1987) Studies on the Ultrastructure of Fibrin Lacking Fibrinopeptide B (Beta-Fibrin). Blood 69: 1073-1081.

(145) Litvinov RI, Gorkun OV, Galanakis DK, Yakovlev S, Medved L, et al. (2007) Polymerization of Fibrin: Direct Observation and Quantification of Individual B:B Knob-Hole Interactions. Blood 109: 130-138.

(146) Doolittle RF, Pandi L (2007) Probing the B-Chain Hole of Fibrinogen with Synthetic Peptides That Differ at Their Amino Termini. Biochemistry 46: 10033-10038.

(147) Kollman JM, Pandi L, Sawaya MR, Riley M, Doolittle RF (2009) Crystal Structure of Human Fibrinogen. Biochemistry 48: 3877-3886.

(148) Medved L, Weisel JW (2009) Recommendations for Nomenclature on Fibrinogen and Fibrin. J Thromb Haemost 7: 355-359.

(149) Burton RA, Tsurupa G, Hantgan RR, Tjandra N, Medved L (2007) Nmr Solution Structure, Stability, and Interaction of the Recombinant Bovine Fibrinogen Ac-Domain Fragment. Biochemistry 46: 8550-8560.

(150) Tsurupa G, Hantgan RR, Burton RA, Pechik I, Tjandra N, et al. (2009) Structure, Stability, and Interaction of the Fibrin(Ogen) Ac-Domains. Biochemistry 48: 12191-12201.

(151) Watt KWK, Cottrell BA, Strong DD, Doolittle RF (1979) Amino Acid Sequence Studies on the A Chain of Human Fibrinogen. Overlapping Sequences Providing the Complete Sequence. Biochemistry 18: 5410-5416.

(152) Šali A, Blundell TL (1993) Comparative Protein Modelling by Satisfaction of Spatial Restraints. J Mol Biol 234: 779-815.

(153) Simons KT, Bonneau R, Ruczinski I, Baker D (1999) Ab Initio Protein Structure Prediction of Casp Iii Targets Using Rosetta. Proteins Suppl 3: 171-176.

(154) Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular Dynamics with Coupling to an External Bath. J Chem Phys 81: 3684.

(155) Bonomi M, Parrinello M (2010) Enhanced Sampling in the Well-Tempered Ensemble. Phys Rev Lett 104: 190601.

(156) Deighan M, Bonomi M, Pfaendtner J (2012) Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble. J Chem Theory Comput 8: 2189-2192.

(157) Deighan M, Pfaendtner J (2013) Exhaustively Sampling Peptide Adsorption with Metadynamics. Langmuir 29: 7999-8009.

(158) Laio A, Parrinello M (2002) Escaping Free-Energy Minima. Proc Natl Acad Sci U S A 99: 12562-12566.

(159) Barducci A, Bussi G, Parrinello M (2008) Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. Phys Rev Lett 100: 020603.

(160) Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of State Calculations by Fast Computing Machines. J Chem Phys 21: 1087.

(161) Bussi G, Gervasio FL, Laio A, Parrinello M (2006) Free-Energy Landscape for B Hairpin Folding from Combined Parallel Tempering and Metadynamics. J Am Chem Soc 128: 13435-13441.

(162) Sutto L, Gervasio FL (2013) Effects of Oncogenic Mutations on the Conformational Free-Energy Landscape of Egfr Kinase. Proc Natl Acad Sci U S A 110: 10616-10621.

(163) Prakash MK, Barducci A, Parrinello M (2011) Replica Temperatures for Uniform Exchange and Efficient Roundtrip Times in Explicit Solvent Parallel Tempering Simulations. J Chem Theory Comput 7: 2025-2027.

(164) Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, et al. (2009) Plumed: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. Comput Phys Commun 180: 1961-1972.

(165) Trzesniak D, Kunz A-PE, van Gunsteren WF (2007) A Comparison of Methods to Compute the Potential of Mean Force. ChemPhysChem 8: 162-169.

