Development and Use of Molecular Simulation Tools to Study the Structure and Function of Biomolecules at Interfaces

Kayla G. Sprenger

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

Doctor of Philosophy

University of Washington

2017

Reading Committee:
Jim Pfaendtner, Chair
Elizabeth Nance
Shaoyi Jiang

Program Authorized to Offer Degree:
Chemical Engineering

University of Washington

Abstract

Development and Use of Molecular Simulation Tools to Study the Structure and Function of Biomolecules at Interfaces

Kayla G. Sprenger

Chair of the Supervisory Committee: Associate Professor Jim Pfaendtner Department of Chemical Engineering

Predicting and controlling the biophysical chemistry of protein/host interactions remains a pressing challenge of high fundamental interest across many subfields in chemistry, engineering, and medicine. Our ability to (experimentally) probe interfacial interactions between a surface, polymer, or solvent and a biomolecule has increased due to high-resolution NMR and other spectroscopic techniques. However, new structural and mechanistic insights coming from experiments alone have been limited due to – in the case of the protein/liquid interface – the compounding challenges of simultaneously studying the protein/host interface and elucidating sequence specific interactions, and in the case of the protein/surface interface – the lack of a single unifying technique that can fully resolve the structure of an adsorbed protein. For the protein/liquid interface, the challenge is even worse in the case of synthetic frameworks such as ionic liquids (ILs) where the combinatorial design space of the solvent and protein sequence explodes beyond what could ever be feasibly considered in a laboratory. The field could be greatly advanced through the use of predictive, physics-based simulations; however, best practices of how to simulate protein interfacial adsorption and extract meaningful information about protein/host interactions from molecular simulations for direct experimental comparison, are still developing. Herein, we aim to address these challenges through the use of a variety of statistical enhanced sampling simulation techniques and novel analytical approaches. A number of systems are simulated, ranging from small model proteins adsorbing onto solid, idealized surfaces, to simulations of full proteins in complex solution environments. Ultimately, these simulations should lead to large improvements in the way we use computers to study multifaceted interfacial processes like protein adsorption, in addition to providing new fundamental insights into protein/host interactions.

Table of Contents

List of Figures 7
List of Tables 18
Acknowledgements 20

Introduction **21**

1 Strong Electrostatic Interactions Lead to Entropically Favorable Binding of Peptides to Charged Surfaces **24**
1.1 Introduction 24
1.2 Methods 26
1.2.1 System Setup 26
1.2.2 Enhanced Sampling 28
1.3 Results and Discussion 29
1.3.1 Convergence of MetaD Simulations 29
1.4 Conclusions 41

2 Essential Slow Degrees of Freedom in Protein-Surface Simulations: A Metadynamics Investigation **43**
2.1 Introduction 43
2.2 Methods 44
2.2.1 Simulation Setup 44
2.2.2 Biasing Methods 45
2.2.3 Well-Tempered Metadynamics (WTM) 45
2.2.4 Parallel Tempering Metadynamics in the Well-Tempered Ensemble (PTMetaD-WTE) 46
2.2.5 Parallel Bias Metadynamics (PBMetaD) 47
2.3 Results and Discussion 48
2.3.1 Choice of Collective Variable 48
2.3.2 Comparison of Sampling Schemes 49
2.3.3 Ion-Surface Distance as a Collective Variable 51
Protein Adsorption Without Excess Ions *51*
Effect of Excess $Na^{0.5+}$ Ions on Peptide Adsorption *52*
Effect of Excess Ca^{2+} Ions on Peptide Adsorption *53*
Effect of Excess Na^{+} Ions on Peptide Adsorption *54*
2.4 Conclusions 55

3 Probing How Defects in Self-Assembled Monolayers Affect Peptide Adsorption with Molecular Simulation **57**
3.1 Introduction 57
3.2 Methods 58
3.2.1 System Setup 58
3.3 Results and Discussion 62
3.3.1 Convergence of MetaD Simulations 62
3.3.2 Clustering of Surface-Bound Structures 63
3.4 Conclusions 67

4 Using Molecular Simulation to Study Biocatalysis in Ionic Liquids 68
4.1 Introduction 68
4.1.1 Motivation for Studying Biomolecules in Ionic Liquids 68
4.1.2 Experiments of Biocatalysis in Ionic Liquids 69
4.1.3 Simulations of Biocatalysis in Ionic Liquids 69
4.2 Methods for Simulating Biomolecules in Ionic Liquids 72
4.2.1 Choice of Force Field/Parameterization Process 72
4.2.2 System Setup Example Using GAFF/Antechamber Tools 74
4.2.3 Electrostatics and Charge Scaling Considerations 75
4.2.4 IL-Specific Aspects of Enzyme MD Simulations 78
4.2.5 Typical Analysis Approaches 78
4.3 Perspective: Challenges and Future Directions 80

5 The General AMBER Force Field (GAFF) Can Accurately Predict Thermodynamic And Transport Properties Of Many Ionic Liquids 82
5.1 Introduction 82
5.2 Methods 85
5.2.1 GAFF 85
Quantum Mechanics Calculations *85*
Ionic Liquid Simulations *85*
5.3 Results and Discussion 86
5.3.1 Density 87
5.3.2 Heat Capacity 89
5.3.3 Molar Enthalpy of Vaporization 91
5.3.4 Self-Diffusivity 93
5.3.5 Shear Viscosity 97
5.3.6 Opportunities for Improvement 100
5.4 Conclusions 100

6 The Lytic Polysaccharide Monooxygenases *Sc*LPMO10B and *Sc*LPMO10C Are Stable in Ionic Liquids as Determined by Molecular Simulations 101
6.1 Introduction 102
6.2 Methods 104
6.2.1 Enzyme/IL Simulations 104
6.2.2 Parameterization of the Cu-Histidine Brace 105
6.3 Results and Discussion 105
6.3.1 IL Effects on Enzyme Structure 106
6.3.2 IL Effects on Dynamical Fluctuations 113
6.3.3 IL Effects on Enzyme Active Sites 115
6.4 Conclusions 117

7 Elucidating Sequence and Solvent Specific Design Targets to Protect and Stabilize Enzymes for Biocatalysis in Ionic Liquids 119
7.1 Introduction 119
7.2 Computational Methods 120
7.2.1 Binding Site Determination 120
7.2.2. Simulation Protocol 121
7.3 Results and Discussion 122
7.3.1. Comparing the Locations of Cation Binding to the Folded State from MD with Experiments 122
7.3.2. Comparing the Locations of Anion Binding to the Folded State from MD with Experiments 125
7.3.3. Analysis of IL Binding to the Unfolded State 126
7.3.4. The Counter-Ion Effect and its Influence on Enzyme Stabilization in Aqueous IL Solutions 128
7.4 Conclusions 130

8 Effect of an Ionic Liquid/Air Interface on the Structure and Dynamics of Amphiphilic Peptides 131
8.1 Introduction 131
8.2 Methods 133
8.3 Results and Discussion 136
8.3.1 IL Effects on LK Peptide Affinity for the Vacuum/Solvent Interface 136
8.3.2 Peptide Secondary Structure - LKα14 142
8.3.3 Peptide Secondary Structure - LKβ15 143
8.4 Conclusions 147

9 Determining Dominant Driving Forces Affecting Controlled Protein Release from Polymeric Nanoparticles 149
9.1 Introduction 149
9.2 Experimental Methods 151
9.2.1 Materials 151
9.2.2 Nanoparticle Formulation 151
9.2.3 Nanoparticle Characterization 151
9.2.4 BSA Release 152
9.3 Computational Methods 152
9.4 Results and Discussion 153
9.4.1 Experimental Determination of Protein Release Rates from Polymer Nanoparticles 153
9.4.2 Aggregation Tendencies of Polymer Surrogates 155
9.4.3 Amino Acid Binding Preferences of Polymer Surrogates 156
9.4.4 Spatiotemporal Resolution of Polymer Surrogate Binding to Protein Surface Residues 158
9.4.5 Tuning Polymer Self-Interactions 159
9.4.6 Effect of Surrogate Length on Binding Convergence and Characteristics 161
9.5 Summary and Conclusions 163

Coda 164
Impact 164
Future Work 164

Appendix I 166
A Slight Tweak to the PTMetaD-WTE Protocol 166
Evaluating the Temperature Dependence of the Potential Energy 167

Appendix II 171

Appendix III 173
Additional Metrics for Binding Site Clarification 174

Appendix IV 180

Appendix V 181

Bibliography 186

List of Figures

Figure 1.1. Convergence of free energy differences between solvated and adsorbed states for PTMetaD-WTE simulations at 300 K. Lines I-X correspond to systems I-X in Table 1.1. Lines VIIa/VIIb and VIIIa/VIIIb correspond to systems VII and VIII in Table 1.1, respectively, where a single ion/water molecule was biased. 30

Figure 1.2. Binding free energy change upon adsorption to SAM surfaces as a function of simulation temperature for systems I, III, IV, VI, IX, and X. Each circle represents a different replica in the system. 31

Figure 1.3. Proposed breakdown of the thermodynamic cycle of the peptide/surface adsorption process. Steps A, B, and C represent the overall thermodynamic cycle, peptide adsorption step, and ion desorption step, respectively. 33

Figure 1.4. Binding free energies calculated for systems I, II, and VIIa/b, corresponding to steps B, A, and C, respectively, of the thermodynamic cycle of adsorption (c.f. Figure 1.3). Binding free energies have been multiplied by -1 for system VIIa to represent ion desorption from the surface. 33

Figure 1.5. Reweighted number of ions displaced from the COOH/COO SAM during LKα14 adsorption as a function of temperature. Lines I and II refer to the peptide adsorption step (step B) and the overall thermodynamic cycle (step A), respectively. Dashed lines refer to the average number of desorbed ions used in each case for later analysis. 35

Figure 1.6. Calculated binding free energies for the three steps of the thermodynamic cycle for LKα14 on COOH/COO. Yellow and green lines represent steps B (system I, corrected for ion desorption) and 3.5·C (system VIIa) of the mechanism proposed in Figure 1.3, respectively, whereas purple and blue lines represent the results from the direct simulation of step A (system II) and the calculated results for step A based on addition of the yellow and green lines (B+3.5·C). 36

Figure 1.7. Reweighted number of ions displaced from the COOH/COO SAM during LKβ15 adsorption as a function of temperature. Lines I and II refer to the peptide adsorption step (step B) and the overall thermodynamic cycle (step A), respectively. Dashed lines refer to the average number of ions used in each case for later analysis. 37

Figure 1.8. Calculated binding free energies for the three steps of the thermodynamic cycle for LKβ15 on COOH/COO. Yellow and green lines represent steps B (system IV, corrected for ion desorption) and 4·C (system VIIa) of the mechanism proposed in Figure 1.3, respectively, whereas purple and blue lines represent the results from the direct simulation of step A (V) and the calculated results for step A based on adding the yellow and green lines (B+4·C). 37

Figure 1.9. 2D free energy profile of the end-to-end distance for LKβ15 versus the separation distance between LKβ15 and A) the COOH/COO-terminated SAM surface (system IV) at 300 K, B) the COOH/COO-terminated SAM surface at 384.5 K, C) the CH_3-terminated SAM surface (system VI) at 300 K, and D) the CH_3-terminated SAM surface at 381 K. Phase space is restricted to include only surface-bound states. Black boxes indicate areas for clustering analyses (each box indicates a separately performed clustering analysis, labeled by the number next to it). 39

Figure 2.1. (top left) Free energy projected onto the distance between the sidechain (SC) nitrogen atom of lysine and silica for the two biasing schemes: bias added to all atoms in the peptide through the center-of-mass (COM), or bias applied to just the SC nitrogen atom. The black line is the result of reweighting. (top right) Change in Helmholtz energy between the solvated and adsorbed state of GGKGG on silica for the two different biasing schemes. A gray box highlights the "convergence region" of +/- k_BT of the average of the final ΔA values of the three methods of ~ -8.6 kJ/mol. (bottom) Snapshots from each of the labeled minima in the top left plot. The surface and sodium ions are shown in gray and pink, respectively, and are restricted to within 1 nm of the peptide. The peptide backbone is shown in purple, and carbon, hydrogen, oxygen, and nitrogen atoms are shown in cyan, white, red, and blue, respectively. Water is not pictured for clarity. 49

Figure 2.2. Free energy (kJ/mol) along the distance of the nitrogen atom of the lysine sidechain from the surface using enhanced sampling method WTM (black), PTMetaD-WTE (blue), and PBMetaD (purple). 51

Figure 2.3. Free energy projected onto the COM-surface distance with: (A) no electrolyte, (B) excess $Na^{0.5+}$ ions, (C) excess Na^+ ions, and (D) excess Ca^{2+} ions. Green lines indicate thermally-averaged ion binding profiles from PBMetaD simulations. Solid and dotted purple lines indicate peptide binding profiles from PBMetaD and WTM simulations, respectively. 52

Figure 2.4. Snapshots from a WTM simulation with added electrolyte $Ca^{2+}(Cl^-)_2$. The surface and peptide are colored as previously described (Figure 2.1). Ca^{2+} ions are shown in green. Explicit water molecules and Cl^- ions are not pictured for clarity. 53

Figure 2.5. Distance of the sidechain nitrogen of lysine from the surface, as sampled by (left) WTM and (top) PBMetaD (bottom). Each walker in the PBMetaD simulation is colored separately, with the x-axis denoting cumulative simulation time. Each simulation starts from independent starting structures and evolves sharing the same bias potential for 166.67 ns each. 55

Figure 3.1. Side view of LKα14 (sidechains shown in space-filling representation and hydrogen not included) on a SAM surface with a Type I substrate defect causing areas of shortened alkyl chain lengths. The +z direction is orthogonal to the SAM surface and the +x direction is out of the plane of the page. Chains are colored to highlight frozen atoms (silver = frozen CH_2 atoms) and head group atoms allowed to remain free during MD simulation (teal = carbon, yellow = hydrogen, and red = oxygen). 59

Figure 3.2. Distribution of healthy to defective (i.e. short) chains for the three Type I defect simulation trials. The +z direction is out of the plane of the page. Cyan and magenta represent healthy and defective chains, respectively. 59

Figure 3.3. Side view of LKα14 (sidechains shown in space-filling representation and hydrogen not included) on a SAM surface with a Type II film self-assembly defect causing inwards and outwards boundary effects. The +z direction is orthogonal to the SAM surface and the +y direction is out of the plane of the page. Chains are colored to highlight frozen atoms (silver = frozen CH_2 atoms) and head group atoms allowed to remain free during MD simulation (teal = carbon, yellow = hydrogen, and red = oxygen). 60

Figure 3.4. Convergence of free energy differences between solvated and adsorbed states for PTMetaD-WTE simulations at 300 K. The negative value implies a decrease in free energy upon adsorption. 63

Figure 3.5. Helmholtz free energy as a function of LKα14/SAM distance for PTMetaD-WTE simulations at 300 K: a) Type I defect simulations, trials I-III; b) Type II defect simulation, energy minima highlighted in inset; and c) control simulation. Note that the relative energy scale is arbitrary owing to the trivial constant introduced in the estimation of the free energy from the MetaD bias potential. 63

Figure 3.6. Top 3 surface-bound cluster center conformations from a clustering analysis of the Type I, trial III defect simulation compared to the control simulation with no chain defects. Secondary structure is indicated by peptide backbone color: purple designates an α-helix, magenta a turn, and cyan a random coil. Silver and pink represent healthy and defective chains, respectively. 65

Figure 3.7. Top 3 surface-bound cluster center conformations from a clustering analysis of the Type II defect simulation for each energy minima highlighted in Figure 3.5b. Secondary structure is indicated by peptide backbone color: purple designates an α-helix, blue a 3_{10}-helix, magenta a turn, and cyan a random coil. 66

Figure 4.1. Flow chart for developing MD simulation files for an enzyme/IL system to be simulated in GROMACS. This workflow is general and could be used with any MD engine compatible with modern force fields. 75

Figure 4.2. Free energy surfaces for alanine dipeptide in: (A) water; (B) [BMIM][Cl] with partial charges scaled by multiplying by 0.8; (C) [BMIM][Cl] with partial charges scaled by multiplying by 0.9; and (D) [BMIM][Cl] with partial charges scaled by multiplying by 1.0 (i.e. not scaled). 77

Figure 5.1. Ionic liquid ions. R = ethyl (Et or E), butyl (B), pentyl (P), hexyl (H), and octyl (O). 84

Figure 5.2. Simulated vs. experimental IL properties: A) density (g/cm^3), square = TIP3P water, triangle = SPC/E water; B) heat capacity (J/mol-K); C) molar enthalpy of vaporization (kJ/mol); and D) shear viscosity (Pa•s). The x and y axes show experimental and simulated results, respectively. Brackets show results for TIP3P and SPC/E water with experimental results listed first. 87

Figure 5.3. Simulated vs. experimental IL densities for this work (black squares) and the work of Sambasivarao and Acevedo[165] using a refined IL-specific OPLS-AA force field (gray circles). 89

Figure 5.4. Simulated vs. experimental IL molar enthalpies of vaporization for [RMIM][Tf_2N] ILs (R=E, B, H, and O). Simulation results are also shown from the work of Santos et al.[287], using a refined OPLS-AA force field, and Kelkar and Maginn[288], using the CHARMM 22 force field.. 93

Figure 5.5. Simulated vs. experimental IL cation and anion self-diffusivities for [RMIM][Tf_2N] ILs (R=E, B, H, and O): A) simulation results of this work compared to experimental values[293] at 298.15 K and 1 bar; B) simulation results from Köddermann and Paschek[292] compared to experimental values[293] at 303 K and 1 bar. 95

Figure 5.6. Simulated vs. experimental[293] cation to anion self-diffusivity ratios for [RMIM][Tf_2N] ILs (R=E, B, H, and O). Simulation results from Köddermann and Paschek[292] are included for comparison. 96

Figure 6.1. Side-by-side comparison of the structures of LPMO enzymes *Sc*LPMO10B, left, and *Sc*LPMO10C, right. Magenta and gray coloring indicate regions with α- and 3_{10}-helical character, respectively, and yellow and cyan coloring indicate regions with β-barrels and turns, respectively. The location of the copper active site within each enzyme is also highlighted. Each copper ion is coordinated to a bidentate histidine residue, a different monodentate histidine residue, and a monodentate acetate ion ligand (substituted for a coordinating water molecule in *Sc*LPMO10C for consistency with *Sc*LPMO10B). Dotted black lines signifying bonds between atoms coordinated to the central copper ion in the active site are included to guide the eye. 104

Figure 6.2. Side-by-side comparison of the time series of the Cα RMSD of *Sc*LPMO10B (left) and *Sc*LPMO10C (right) from the crystal structure for each enzyme/solvent combination. Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer). 106

Figure 6.3. Composition of the residues within the 3-helix bundles of enzymes *Sc*LPMO10B (A) and *Sc*LPMO10C (C), with positively charged amino acids highlighted in blue, negatively charged amino acids in pink, hydrophobic amino acids in red, and polar amino acids in green. Zoomed in representations of the enzymes' 3-helix bundles (B, D) are also shown, with residues colored to correspond to the plots at the left. 108

Figure 6.4. Side-by-side comparison of the stability of the 3-helix bundle of each enzyme: A) RMSD of the helical regions of *Sc*LPMO10B in water and in each IL; and B) RMSD of the helical

regions of *Sc*LPMO10C in water and in each IL. Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer). 109

Figure 6.5. Side-by-side comparison of the Cα RMSD of the helical regions compared to the crystal structure for extended systems *Sc*LPMO10B/20%[BMIM][Cl]-trial 1 (left) and *Sc*LPMO10B/Water-trial 1 (right). Plot insets show snapshots from the simulations of the enzyme's structure compared to the crystal structure (zoomed in), with letters marking the time at which the snapshots were extracted from the trajectories. The enzyme's crystal structure is colored entirely transparent silver, whereas the NPT structures are colored transparent dark gray, with α-helical regions colored purple to highlight structural differences in this region. 110