(166) Jarvis RA, Patrick EA (1973) Clustering Using a Similarity Measure Based on Shared near Neighbors. IEEE Trans Comput C-22: 1025-1034.

(167) Devore JL (1995) Probability and Statistics for Engineering and the Sciences. Belmont: Duxbury Press.

(168) Tsurupa G, Pechik I, Litvinov RI, Hantgan RR, Tjandra N, et al. (2012) On the Mechanism of Ac Polymer Formation in Fibrin. Biochemistry 51: 2526-2538.

(169) Tsurupa G, Mahid A, Veklich Y, Weisel JW, Medved L (2011) Structure, Stability, and Interaction of Fibrin Ac-Domain Polymers. Biochemistry 50: 8028-8037.

(170) Ping L, Huang L, Cardinali B, Profumo A, Gorkun OV, et al. (2011) Substitution of the Human Ac Region with the Analogous Chicken Domain Generates a Fibrinogen with Severely Impaired Lateral Aggregation: Fibrin Monomers Assemble into Protofibrils but Protofibrils Do Not Assemble into Fibers. Biochemistry 50: 9066-9075.

(171) Rosen H, Klebanoff SJ, Wang Y, Brot N, Heinecke JW, et al. (2009) Methionine Oxidation Contributes to Bacterial Killing by the Myeloperoxidase System of Neutrophils. Proc Natl Acad Sci U S A 106: 18686-18691.

(172) Heffron SP, Parastatidis I, Cuchel M, Wolfe ML, Tadesse MG, et al. (2009) Inflammation Induces Fibrinogen Nitration in Experimental Human Endotoxemia. Free Radical Biol Med 47: 1140-1146.

(173) Nalian A, Iakhiaev AV (2008) Possible Mechanisms Contributing to Oxidative Inactivation of Activated Protein C: Molecular Dynamics Study. Thromb Haemostasis 100: 18-25.

(174) Glaser CB, Morser J, Clarke JH, Blasko E, McLean K, et al. (1992) Oxidation of a Specific Methionine in Thrombomodulin by Activated Neutrophil Products Blocks Cofactor Activity. A Potential Rapid Mechanism for Modulation of Coagulation. J Clin Invest 90: 2565-2573.

(175) Shaw DE, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, et al. (2007) Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. SIGARCH Comput Archit News 35: 1-12.

(176) Turrens JF (2003) Mitochondrial Formation of Reactive Oxygen Species. J Physiol 552: 335-344.

(177) Frey RS, Ushio–Fukai M, Malik AB (2009) Nadph Oxidase-Dependent Signaling in Endothelial Cells: Role in Physiology and Pathophysiology. Antioxid Redox Signaling 11: 791-810.

(178) Brohi K, Singh J, Heron M, Coats T (2003) Acute Traumatic Coagulopathy. J Trauma 54: 1127-1130.

(179) MacLeod JBA, Lynn M, McKenney MG, Cohn SM, Murtha M (2003) Early Coagulopathy Predicts Mortality in Trauma. J Trauma 55: 39-44.

(180) Mitra B, Cameron PA, Mori A, Fitzgerald M (2012) Acute Coagulopathy and Early Deaths Post Major Trauma. Injury 43: 22-25.

(181) Frith D, Goslings JC, Gaarder C, Maegele M, Cohen MJ, et al. (2010) Definition and Drivers of Acute Traumatic Coagulopathy: Clinical and Experimental Investigations. J Thromb Haemost 8: 1919-1925.

(182) Brohi K, Cohen MJ, Davenport RA (2007) Acute Coagulopathy of Trauma: Mechanism, Identification and Effect. Curr Opin Crit Care 13: 680-685.

(183) Rourke C, Curry N, Khan S, Taylor R, Raza I, et al. (2012) Fibrinogen Levels During Trauma Hemorrhage, Response to Replacement Therapy, and Association with Patient Outcomes. J Thromb Haemost 10: 1342-1351.