Figure 6.6. Zoomed out versions of (left) the bottom inset of the left plot of Figure 6.5, corresponding to the *Sc*LPMO10B/20%[BMIM][Cl]-trial 1 simulation, and (right) the inset of the right plot of Figure 6.5, corresponding to the *Sc*LPMO10B/Water-trial 1 simulation. The snapshots were extracted from points in the simulations at which maximum structural fluctuations occurred (c.f. Figure 6.5). 110

Figure 6.7. Smoothed pairwise distribution functions between the COM of the Cα atoms of the 3-helix bundle in: A) *Sc*LPMO10B, and surrounding cations; B) *Sc*LPMO10C, and surrounding cations; C) *Sc*LPMO10B, and surrounding anions; and D) *Sc*LPMO10C, and surrounding anions. Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer)..... 112

Figure 6.8. Smoothed pairwise distribution functions between the oxygen atoms in water molecules and the centers-of-mass of the 3-helix bundles in *Sc*LPMO10B (black) and *Sc*LPMO10C (gray). The two trials with each enzyme are shown with solid and dotted lines.. 113

Figure 6.9. Side-by-side comparison of the difference in RMSF between each simulation and the average RMSF of the two water simulations, for *Sc*LPMO10B (left) and *Sc*LPMO10C (right). Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer)..... 114

Figure 6.10. Crystal structures of *Sc*LPMO10B (left) and *Sc*LPMO10C (right), highlighting regions of high RMSF values in accordance with Figure 6.9. Number ranges next to highlighted areas indicate the corresponding sequence range. 115

Figure 6.11. Side-by-side comparison of smoothed pairwise distribution functions between the copper ion in the active site and surrounding water for *Sc*LPMO10B (left) and *Sc*LPMO10C (right). Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer)..... 116

Figure 6.12. Side-by-side comparison of the distance between the copper ion and alanine residue (Cβ atom) near the active site for *Sc*LPMO10B (left) and *Sc*LPMO10C (right). Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer). 117

Figure 7.1. Visual representation of the 12 IL cation binding sites on the surface of WT-lipA determined from crystal structures derived from experiments. Binding site residues are shown in both licorice and space-filling representation, and hydrogen atoms are excluded for clarity. ... 122

Figure 7.2. Percent occupancy of residues in WT-lipA by IL cations as a function of IL concentration in water. Colored circles represent experimentally identified binding sites, and uncolored circles represent all other protein residues. ... 123

Figure 7.3. Superimposed IL anion binding poses from simulations of 20% IL around residues in (left) wild-type lipase and (right) mutant lipase. Anions are shown as green spheres. C, O, and N atoms are shown in in cyan, red, and blue, respectively, and protein secondary structure is shown in gray. Hydrogen atoms are not pictured for clarity. ... 126

Figure 7.4. Determining the role of mutations in changing residue specific binding and protein folding. The top panels (A/B) show the percent occupancy by IL cations (A) and anions (B) in the wild-type (WT-lipA) (x-axis) vs the change in percent occupancy (mutant minus wild-type) from a 20% aqueous [BMIM][Cl] solution at 300 K. The bottom panels (C/D) show the effect of ion binding on the denatured lipase (WT-Den-lipA). Shaded circles represent experimentally identified binding sites, open circles represent all other protein residues, and squares represent mutated residues. All y-axes (change in ion binding occupancy) are presented as mutant minus wild-type. ... 127

Figure 7.5. Pair correlation functions for cations and anions within 4 Å of (left) the folded state and (right) the unfolded state of wild-type and mutant lipase. ... 129

Figure 8.1. Example snapshot of an interfacial simulation of LKα14 peptides in water. Peptide secondary structures are shown in gray, atoms in leucine residues are shown in yellow, and atoms in lysine residues are shown in cyan (CPK format). Solvent oxygen atoms are represented as red beads. ... 133

Figure 8.2. Comparison of the center-of-mass position over time of LKα14 and LKβ15 systems in water and 50% aqueous [BMIM][Ace]. Data for a single trial is shown of the three total trials for each simulation. Note that solvent is present between 0 and 6 nm, while the area above 6 nm represents vacuum. The different colors represent the five different peptides in each system. . 137

Figure 8.3. Convergence of the distance between LKα14 and the vacuum interface with water and [BMIM][Ace] for metadynamics simulations. Note that the system is still diffusing in *x* during the last 50-100 ns of each run. ... 138

Figure 8.4. Free energy profiles with respect to distance from the interface for LKα14 in water and in aqueous [BMIM][Ace]. The interface is located at approximately 0.5 nm and the free energy is truncated at 4 nm due to the presence of a harmonic wall potential. ... 139

Figure 8.5. Atomic species densities for LKα14 and LKβ15 in aqueous solutions with 50% [BMIM][Ace] or [BMIM][Cl], and in pure water. Data for a single trial is shown of the three trials for each simulation. Lines are shown for water (blue), peptide backbone Cα carbons (red), peptide sidechain Cδ carbons (green), [BMIM] cations (purple), and either [ACE] or [Cl] anions (cyan). Note that (1) both types of carbons are included to show peptide orientation, (2) solvent is present between 0 and 6 nm, while the area below 0 nm and above 6 nm represents vacuum, and (3) that densities for different species are normalized only with respect to themselves for easier viewing; in reality, solvent densities are universally much higher than peptide densities. 141

Figure 8.6. Atomic species densities for the [BMIM] cation in an interfacial system with LKα14 peptides. The black line represents the density of the last two carbons on the hydrophobic tail of the [BMIM] cation, whereas the orange line represents the density of the three carbons present in the ring of the [BMIM] cation. The y-axis is the same as in Figure 8.5 for easy comparison across species. 142

Figure 8.7. Tilt angles with respect to the vacuum/solvent interface of leucine sidechains for interfacially adsorbed LKα14 peptides in the three different solvents. Data represent averages and standard deviations across all peptides and replicate simulations. 143

Figure 8.8. Comparison of conformations adopted by LKβ15 at the vacuum/water interface (left) and the interface of vacuum with 50% aqueous [BMIM][Ace] (right), along with corresponding Ramachandran plots. Angle data was calculated for a single peptide over the last 10 ns of a simulation – the representative snapshots were taken from a single trajectory frame from this range. Green, cyan, silver, white, and blue coloring indicate β-sheet character, turns, random coils, leucine, and lysine residues, respectively. Acetyl capping groups are shown in light blue. High-density regions are outlined in red; low density regions in blue..... 144

Figure 8.9. Comparison of secondary structures formed by LKβ15 both at the interface and in bulk in pure water and 50% [BMIM][Ace]. High-density regions are outlined in red; low density regions in blue. For each plot, the data was obtained by averaging angle data for all peptides over the last 10 ns of a single simulation. 146

Figure 8.10. Tilt angles with respect to the vacuum/solvent interface of leucine sidechains for interfacially adsorbed LKβ15 peptides in the three different solvents. Data represent averages and standard deviations across all peptides and replicate simulations. 147

Figure 9.1. Protein and surrogates used to model polymer nanoparticles in MD simulations. Surrogates include trimers of PLGA, PS-PLA, and PLA. Purple, blue, yellow, cyan, and white coloring indicate an α-helix, α_{310}-helix, β-sheet, turn, and random coil, respectively, and the protein (BSA) is also shown in surface representation in gray..... 153

Figure 9.2. Percent BSA release profile of BSA released against total BSA initially encapsulated for PLA (black), PS-PLA (red), and PLGA (blue) nanoparticles (n=3 batches for all samples). 154

Figure 9.3. Final snapshots from simulation trajectories of BSA in aqueous solutions with 2 wt.% PS, without (A) and with (B) restraints between PS atoms in neighboring molecules to prevent self-aggregation. BSA is shown in surface representation in transparent gray coloring to show PS binding to all sides of the surface. PS molecules are shown with carbon and hydrogen atoms colored cyan and white, respectively. Water and neutralizing counterions are not pictured..... 156

Figure 9.4. Fraction of contacts to the different amino acid types on the surface of BSA by polymer surrogates. Error bars represent the average standard deviation of calculations performed on consecutive 25 ns portions over the second half of each 200 ns simulation trajectory (i.e., sample size of n=4). 157

Figure 9.5. Occupancy heat maps for BSA residues: (left) BSA residue occupancy for PLGA, PLA, PS-PLA, and PS simulations. PS simulations include a coordination potential to discourage aggregation. Occupancy was calculated as the fraction of surrogate contacts for each residue over the last 100 ns; (right) PLA, PS-PLA, and PS occupancy from the PS-PLA simulation. 158

Figure 9.6. Top: Representative structure of poly(styrene)-acrylate (PSAC) copolymer. Bottom: Snapshot from the end of a 200 ns simulation trajectory of BSA in an aqueous solution with 2 wt.% PSAC, with no restraints to discourage PSAC self-aggregation. 160

Figure 9.7. Fraction of contacts for different amino acid types on the surface of BSA in an aqueous solution with PSAC compared to with PS-PLA or PLGA. Error bars represent the average standard deviation of calculations performed on consecutive 25 ns portions over the second half of each 200 ns simulation trajectory (i.e., sample size of n=4). 161

Figure 9.8. Fraction of contacts for different amino acid types on the surface of BSA in solutions with monomeric or trimeric LGA or PLGA copolymer, or SAC or PSAC copolymer, respectively. Error bars represent the average standard deviation of calculations performed on consecutive 25 ns portions over the second half of each 200 ns simulation trajectory (i.e., sample size of n=4). 162

Figure A1.1. Reweighted potential energies averaged over discrete peptide-surface separation distances (SSD) for LKα14 on the COO/COOH SAM (n=51 slices over the range of x-axis values shown above) as a function of temperature. 168

Figure A1.2. Calculated change in potential energy upon adsorption (adsorbed - solvated state) of LKα14 to the COO/COOH SAM as a function of temperature. Adsorbed potential energies correspond to an SSD value of ~0.7 nm (c.f. Figure A1.1) and solvated potential energies were calculated as averages over a range in SSD of 1-3.5 nm. Error bars represent the standard deviation in potential energy over a range in SSD of 1-3.5 nm. A linear fit of the data is shown as a dotted black line. 168