(184) Kashuk JL, Moore EE, Sawyer M, Wohlauer M, Pezold M, et al. (2010) Primary Fibrinolysis Is Integral in the Pathogenesis of the Acute Coagulopathy of Trauma. Ann Surg 252: 434-444.

(185) Limongelli V, Bonomi M, Parrinello M (2013) Funnel Metadynamics as Accurate Binding Free-Energy Method. Proceedings of the National Academy of Sciences 110: 6358-6363.

(186) (2005) Nist Chemistry Webbook, Nist Standard Reference Database Number 69; Linstrom PJ, Mallard WG, editors: National Institute of Standards and Technology.

(187) Mokhtarani B, Sharifi A, Mortaheb HR, Mirzaei M, Mafi M, et al. (2009) Density and Viscosity of 1-Butyl-3-Methylimidazolium Nitrate with Ethanol, 1-Propanol, or 1-Butanol at Several Temperatures. J Chem Thermodyn 41: 1432-1438.

(188) Jacquemin J, Husson P, Padua AAH, Majer V (2006) Density and Viscosity of Several Pure and Water-Saturated Ionic Liquids. Green Chem 8: 172-180.

(189) Strechan AA, Kabo AG, Paulechka YU, Blokhin AV, Kabo GJ, et al. (2008) Thermochemical Properties of 1-Butyl-3-Methylimidazolium Nitrate. Thermochim Acta 474: 25-31.

(190) Nieto de Castro CA, Lourenço MJV, Ribeiro APC, Langa E, Vieira SIC, et al. (2010) Thermal Properties of Ionic Liquids and Ionanofluids of Imidazolium and Pyrrolidinium Liquids. J Chem Eng Data 55: 653-661.

VITA

Patrick was born in Spokane, WA and spent his formative years in Arlington, TX and Tampa, FL. Patrick attended the University of Florida in Gainesville, FL under the Florida Bright Futures scholarship program. Patrick received his Bachelor of Science in Chemical Engineering in 2009 from the University of Florida then proceeded to the University of Washington, becoming the first student to join Professor Jim Pfaendtner's research group.

Master of Science, Chemical Engineering (2013)
University of Washington, Seattle, WA

Bachelor of Science, Chemical Engineering (2009) *Cum Laude*
University of Florida, Gainesville, FL

Honors:
Thomas G. Thompson Endowed Fellowship Recipient, 2009-2010
UW College of Engineering: Top Scholar Fellowship Recipient, 2009-2010

Publications

1. **Burney PR**, Nordwald EM, Hickman K, Kaar JL, Pfaendtner J (2014) Molecular Dynamics Simulations of the Role of Enzyme Surface Charge on Mediating Ion-Specific Interations in Ionic Liquids**.** *In Prep.*
2. Jaeger VW, **Burney PR**, Pfaendtner J (2014) Comparison of Three Ionic Liquid Tolerant Cellulases by Molecular Dynamic. *Submitted.*
3. **Burney PR**, White N,Pfaendtner J (2014) Structural Effects of Methionine Oxidation on Isolated Subdomains of Human Fibrin D and αC Regions. PLOS One 9: e86981
4. **Burney PR**, Pfaendtner J (2013) Structural and Dynamic Features of Candida Rugosa Lipase 1 in Water, Octane, Toluene, and Ionic Liquids BMIM-PF6 and BMIM-NO3. Journal of Physical Chemistry B 117: 2662-2670.
5. Wu P, Arce SH, **Burney PR**, Tseng Y (2009) A Novel Approach to High Accuracy of Video-Based Microrheology. Biophysical Journal 96:5103-5111