Figure A1.3. Calculated binding free energies versus replica temperature for systems VIIa/b and VIIIa/b, corresponding to systems VII (COOH/COO SAM) and VIII (CH_3 SAM) listed in Table 1

(main text), respectively, where a single ion/water molecule was biased. The lines represent binding free energies of step C of the thermodynamic cycle of adsorption shown in Figure 1.3 (main text). 169

Figure A1.4. Calculated binding free energies for the three steps of the thermodynamic cycle for LKα14 on COOH/COO, using a value of *x,* or the average number of ions desorbing from the surface, unique to each replica temperature. Yellow and green lines represent steps B (system I, corrected for ion desorption) and #Ions·C (system VIIa) of the mechanism proposed in Figure 1.3 (main text), respectively, whereas purple and blue lines represent the results from the direct simulation of step A (system II) and the calculated results for step A based on adding the yellow and green lines (B+#Ions·C). 169

Figure A1.5. Calculated binding free energies for the three steps of the thermodynamic cycle for LKβ15 on COOH/COO, using a value of *x,* or the reweighted number of ions desorbing from the surface, unique to each replica temperature. Yellow and green lines represent steps B (system IV, corrected for ion desorption) and #Ions·C (system VIIa) of the proposed mechanism, respectively, whereas purple and blue lines represent the results from the direct simulation of step A (system V) and the calculated results for step A based on adding the yellow and green lines (B+#Ions·C). 170

Figure A1.6. Unidirectional density of water above: 1) the COOH/COO surface (pink curve, system VII); 2) the COOH/COO surface with 100 additional sodium and chlorine ions each (blue curve); and 3) the CH_3 surface (green curve, system VIII). The z-component of the topmost frozen carbon atom (C10) in each surface was used as the point of reference. 170

Figure A2.1. Atoms used in quantum mechanics calculations of the copper-histidine coordination complex in enzymes *Sc*LPMO10B and *Sc*LPMO10C. 171

Figure A2.2. Time series of the Cα RMSD from the crystal structure for extended systems *Sc*LPMO10B/20%[BMIM][Cl]-trial 1 (green) and *Sc*LPMO10C/10%[AMIM][Cl]-trial 1 (red). 172

Figure A2.3. Side-by-side comparison of the time series of the Cα RMSF of *Sc*LPMO10B (left) and *Sc*LPMO10C (right) from the crystal structure for each enzyme/solvent combination. Legend entries indicate the IL, wt% in water, and trial number. 172

Figure A3.1. Percent occupancy of WT-lipA residues by IL cations, calculated over increasing periods of simulation time, for the residues in the top five cation-binding sites identified from the experimental crystal structures. 173

Figure A3.2. Snapshot of an IL cation stabilized between two proteins in the crystal structure of lipA in 20% aqueous IL solution (PDB code 5CT6). Residues in binding sites 5 (yellow) and 6 (cyan) are shown in licorice representation. The IL cation is shown in van der Waals representation with carbon, nitrogen, and hydrogen atoms shown in cyan, blue, and pink, respectively..... 174

Figure A3.3. Residence time of IL cation binding to residues in WT-lipA as a function of IL concentration in water. Colored circles represent experimentally identified binding sites, and uncolored circles represent all other protein residues. 175

Figure A3.4. Residence time of IL cation binding to residues in QM-lipA as a function of IL concentration in water. Colored circles represent experimentally identified binding sites, and uncolored circles represent all other protein residues. Vertical gray lines correspond to residues mutated in mutagenesis experiments. 175

Figure A3.5. Simulation snapshot from the 5% IL solution simulation highlighting the location of residue 62 (valine) on the surface of lipase (in gray) and its interaction with an IL cation. Hydrogen atoms not pictured. 176

Figure A3.6. Time series of the root-mean-squared deviation (RMSD) of distances between Cα atoms in WT-lipA for each of the different IL concentrations studied. Data was collected from the trajectory every 20 ps and plotted every 10^{th} point. 177

Figure A3.7. (left) Time series of the root-mean-squared fluctuation (RMSF) of Cα atoms (sequence) in WT-lipA for each of the different IL concentrations studied. (right) crystal structure of WT-lipA, highlighting regions of high RMSF values. Number ranges next to highlighted areas indicate the corresponding sequence range. Magenta and blue coloring indicate α- and 3_{10}-helices, respectively, and yellow, cyan, and dark gray coloring indicate β-sheets, turns, and random coils, respectively. 177

Figure A3.8. RMSF vs. percent occupancy of residues in WT-lipA by IL cations as a function of IL concentration in water. Colored circles represent experimentally identified binding sites, and uncolored circles represent all other protein residues. 178

Figure A3.9. Percent occupancy of residues in QM-lipA by IL cations as a function of IL concentration in water. Colored circles represent experimentally identified binding sites, and uncolored circles represent all other protein residues. Vertical gray lines correspond to residues mutated in mutagenesis experiments. 178

Figure A3.10. Radius of gyration and RMSD (relative to the crystal structure) of denatured WT-lipA versus time with corresponding simulation snapshots (insets). 179

Figure A4.1. Average x-position of the five highest central ring carbons in [BMIM] (from the bottom of the simulation box) as a function of simulation time, for the simulation of (left) LKα14 and (right) LKβ15 in the IL solutions. 180

Figure A4.2. Atomic species densities for LKα14 in 50% aqueous [BMIM][Cl]. Data for a single trial is shown of the three total trials. Note that solvent is present between 0 and 6 nm, while the

area below 0 nm and above 6 nm represents vacuum. The y-axis is the same as in Figure 8.5 (main text) for easy comparison across species. 180

Figure A4.3. Atomic species densities for LKβ15 in 50% aqueous [BMIM][Cl]. Data for a single trial is shown of the three total trials. Note that solvent is present between 0 and 6 nm, while the area below 0 nm and above 6 nm represents vacuum. The y-axis is the same as in Figure 8.5 (main text) for easy comparison across species. 181

Figure A5.1. Convergence of amino acid contacts on BSA in polymer surrogate solutions..... 181

Figure A5.2. Convergence of amino acid contacts on BSA in a 2 wt.% PLGA solution over an extended period of 500 ns. 182

Figure A5.3. Root mean squared deviation (RMSD) of α-carbons (Cα) in BSA compared to the crystal structure in the different aqueous polymer solutions and in pure water. 182

Figure A5.4. Occupancy heat maps for PS simulation with and without coordination constraint. Aggregation in unconstrained simulation (top) results in relatively sparse binding. 183

Figure A5.5. Convergence of amino acid contacts on the surface of BSA in an aqueous solution with 2 wt.% PSAC. 183

Figure A5.6. Root mean squared deviation (RMSD) of Cα atoms in BSA compared to the crystal structures in PSAC and in pure water. 184

Figure A5.7. Snapshots from simulation trajectories of BSA in aqueous solutions with 2 wt.% PSAC, with: (A) PSAC molecules with protonated acrylate side chains, leading to self-aggregation, and (B) deprotonated acrylate side chains, approximating a solution at the intracellular pH of the brain of around 7.2. 184

Figure A5.8. Root mean squared deviation (RMSD) of Cα atoms in BSA compared to the crystal structure in monomeric vs. trimeric polymer surrogate solutions and in pure water. 185

Figure A5.9. Convergence of amino acid contacts on BSA in (left) monomeric vs. trimeric LGA or PLGA copolymer, and (right) monomeric vs. trimeric SAC or PSAC copolymer, respectively. 185

List of Tables

Table 1.1. Setup of PTMetaD-WTE simulations. 27

Table 1.2. Clustering analysis results for surface-bound structures of LKβ15. 40

Table 2.1. List of metadynamics parameters and collective variables used for different simulations of GGKGG on silica at pH 7.5 47

Table 3.1. Setup of PTMetaD-WTE simulations. 60

Table 4.1. Overview of enzyme/IL systems studied to date with molecular simulations. 71

Table 4.2. Abbreviations and formulas for IL cations and anions listed in Table 4.1. 72

Table 4.3. Mean escape time data for the alanine dipeptide $\alpha \rightarrow \beta$ transition. 78

Table 4.4. Common methods of analysis for enzyme/IL systems. 80

Table 5.1. Experimental and simulated IL densities (g/cm^3). Values in parentheses represent simulated densities from Vega and Abascal.[224] 88

Table 5.2. Experimental and simulated IL heat capacities (J/mol-K). Values in parentheses represent simulated heat capacities from Vega and Abascal[224]. 90

Table 5.3. Experimental and simulated IL molar enthalpies of vaporization (kJ/mol). Values in parentheses represent simulated molar enthalpies of vaporization from Santos et al.[287], Kelkar and Maginn[288], and Vega and Abascal[224]. 92

Table 5.4. Experimental and simulated IL self-diffusivities (10^{-11} m^2/s). Simulated self-diffusivities from Köddermann and Paschek[292] and Vega and Abascal[224] are included for comparison. 94

Table 5.5. Experimental and simulated IL shear viscosities (Pa•s) using a shear rate of 0.03 nm/ps^2. Values in parentheses represent simulated shear viscosities from Vega and Abascal[224]. Bolded ILs are not included in Figure 5.2D, which is restricted to a range that highlights good fits. 99

Table 5.6. Experimental and simulated IL shear viscosities (Pa•s) using a shear rate of 0.03 nm/ps^2 and temperatures above room condition. 100

Table 6.1. Size of each simulation box and the number of all solvent molecules for each system. 104

Table 6.2. Average RMSD and standard deviation of Cα atoms from the crystal structure for each enzyme/solvent system, calculated over the second half of the 250 ns trajectories. The error bars are one standard deviation of the RMSD over the same period. 106

Table 7.1. Percent occupancy of IL cations and anions to residues in WT-lipA before mutation for all IL solutions. 125

Table 7.2. Percent occupancy of IL cations and anions to mutated residues in QM-lipA for all IL solutions. 125

Table 8.1. Summary of simulation parameters. Note that all unbiased simulations were run in triplicate. The average number of particles in unbiased simulations is ~18,500 and is ~24,000 in biased simulations. 134

Table 8.2. Diffusivities for all unbiased, bulk systems in 10^{-5} cm^2/s, averaged across the five peptides in all three trials of each system and fit over the first 40 ns of the 2nd half of each trajectory. Error bars represent the standard error of the mean, where the sample size (n=15) is the total number of individual peptide trajectories for the three replicate simulations. 137

Table 8.3. LKα14 degree of helicity as calculated by DSSP in IL solvents and water..... 142

Table 8.4. Degree of secondary structure formation of LKβ15 in IL solvents and water, as calculated by DSSP. The novel helix structure described in the text is not shown, since it does not contain any hydrogen bonding and therefore cannot be detected by DSSP. 145

Table 9.1. Nanoparticle characterization before and after BSA release experiments (n=3 batches for all measurements). 154

Table 9.2. Trimer mean occupancy per BSA residue. 159

Table A1.1. Setup of PTMetaD-WTE Simulations..... 166

Table A2.1. Calculated partial charges for the atoms in Figure A2.1. 171

Table A3.1. Characterization of cation binding sites on the surface of WT-lipA as determined through a combined experimental and computational approach (see Computational Methods, main text). 173

Acknowledgements

Unlike in Hollywood where actors have numerous award ceremonies and red carpet events where they can thank the people who have helped them get to where they are, academia has few such opportunities (unless you receive a Nobel Prize, for example!). Therefore, I would like to take this rare opportunity to thank the many people who have been instrumental in my journey to and through graduate school, and in helping me to realize my dream of becoming a professor.

First of all, I want to say thank you to my dissertation committee members Jim Pfaendtner (Chair), François Baneyx, Shaoyi Jiang, Elizabeth Nance, Eve Riskin, and Gary Drobny, for your advice, feedback, and career insight (and putting up with reading this incredibly long document!).

Thank you to my friends in the department for helping me through the emotional roller coaster that is graduate school, at times. Thank you, Elisa Harrison, Alaina Floyd, and Kelly Fleming for the many coffee dates and Portage Bay lunches that always lifted my spirits. Thank you, Vance Jaeger, Blake Hough, Gaunqun Luo, and Arushi Prakash for being such wonderful and supportive coworkers – and of course, friends – who I am confident will be a part of my life for many years to come. My time in graduate school represents some of the best years of my life, in large part to all of you. Thank you, in addition, to every member of the Pfaendtner Research Group (PRG), both past and present, who I have had the opportunity and pleasure of working with during graduate school. I have so much respect for each and every one of you, and have truly loved working alongside many of you on research projects, writing papers, co-supervising undergraduates, or just attending conferences together. I am very proud to have been a member of PRG; during my 7 years in the group, PRGs have always held to the highest quality standards of research that have pushed the group to excellence and inspired me to perform at my best. PRG has felt like a family to me for many years now, and one that I will miss dearly.

Thank you, Jim Pfaendtner, for being my biggest supporter – you have been so much more than just my graduate school advisor, but also a mentor, life coach, and friend. Thank you for always believing in me, starting back when I was an undergrad and was unaware of what academic research even was, and continuing to believe in me when I applied to the same program at UW for graduate school – a very unorthodox move – and then when I reached for the stars in obtaining a postdoc. Thank you for being patient with me when I doubted my own abilities, and for helping me see that my weaknesses can actually be used to my advantage, like a "secret power". And lastly, thank you for being an amazing role model; apart from your inspiring drive to be the best research mentor and teacher possible, your strong support of women in STEM has led me to feel empowered in my daily life and excited about my future in academia. I am profoundly grateful for all you have taught me, and will continue to teach me, as I aspire to follow in your footsteps.

Lastly, thank you to my family: my sister Matreya, my sister-in-law Jule and her husband Ben, my father-in-law Ekke and his partner Ellen, my parents Lonnie and Gary, and my husband Ole. Thank you all for your unconditional love and support through all of my triumphs and tribulations. I love you all. My parents deserve special recognition for shaping me into the person that I am today; thank you, mom and dad, for always encouraging me to pursue my dreams, and to follow my heart. And thank you, Ole, for being my shoulder to lean on through it all, and for sharing in my accomplishments. I look forward to experiencing all of life's greatest adventures with you, starting with a bold move across the country! Thank you for everything.

Introduction

Some of the greatest innovations in human history have been inspired by natural processes and the molecular machinery that drives them. From early studies of birds to enable human flight[1] to studies of animal diets to discover miracle drugs[2]; from photosynthesis-inspired solar power research[3] to harnessing the coding power of DNA to design the world's fastest computers[2]; and from studies of the synthesis of cell walls and structures of incredible strength by living organisms to enable new nanotechnologies and biomaterials[4], nature has played a crucial role in solving many of the toughest technological challenges that have faced humanity to date.

In the drive to understand the molecular mechanisms and machines responsible for governing such nature-driven processes, proteins have largely dominated the discussion. Proteins are primarily made up of the 20 naturally-occurring amino acids, which can be arranged in any sequence or order for any given chain length L, and then folds into a unique protein structure. This incredible diversity in protein sequence and structure makes possible a wide range of functions that, collectively, allow proteins to carry out all of the nanoscale processes of life. It may be intuitive, then, to think that specific proteins could be readily chosen for use in specific applications based on their individual functions. In early studies of proteins, however, it was discovered that the relationship between a protein's sequence, structure, and function is hardly straightforward, and can instead depend heavily on the protein's surrounding environment[5-6]. Indeed, some researchers have gone so far as to call the surrounding host environment the "21st amino acid."[7]

The mechanism through which a protein interacts with its surrounding microenvironment can be characterized as an interfacial process, and the points of interaction between them as an interface. In a biological context, interfaces can exist between proteins and a surrounding liquid, such as water or blood plasma, and between proteins and a solid surface, such as the flexible membrane of a cell or the tough mineral structures of bones and teeth. Many complex behaviors can take place at these biological interfaces, including molecular recognition and protein conformational change. These complexities are further exacerbated in the presence of a solid surface, where not just one but three interfaces exist simultaneously that can influence each other's behavior (i.e. the protein/liquid, protein/solid, and liquid/solid interfaces). The ability to understand and predict how a protein concurrently interacts with multiple interfaces at the molecular level, or even with just a single interface, is an extremely challenging problem. Yet it must be solved if nature's best and brightest ideas are to be harnessed to address today's technological problems and demands.

Experimental techniques such as nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography have been developed – and successfully applied in many cases – to determine the structure of proteins in solution or in the crystal state, respectively. After obtaining a structure with atomistic resolution, protein coordinates are generally deposited in the Protein Data Bank (PDB). Despite over 100,000 entries, the PDB currently contains no entries of a protein adsorbed to a solid surface, to our knowledge. This is because no single experimental technique exists today that can fully characterize the structure of a surface-bound protein. In addition, even in cases where NMR or X-ray crystallography have been successful in ascertaining the structure of a protein, the coordinates represent "static" snapshots. Essentially, we gain little mechanistic or dynamic insight into the system from looking at these snapshots, e.g., insight into the different configurations explored by the protein before reaching the final coordinates deposited in the PDB, or how long

on average it took for the protein to reach those final coordinates in response to its environment. We also obtain next-to-no information about key interactions between the solution and specific protein residues or moieties, except in rare cases where the coordinates of bound solvent molecules were able to be determined during the experiments in addition to the protein coordinates[8].

Computer simulations, in particular classical molecular dynamics (MD) simulations, are a natural partner to both build upon and complement experiments of protein interfacial processes. Using coordinates from the PDB or other sources, as well as empirically or theoretically-derived data to describe molecular interactions and a set of elementary physical rules to propagate these interactions in time, MD has the potential to provide unprecedented insight into the behavior of proteins in complex interfacial environments. For example, by watching the "trajectory" of an MD simulation, we can directly observe how a protein changes shape upon adsorption to a surface to minimize the overall energy of the system, or how the structure of a protein changes over time (e.g. degrades or unfolds) in the presence of a given fluid. Additionally, MD can be used to provide quantitative details about the preferred structure of an adsorbed protein to supplement experimental data. In cases where we want to simulate the protein/solid interface, however, special simulation techniques must be employed to ensure the results are accurate. In reality, countless proteins sampling all conceivable configurations are present in solution, and then adsorb onto the surface. Thus, the protein coordinates determined via experimental techniques represent an "ensemble average". Since it would be too expensive to perform a simulation with anything more than a few proteins at once, simulations must instead be run for a very long time in order to sample this same set of configurations. Alternatively, a number of what are called "enhanced sampling" simulation techniques have been developed in recent years to address this issue that allow for faster convergence of thermodynamic and structural properties in complex systems. Nonetheless, best practices of how to simulate protein adsorption are still developing and so remains an active and important area of research in the field of computational biophysics.

The work presented in this document employs many of these advanced simulation sampling schemes to study several different aspects of interfacial protein adsorption, both at the protein/solid and protein/liquid interface. The overall goal of this work is to improve the way we use computers to discover and control molecular driving forces at engineered interfaces to optimize the protein or host environment for specific applications. The first three chapters of the document focus on the investigation of peptide behavior – a proxy for protein behavior with reduced computational expense – at the protein/solid interface. In Chapter 1, the binding of model peptides to model surfaces is explored from a mostly thermodynamic perspective, wherein it was discovered that the surface-bound ion content has a large influence on biomolecular binding free energies derived from molecular simulations. This information was the basis for the work described in Chapter 2, which further investigates the effects of different types and concentrations of ions on peptide adsorption. Chapter 3 investigates a slightly more realistic scenario wherein peptides adsorb to surfaces incorporated with different types of experimentally-observed defects to investigate the role of the surface in modulating biomolecular structure. At this point, the document switches to focus on the investigation of proteins in contact with non-solid interfaces. Chapter 4 is an overview of the molecular simulation approaches ("force fields") and analysis methods used to study protein interactions with a particular type of solvent called ionic liquids, or ILs, which will be described in detail throughout the document. In Chapter 5, GAFF – a common force field used to model ILs – is extensively tested for its ability to reproduce thermodynamic and transport properties of a wide range of ILs, comparing directly against available experimental data. Chapter 6 is a foray into the realm of prediction. Simulation methods and analysis techniques

described in the previous two chapters are used to study the behavior of two proteins recently discovered in cellulolytic cocktails, in ILs. As there is currently no experimental data to compare against for these particular systems, the results from our simulations – that the enzymes are highly stable in a wide array of ILs and thus are strong candidates for biomass conversion in ILs – can hopefully serve as a guide for future experiments. In stark contrast to Chapter 6, Chapter 7 is a detailed quantitative validation using experimentally measured structures of the accuracy of MD in predicting protein/IL interactions. We find MD to be highly accurate in this respect, with easily explainable differences. The final chapters explore biomolecular behavior in slightly different environments, though still at interfaces. Chapter 8 is an investigation of peptide structure and dynamics at the air/water interface in the presence of ILs. Finally, utilizing a combined computational and experimental approach, Chapter 9 explores the interactions between enzymes and small polymer surrogate molecules to provide new insights into the protein encapsulation process for biomedical applications. The chapter concludes with a discussion of the overall significance and future directions of this work.