Contributed Presentations

1. Burney, PR, White N, Pfaendtner J (2014) Structural Effects of Methionine Oxidation on Isolated Subdomains of Human Fibrin. Biophysical Society Annual Meeting. San Francisco, CA
2. Burney, PR, Pfaendtner J (2012) Candida Rugosa Lipase Simulations in Water & Non-aqueous Solvents. AIChE Annual Meeting. Pittsburgh, PA
3. Burney PR, Pfaendtner J (2012) Simulation Reveals Mechanistic Details of Lipase Lid Closing. UW Chemical Engineering Graduate Symposium. Seattle, WA
4. Burney PR, Pfaendtner J (2011) MD and Coarse Grained Simulation Reveals Mechanistic Details of Lipase Lid Closing. AICHE Annual Meeting. Minneapolis, MN

Appendix 1 – Chapter 2 Supporting Material

Testing Non-aqueous Solvent Force Fields

We performed characterization of pure ILs and testing of the force fields before coupling them with the enzyme. For the two chosen ILs, there are few experimental observables known for these ILs when they are near pure and which are easily calculable using MD simulations. The two observables we found available and simple to calculate for our ILs were density and heat capacity. Using 33x33x33 Å boxes of each of our pure non-aqueous solvents, we performed 10 duplicate 10ns NPT simulations at 330 K and 1 bar, followed by 10ns NVT simulations at 330K from which we calculated the density and heat capacity. Of the several force fields tested, we found Amber's GAFF performed admirably when combined with un-scaled RESP partial charges, as obtained from Gaussian 09 calculations. However, the hexafluorophosphate anion from BMIM-PF_6 could not be parameterized using GAFF in Amber10. After testing several force fields, we found that treating PF6 as rigid at the bond lengths and angles we obtained from a Gaussian 09 geometry optimization combined with the RESP partial charges reproduced the experimental observables. We have tabulated (Table A1.1) the average density and heat capacity from these 10 replicate simulations for each solvent along with literature reported experimental values at 330K. Overall, GAFF suited the organic solvents the best and the ILs were < 5% and < 7% different from literature values of density and heat capacity, respectively.

Table A1.1 Comparison of literature values for both the density and heat capacity of these pure liquid systems with those calculated using MD simulation with the force fields employed in our production simulations. Simulated values of density displayed here are an average value over 10 separate 10ns NPT simulations. Similarly, heat capacity values are average values calculated from the last 1 ns of 10 separate 10ns NVT simulations, which immediately followed the 10ns NPT simulations.

Solvent	Density (Simulation) [g/cc]	Density (Experiment)[a] [g/cc]	Cv (Simulation) [J/mol K]	Cv (Experiment)[b] [J/mol K]
Octane	0.65	0.67	276	268
Toluene	0.79	0.83	168	167
BMIM-NO_3	1.15	1.14	387	366
BMIM-PF_6	1.36	1.34	397	374

[a] Experimental data for densities were obtained from previously published studies for octane [186], toluene [186], BMIM-NO_3 [187], and BMIM-PF_6 [188].
[b] Experimental data for heat capacities were obtained from previously published studies for octane [186], toluene [186], BMIM-NO_3 [189], and BMIM-PF_6 [190].

MD Simulation Setup

Preparation of the CRL in water systems was carried out in the Amber10 leap environment. This included surrounding the protein with a pre-equilibrated box of TIP3P water. To emulate this basic function of leap for our non-aqueous systems, we generated randomized boxes of pure solvent molecules using Packmol [101] and ran several MD simulations in NAMD to equilibrate these boxes before inserting the enzyme. The initial box size as input to Packmol was 109x109x109 Å and the number of atoms was chosen as 10% lower than the reported experimental density at 310K. Sequentially, we used 5000 steps of conjugate gradient minimization, 4 ns of NVT simulation at 500 K, 0.5 ns of incrementally lowering the temperature to 310K, and a final 10 ns NPT simulation at 1 atm and 350 K. The protein, using the coordinates obtained for the corresponding crystal structure, was then placed in the center of the final frame from the NPT equilibration and solvent molecules were deleted within 3.5 Å of any protein atom. Sodium counter ions from the water simulations were also introduced into the non-aqueous systems to neutralize the protein. This also required deletion of solvent molecules. In the ionic liquid systems these deletions always led to a charge imbalance, so we also randomly deleted excess

ions. For the 10 systems simulated, we have tabulated here (Table A1.2) details of the initial and final box dimensions as well as the number of atoms.

Table A1.2 Details on the scale of our production NPT simulations. The initial box volume reported here corresponds to the volume of the system used as the initial coordinates of the production NPT simulations. The equilibrated box volumes correspond to an average volume over each frame from the 98th through 100th ns of production NPT simulations.

Solvent	Initial Structure	Num. of Atoms	Initial Box Volume [Å^3]	Equilibrated Box Volume (310K / 375K)
Water	CRLO	77160	830079	907089 / 838384
Water	CRLC	73278	794477	741329 / 795520
Octane	CRLO	86530	1272568	927418 / 1025740
Octane	CRLC	88167	1272568	947473 / 1046292
Toluene	CRLO	58521	913908	695698 / 770895
Toluene	CRLC	60413	913908	719283 / 797344
BMIM-NO3	CRLO	79474	1102397	778513 / 805216
BMIM-NO3	CRLC	79503	1102397	779158 / 806137
BMIM-PF6	CRLO	63826	959284	664286 / 684382
BMIM-PF6	CRLC	65207	959284	678058 / 699886

Closed Structure Comparison

During the simulations in water at both 310K and 375K starting from 1CRL [51], the open crystal structure (CRLO), we observed protein motions encompassing a closing of the lid onto the protein bulk. Plotting the RMSD of these trajectories compared to the 1TRH [44], the closed crystal structure (CRLC), reveals that these new stable states are significantly different from the CRLC (c.f., Figure 2.7B in Section 2.3.4). Analysis of the trajectories from these two simulations reveals that neither of these states exhibits proper packing of hydrophobic residues into the active site, the correct cis configuration of proline 92, or complete coverage of the active site. Shown here are surface representations of the protein for the final structures from both the water simulations and the CRLC structure. These renderings were made using VMD [68]. While these structures do show moderate coverage of the active site, analysis of the trajectory shows this coverage is somewhat transient and not as complete as seen in the CRLC structure.