Chapter 1

Strong Electrostatic Interactions Lead to Entropically Favorable Binding of Peptides to Charged Surfaces[1]

Introduction

Protein adsorption at solid surfaces has been extensively researched due to its widespread application in the fields of biocatalysis, biomedicine, biomaterials, and in the broader area of biotechnology. Prior technological progress in these applications has benefited from increased molecular-level understanding of the interactions of proteins and surfaces such that protein adsorption can be controlled and tuned to a desired degree. To achieve this goal of rationally designing the protein/surface interface with specific functionalities, a comprehensive understanding of the interfacial structural and thermodynamic behavior of adsorbed proteins and the underlying mechanisms they govern is essential. Due to a number of long-standing challenges, experiments have been limited in their ability to obtain all the information necessary to characterize protein/surface interactions; no single experimental technique exists today that can fully resolve the structure or dynamics of proteins adsorbed at interfaces. This creates a unique window for molecular simulations to fill the gaps in experimental knowledge. The combination of the two allows us to obtain a significantly more detailed picture of the adsorption process.

The lack of experimental data also leads to difficulties in investigating molecular level hypotheses and validating molecular models regarding protein adsorption.[9] Thus, there is a pressing need for model systems that are experimentally well characterized.[10] Self-assembled monolayer (SAM) surfaces have long been regarded as strong candidates for fundamental studies to examine protein/surface interactions in detail.[9] Additionally, LK peptides, originally designed to permit study of adsorption-induced conformational changes[11], also fall into the category of being excellent contenders for fundamental modeling studies of protein adsorption. A great number of experimental studies employing a wide variety of techniques like solid-state nuclear magnetic resonance (ssNMR)[12], near edge X-ray adsorption fine structure (NEXAFS)[13-14], sum frequency generation (SFG)[15-18], X-ray photoelectron spectroscopy (XPS)[19], and time-of-flight

[1]Reproduced in part with permission from K.G. Sprenger and J. Pfaendtner. Strong Electrostatic Interactions Lead to Entropically Favorable Binding of Peptides to Charged Surfaces. Langmuir, 32:5690-5701, 2016. Copyright 2016 American Chemical Society.

secondary ion mass spectrometry (ToF-SIMS)[19] have been performed on systems of LK peptides adsorbed to solid surfaces, many with SAM surfaces specifically, to determine interfacial conformations and orientations of the adsorbed peptides. Some studies have gone a step further to investigate mechanisms of LK peptide aggregation[4, 20-21] and complex/film formation at interfaces[13, 22]. Surprisingly fewer experimental studies have focused on the kinetic[23] or thermodynamic aspects of LK peptide adsorption. Thermodynamic analyses are important because they can provide key mechanistic insights into protein self-assembly on surfaces and the basis of protein function and stability in general.[24] Calculating the change in the standard-state adsorption free energy is a convenient way to directly evaluate the primary thermodynamic driving force for adsorption[25]. Surface plasmon resonance (SPR) and atomic force microscopy (AFM) can be used together to obtain peptide adsorption free energies[25-26], as can a combination of isothermal titration calorimetry and equilibrium adsorption isotherm analysis[24], which can also be used to break down the adsorption free energy change into the separate enthalpic and entropic contributions. These types of experimental analyses have been applied to systems such as zwitterionic host-guest peptides adsorbing on SAM surfaces[27] and to statherin adsorption on hydroxyapatite[24], but to our knowledge they have never been applied to LK peptides adsorbing on SAMs. Developing the tools to carry out such analysis computationally is a logical step in studying model systems.

Though molecular simulations have the potential to provide atomic level details about the adsorption process, computational progress to date has been hindered by a number of problems. One issue is the lack of available experimental data that can be directly compared to simulation results to validate force fields[28], and thus force field accuracy in regards to peptide/protein adsorption remains an open question. On the other hand, prior work from our group on similar systems showed similar structure and dynamics across three commonly used molecular dynamics (MD) force fields[29], and thus a standard force field was deemed sufficient for use in this study. Other challenges of simulating interfacial biomolecular adsorption include timescale limitations that limit relevant exploration of conformational phase space, and strong surface binding affinities that lead to kinetic trapping of the biomolecules at the surface and difficult (or impossible) determination of adsorption free energies. These issues are especially true for classical MD simulations. Fortunately, a great deal of work has gone into developing new algorithms that can circumvent the timescale and sampling limitations of MD, and many of these methods have now been applied to study peptide adsorption at interfaces. These techniques work by biasing one or more relevant, slow degrees of freedom (collective variables (CVs)) that can effectively describe the system in reduced dimension. Applications of enhanced sampling techniques to peptide adsorption include parallel tempering (PT) and its variations[30-33], umbrella sampling (US)[29, 34-37], and steered molecular dynamics (SMD)[38-41]. Although there are many choices that work well for sampling peptide adsorption, our approach utilizes a variation of Metadynamics (MetaD)[42], which is particularly powerful since the peptide's degrees of freedom can also be taken into account. In some cases neglecting to account for these degrees of freedom can lead to widely varying results, for example in an umbrella sampling-based scheme, if enormous simulation times per window are not used to relax the other degrees of freedom[29]. In this study we combine well-tempered Metadynamics with parallel tempering (PTMetaD)[43]. Simulations are performed in the well-tempered ensemble (WTE)[44] to increase the potential energy fluctuations between replicas and thus the overall efficiency of the PT sampling. This technique, called PTMetaD-WTE, was previously employed to study the adsorption of model LK peptides on SAMs[29], DNA adsorption and folding on model surfaces[45], and folding of the model miniprotein tryptophan cage (trpcage) on SAMs.[46] Because of the large system sizes typically needed in peptide or protein adsorption

simulations, PTMetaD-WTE has proven to be particularly useful in its ability to reduce the number of required replicas needed to span a meaningful temperature range (and thereby make use of temperature to overcome energetic barriers). The WTE boost, in many situations, can also reduce the round trip time of the random walk in temperature space and thereby further increase sampling efficiency.

Building off our past work[29], which focused on determining side chain orientations and preferred conformations of adsorbed LK peptides on SAMs, PTMetaD-WTE is used to simulate similar systems of LK peptides adsorbing to SAMs, but with a focus on determining the role of peptide/surface charge and ion concentration on binding thermodynamics. In particular, we seek to decompose the obtained free energies and determine the respective energetic and entropic components of adsorption. As we show below, the simulations present a very surprising result, namely an apparent increase in peptide binding strength to a charged surface with increasing temperature. This result is completely understood via a combination of independent MetaD simulations that break down the different parts of the thermodynamic cycle of adsorption (highlighting the critical role of ionic solution strength) and a configurational entropy analysis of both the peptide and solvent degrees of freedom upon adsorption. Clustering and reweighting analysis techniques are employed to estimate configurational entropy from a biased simulation. This study also demonstrates one way in which a complicated result from a single PTMetaD-WTE simulation (i.e., one representing a multi-step thermodynamic cycle) can be decomposed through multiple independent simulations to glean crucial information about the adsorption process and provide needed molecular level insight to macroscopic behavior. For consistency with past simulations[29], carboxylic acid/carboxylate- and methyl-terminated SAMs were chosen as the model hydrophilic and hydrophobic surfaces for this study, respectively, and LKα14 and LKβ15 were chosen as the model peptides[11]. Though both peptides consist only of leucine and lysine residues, their different sequences give rise to unique secondary structures presenting interesting opportunities for behavioral comparison: LKα14 with an amino acid sequence of LKKLLKLLKKLLKL adopts an α-helical conformation at interfaces, whereas LKβ15 with an amino acid sequence of LKLKLKLKLKLKLKL displays β-strand character.

Methods

1.2.1 System Setup

Specifications of each simulation are given in Table 1.1. Data from systems III and VI, and IX and X were taken from previously published papers from our group.[29, 46] Unless otherwise noted, discussion of simulation parameters and methodology refers only to new simulations performed for this study. A full list of parameters for systems III, VI, IX, and X can be found in Table A1.1 in the Supporting Information (SI) and other minor methodological details can be found elsewhere.[29, 46] Systems consisted of one LK peptide, either LKα14 or LKβ15, a SAM surface with either a carboxylic acid/carboxylate- (COOH/COO) or methyl-terminated (CH_3) dodecane thiol chain, TIP3P waters, and chlorine or sodium ions as necessary to achieve system charge neutrality. Results for trpcage miniprotein on both of these surfaces are also later included for comparison.[46] LK peptide structures were generated with the VMD psfgen plugin[47] and the SAM

surfaces were downloaded from the website of the Latour Research Group (http://www.clemson.edu/ces/latourlabs/Jmol/Surfaces.html). To match typical experimental conditions, both LK peptides were capped with a deprotonated carboxylate group, leading to overall charges of +5 and +6 for LKα14 and LKβ15, respectively.[12-19, 23] The SAM chains were spaced 5Å apart in an R3 (√3 x √3) geometry and at a tilt angle of 30° to the z-axis to mimic packing of the alkanethiol chains on a gold surface.[48] In contrast to the neutrally charged CH_3-terminated SAM, the COOH/COO-terminated SAM was made up of randomly alternating protonated and deprotonated chains to mimic a bulk pH of 7.4.[49]

Table 1.1. Setup of PTMetaD-WTE simulations.