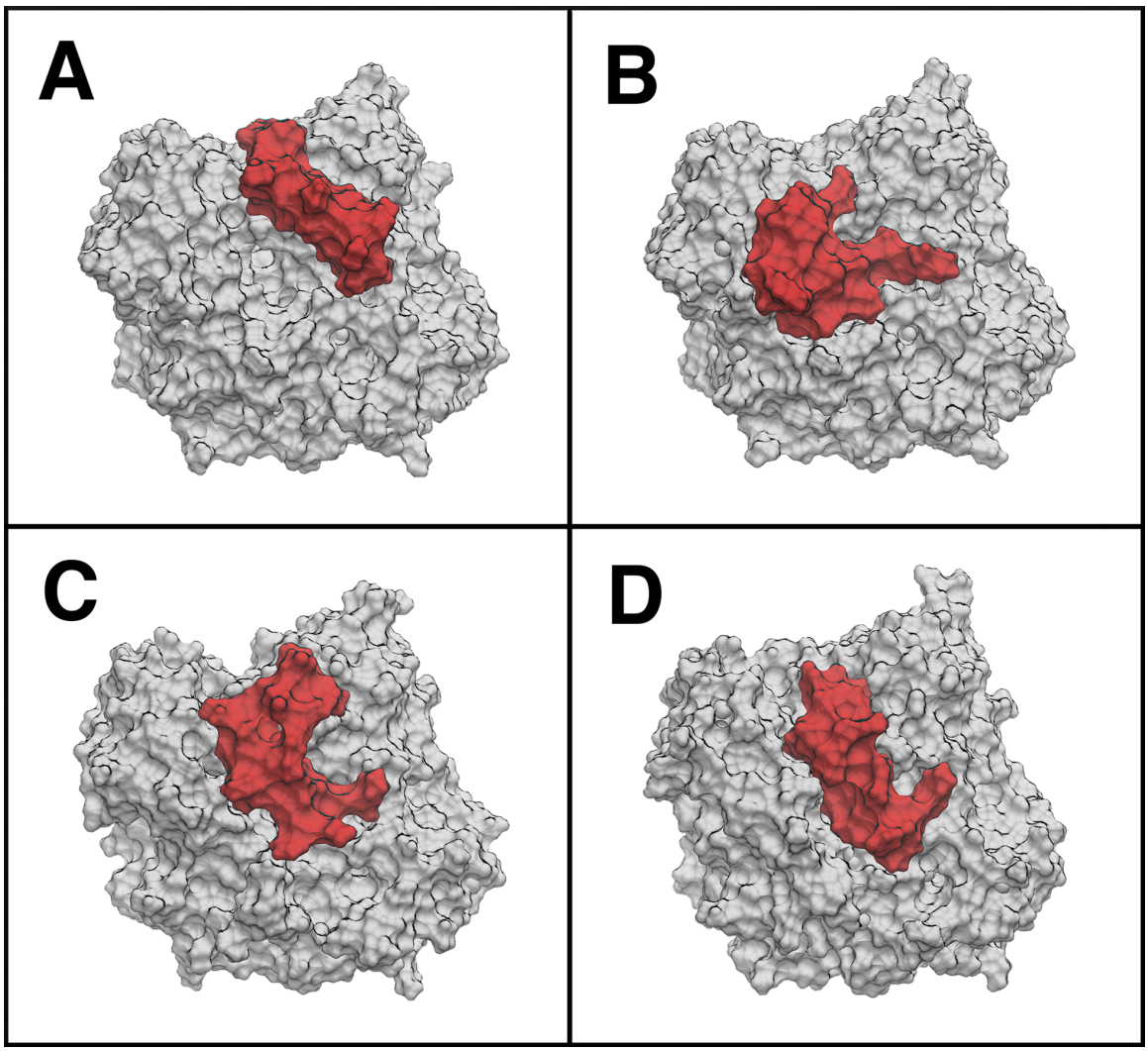


Figure A1.1 Renderings for (A) CRLO, (B) CRLC, (C) the final structure from the 310K simulation and (D) the final structure from the 375K simulation. The bulk of the protein is shown in grey while lid residues, here displayed as residues 60-95, are shown in red.

Appendix 2 – Chapter 3 Supporting Material

This work included analysis of 54 molecular dynamics simulations. The simulated systems are composed of: the enzymes *Trichoderma viride* (VIR), *Thermogata 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 is simulated in triplicate to increase statistics and help to verify observations. 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 (Chapter 3). Bezier curve fitting is also used to help distinguish each replicate trajectory in the figures below.