System	Peptide	SAM	Peptide Charge	Surface Charge	Na^+	Cl^-	Approximate Number of Particles	Approximate Box Dims. (nm^3)
I	LKα14	COO/H	+5	-50	45	0	16,000	4x5x8
II	LKα14	COO/H	+5	-50	51	6	16,000	4x5x8
III[a]	LKα14	CH_3	+5	0	0	5	40,000	8x7x8
IV	LKβ15	COO/H	+6	-116	110	0	40,000	8x7x8
V	LKβ15	COO/H	+6	-116	116	6	40,000	8x7x8
VI[a]	LKβ15	CH_3	+6	0	0	6	40,000	8x7x8
VII	none	COO/H	n/a	-50	50	0	16,000	4x5x8
VIII	none	CH_3	n/a	0	1	1	16,000	4x5x8
IX[b]	Trpcage	COO/H	+1	-50	49	0	16,000	4x5x8
X[b]	Trpcage	CH_3	+1	0	0	1	16,000	4x5x7

[a]Data from Deighan et al.[29]
[b]Data from Levine et al.[46]

The GROMACS 4.6.5 MD engine[50] with the AMBER99SB-ILDN force field[51] was used in conjunction with the PLUMED 2.0 plugin[52]. Force field parameters for the head groups of the SAMs were taken from similar amino acid residues in the AMBER99SB-ILDN force field (i.e. CH_3 from leucine, COOH/COO from glutamic acid/glutamate). The length of the boxes in the z-dimension was chosen such that a peptide could diffuse far enough beyond the short-range van der Waals and Coulombic cutoff distances of 10 Å to allow approximation of a bulk-like state. A harmonic restraint was placed on the LK peptides' centers-of-mass to keep them from interacting with the bottom of the surface (i.e. at the top of the simulation box); the restraining force of this "wall" began acting on the peptides at a z-distance of 45 Å from the top frozen carbon atom on the surface (C10), or about 20 Å from the top of the simulation box. The different box lengths in the x and y dimensions for the LKα14 and LKβ15 simulations reflect the greatly different secondary structures of the two peptides. Preliminary LKβ15 simulations performed with the smaller box parameters showed the peptide's more extended conformations led to potential self-interactions across the periodic boundary. Conversely, initial testing did not reveal any such interactions in the smaller LKα14 system, allowing us to save computational costs (data from system III with LKα14 was taken from past simulations by Deighan et al.,[29] which used the larger box size). Extensive preliminary characterization was performed to ensure structure and thermodynamics were unaffected for LKα14 simulations in the smaller simulation cell. All simulations used periodic

boundary conditions in the three dimensions to allow for the calculation of electrostatic forces using the particle mesh Ewald (PME) method[53].

Energy minimization was performed on each simulation box using a steepest descent algorithm for 40,000 steps. The same system temperature and a global stochastic thermostat[54] was then used in a 1 ns NVT simulation with a time step of 2 fs, where the thiol group and first 10 CH_2 groups of the SAM chains were frozen to prevent diffusion or melting at high temperatures. The chains were frozen as such in all subsequent simulations, while the head groups remained unfrozen.

1.2.2 Enhanced Sampling

The method Parallel Tempering Metadynamics in the Well-Tempered Ensemble, or PTMetaD-WTE, is an enhanced sampling algorithm that can effectively overcome the challenges of simulating peptide/protein adsorption, such as timescale limitations or strong substrate/surface binding forces.[29, 55] An overview of the method will be presented here; the full protocol for employing this method is described elsewhere.[44, 55] Parallel tempering replica exchange[56], or PT, was employed to ensure the systems could overcome hidden energy barriers that likely exist at the interface. Twelve configurationally identical replicas spanning a temperature range of 300-450 K were simulated. The algorithm developed by Prakash and coworkers[57] was used to determine the temperature spacing that would lead to the best exchange rate between replicas. A short 1 ns NVT PT simulation was first performed to equilibrate each replica at the appropriate temperature. Following this simulation, a 10 ns well-tempered Metadynamics (WTM)[58] simulation was performed, biasing the overall potential energy of each replica to increase the energy fluctuations (while maintaining the same average values). This led to an increase in the potential energy overlap and thus exchange rate between adjacent replicas, reducing the round trip time (RTT) of any given replica and thus greatly enhancing the overall sampling efficiency. Ten ns was found to be the approximate amount of time needed for the systems' potential energy fluctuations to cease increasing at a noticeable rate; after this transient period, a nearly constant exchange rate between replicas of 30-35% was achieved via sampling the well-tempered ensemble (WTE)[44]. We note that, as in our prior examples, it is impossible to simulate with PT a temperature span of 300-450 K with only 12 replicas as zero overlap is achieved. A bias factor of 20 was used in all WTE setup simulations with LKα14 or with no peptide present, whereas the larger system sizes of the LKβ15 simulations used a bias factor of 50 to achieve similar exchange success probability (n.b., we desired to achieve similar RTT values in order to simplify comparisons between simulations and obtain similar sampling efficiencies and convergence times). Gaussian hills were added every ps with an initial height of 2.0 kJ/mol in all WTE setup simulations. Gaussian hill widths, or sigma values, of 200 and 320 kJ/mol were used for the LKα14/pure solution (i.e. water and ions) and LKβ15 simulations, respectively; the sigma values were determined by halving the natural fluctuations of each replica's potential energy from the lowest temperature of the NVT PT equilibration runs. As described in the SI, a new functionality in PLUMED2[52] was also used in this study to provide a modest improvement to the PTMetaD-WTE method.

Following the WTE setup step, two CVs were biased for each of the peptides with an additional two-dimensional MetaD bias potential. A common CV to both LKα14 and LKβ15 was the distance between the surface and the center-of-mass of the peptides, biased with a Gaussian hill width of 0.05 nm. The second CV for each of the peptides was chosen to maximize conformational sampling in the context of their different secondary structures. As with our

previous work on LKα14[29], the number of backbone α-helical hydrogen bond contacts was biased. Contacts were defined between oxygen and hydrogen atoms four residues apart within the peptide (i, i+4), and a switching function ranging from 0 to 1 with a reference bond length of 2.5 Å was used to define the degree of the contact. Previous simulations with LKβ15 biased the Φ and ψ backbone dihedral angles to measure the relative orientation of leucine and lysine residues[29]. Preliminary simulations performed for this study indicated sampling might be more effective by biasing the radius of gyration. We note that as the results in this paper generally only make use of the 1D free energy projections onto the distance CV, there is no issue comparing amongst all the simulations. The conformational CVs for LKα14 and LKβ15 were biased with a Gaussian hill width of 0.1 nm, respectively. For all MetaD CVs, a bias factor of 10 was used with Gaussian hills added every ps. An initial hill height deposition rate of 3.0 kJ/mol/ps was used for all simulations. Production simulations were run until convergence was reached, which will be discussed in the next section.

For systems VII and VIII where only solvent and counterions existed above the SAM surfaces, two separate simulations were run, one in which a single water molecule was biased and the other in which a single sodium ion was biased. In each of the two simulations, the only CV was the z-distance between the ion or water molecule and the surface.

Results and Discussion

1.3.1 Convergence of MetaD Simulations

Assessing convergence of the biased simulations is confounded by the absence of thermodynamic data to compare to as there is, to our knowledge, no quantitative data of free energies of adsorption for LKα14 or LKβ15 adsorbed to SAM surfaces. Instead, the metric used to assess convergence was the free energy difference between the adsorbed and solvated states, and when this value ceased changing with time at a noticeable rate, convergence was established. This is demonstrated in Figure 1.1, which shows the change in the Helmholtz binding energy as a function of simulation time per replica for systems I-X. Two lines each are shown for systems VII and VIII, which refer to separate simulations where a single ion (VIIa, VIIIa) or water molecule (VIIb, VIIIb) was biased. As expected, we see the systems converged on different timescales; systems I and II with LKα14 and IX and X with trpcage miniprotein[46] converged at least 100 ns faster than systems IV and V with LKβ15, most likely due to the differences in box sizes. We note that systems III and VI with LKα14 and LKβ15 on the CH_3-terminated SAM, respectively, were run by Deighan et al.[29] for much less time per replica than their counterparts here but were simulated with twice the number of replicas, thus achieving an approximately similar amount of sampling. For reference, systems VII and VIII with only solution above the surfaces were run for 150 ns each but converged in much less time than any of the peptide simulations.

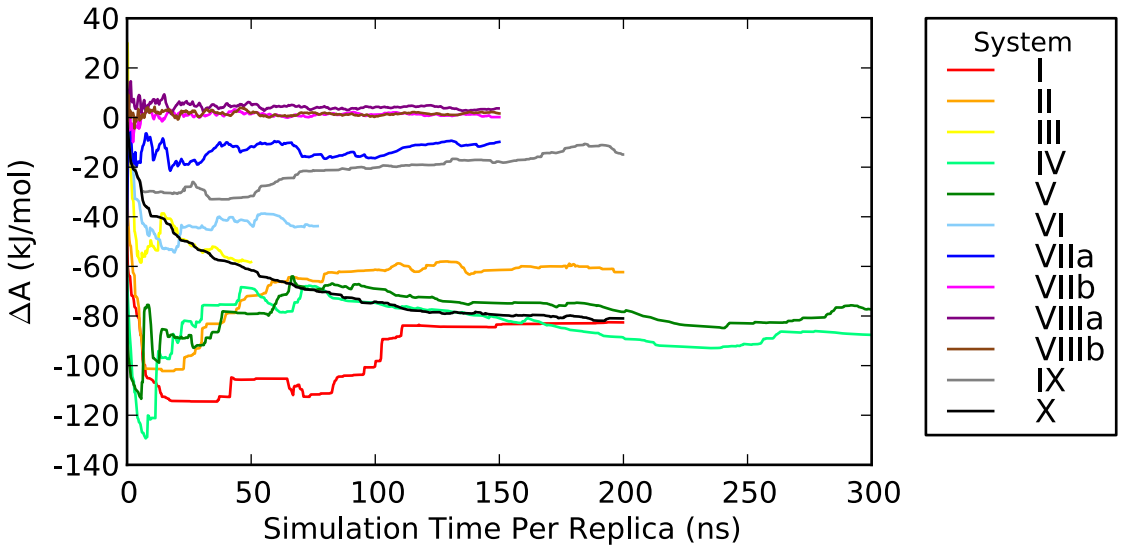

Figure 1.1. Convergence of free energy differences between solvated and adsorbed states for PTMetaD-WTE simulations at 300 K. Lines I-X correspond to systems I-X in Table 1.1. Lines VIIa/VIIb and VIIIa/VIIIb correspond to systems VII and VIII in Table 1.1, respectively, where a single ion/water molecule was biased.