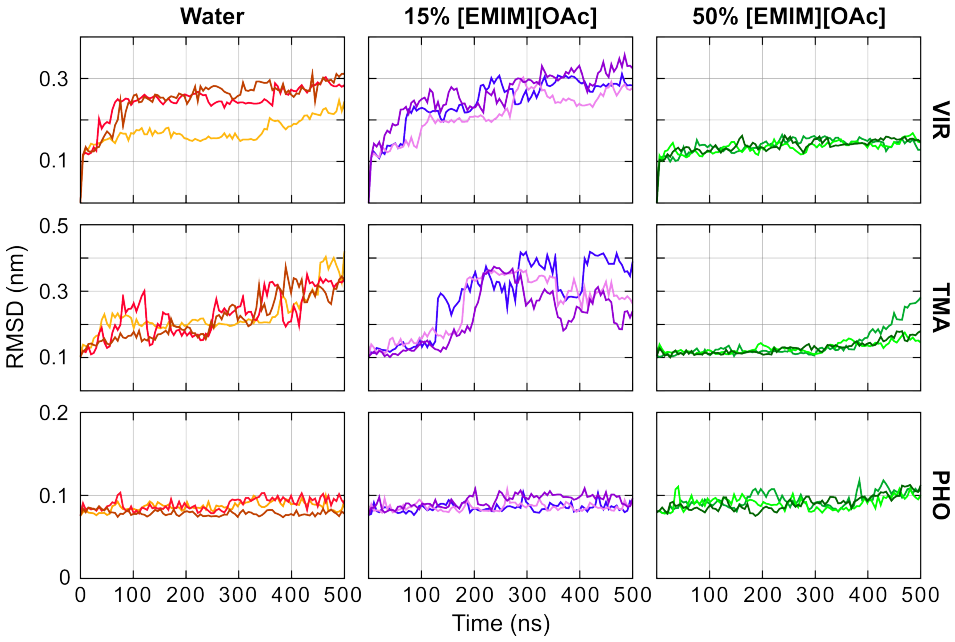


Figure A2.1: Bezier curve fitting of the Cα RMSD for each of the simulated systems at 310 K. The RMSD was calculated from (after alignment to) the corresponding experimental structure for each enzyme. Note the change in y-axis scale between each row.

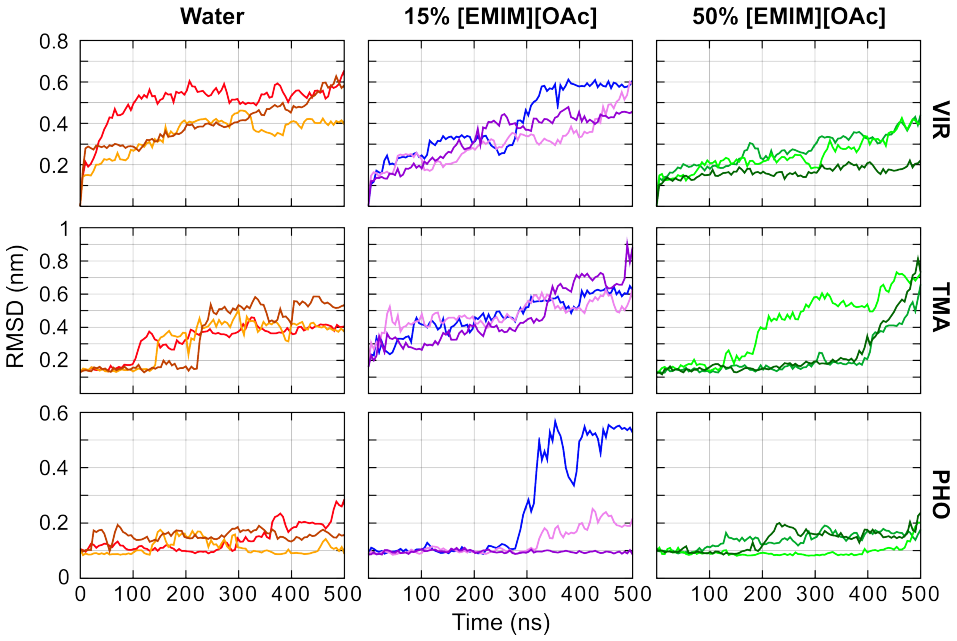


Figure A2.2: Bezier curve fitting of the Cα RMSD for each of the simulated systems at 353 K. The RMSD was calculated from (after alignment to) the corresponding experimental structure for each enzyme. Note the change in y-axis scale between each row.

RMSD was calculated using the g_rms tool in Gromacs 4.6 [124]. The reference structure for VIR was obtained by modifying GLU53 to ASP in the *Trichoderma reesei* (PDB: 3QR3) [107] x-ray structure in order to match the known sequence for T. viride endoglucanase [108]. The reference structures for TMA and PHO were modeled using x-ray structures (PDB: 3AMD, 3AXX) [109,110].