1.3.2 Binding Free Energies

The use of high temperature simulations to extract valuable thermodynamic, structural, and kinetic information about a molecular system is quite common. High temperature data has frequently been used, for example, to analyze both the folding mechanisms and folding rates of proteins and peptides[59-67], information that is generally inaccessible from room temperature simulations. While often times only used as a means to accelerate sampling, high temperature data from tempering-based methods like PTMetaD-WTE has also been utilized, for instance, to analyze the stability of individual secondary structure motifs within a protein[68], as well as the thermodynamic properties of water across a wide temperature range.[57] The utility of high temperature simulations has led to many published studies that have specifically evaluated the accuracy of force fields in predicting properties from high temperature biomolecular simulations in comparison to experiments.[69-70] Following a stringent set of tests, a study published in 2012 found AMBER99SB-ILDN to be overall a robust force field for biomolecular simulations over a wide range of temperatures.[69]

We used the high temperature data to study the free energies of adsorption for all the simulated replicas. Figure 1.2 shows binding free energies calculated as a function of temperature for systems I (LKα14 on COOH/COO), III (LKα14 on CH_3), IV (LKβ15 on COOH/COO), VI (LKβ15 on CH_3), IX (trpcage on COOH/COO) and X (trpcage on CH_3). When comparing the results of the six simulations, a trend quickly becomes apparent: the three lines with increasing slopes, indicating weaker surface binding with increasing temperature, all correspond to systems with the neutral, hydrophobic CH_3 SAM, and conversely the three negatively-trending lines that indicate stronger binding with increasing temperature correspond to systems with the charged,

hydrophilic COOH/COO SAM. Furthermore, the two sets of lines (i.e. increasing or decreasing slopes) show roughly the same rate of change in binding free energy with increasing temperature. We note that for the majority of the lines in Figure 1.2, all of these trends could be similarly realized from analyzing the data over a much narrower range in temperature, e.g., by just considering the first few replicas or so. We found the results from the MetaD simulations indicating increasing binding strength with increasing temperature to be particularly surprising.

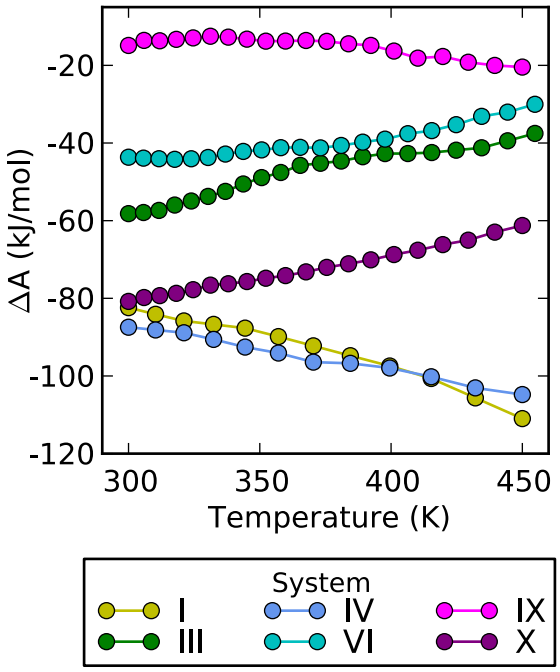

Figure 1.2. Binding free energy change upon adsorption to SAM surfaces as a function of simulation temperature for systems I, III, IV, VI, IX, and X. Each circle represents a different replica in the system.

Since LKα14, LKβ15, and trpcage all show the same effect, it is clear that this effect is not attributable to the peptides. Insight can be gained from looking at the constituent energetic and entropic components of the binding free energy as a function of temperature. The Helmholtz energy is the relevant thermodynamic quantity for the canonical ensemble,

$$\Delta A_{ads} = \Delta U_{ads} - T\Delta S_{ads} \tag{1.1}$$

where ΔA_{abs} is the change in the Helmholtz energy of the system upon binding, ΔU_{abs} is the change in internal energy upon binding, T is the replica temperature, and ΔS_{abs} is the entropy change upon binding. Assuming ΔU_{abs} is independent of temperature, according to Equation 1 the slope of the lines in Figure 1.1 can be used to determine $\Delta S_{abs.}$ This approach has been successfully used by Bussi et al.[43] to extract changes in enthalpy and entropy upon folding of a β hairpin in explicit water, with good agreement to the Go model and to experiments, and by Zhai et al.[71] to calculate

the change in entropy of two states of a water nonamer. We used our simulation results to ensure that the binding potential energy is independent of temperature for the adsorption of LKα14 to the COOH/COO SAM as a test case. This is described in the SI, and the results (Figures A1.1 and A1.2) show no dominant temperature trend exists, thus validating the key assumption underlying the analysis of our results with Equation 1.

As the slopes trend the same for all three simulations with the CH_3 SAM regardless of the peptide present, and similarly for the three simulations with the COOH/COO SAM, we conclude that the entropic contribution to the binding free energy is dominated by the choice of surface. Equation 1 also indicates that ΔU_{abs} corresponds to the to y-intercepts of the six lines in Figure 1.2. Since each of the different peptide/surface combinations results in a unique y-intercept, we can conclude that the energetic contribution to the binding free energy is synergistically dominated by the specific sidechain/surface interactions at play. Recent experiments by Deshapriya and Kumar of enzymes adsorbing to inorganic surfaces validate the finding that positively charged sidechains at the interface are a major driving force for binding.[72] Structures, orientations, and conformations of the LK peptides on the SAM surfaces were previously investigated in detail by Deighan et al.[29]

The result that the binding free energy can strengthen with increasing temperature is not intuitive and calls for a deeper analysis. Visual inspection of the simulations, combined with the fact that the chosen CVs for the six systems only bias the peptide degrees of freedom (i.e. peptide distance from the surface and an internal conformational descriptor), suggested that solution effects (i.e. water or ion surface binding) might play an important role.

1.3.3 Breakdown of the Thermodynamic Cycle of Adsorption

We proposed an overall thermodynamic cycle that takes into account both peptide and ion binding effects in Figure 1.3. In writing the equations for each step in Figure 1.3, we assumed that water would contribute negligibly to the binding thermodynamics. To test the validity of the mechanism and its assumptions, additional MetaD simulations were required. We chose LKα14 adsorbed to the COOH/COO SAM as the first test case.

Mechanism Validation with LKα14. Upon closer inspection of the simulations showing the peculiar temperature behavior, we saw that the systems and thus the binding profiles shown in Figure 1.2 were actually representative of the peptide adsorption process (step B), rather than of the full thermodynamic cycle (step A). This is due to the common practice of charge neutralization without introduction of excess ions: with a surface charge of -50 and 45 positively charged ions bound to surface sites, LKα14 with a charge of +5 should be able to adsorb to the surface without displacing ions. In light of this, system II was designed with excess sodium ions to cover all 50 surface sites prior to adsorption (with chlorine ions added to maintain charge neutrality), which should represent the true step A of the thermodynamic cycle. The same CVs were biased and parameters used as in system I. Additionally, step C is represented by system VII, where in two separate MetaD simulations a single sodium ion (VIIa) and water molecule (VIIb) where biased on and off the COOH/COO SAM surface. Calculated binding free energies for these four separate MetaD runs are shown in Figure 1.4. The binding free energies of system VIIa have been multiplied by -1 to represent desorption of the ions from the surface.

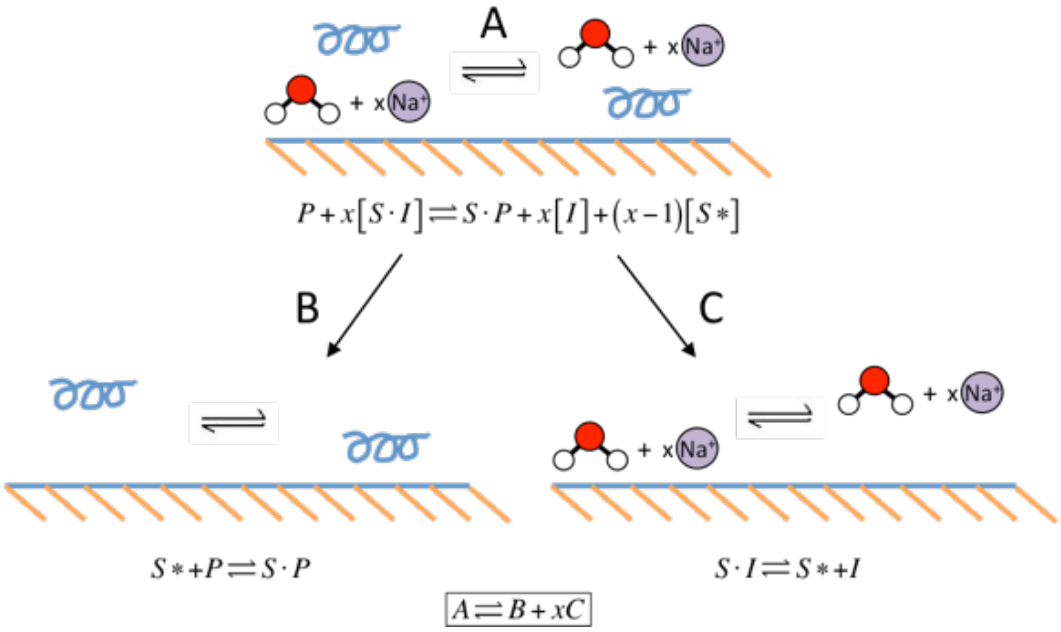

Figure 1.3. Proposed breakdown of the thermodynamic cycle of the peptide/surface adsorption process. Steps A, B, and C represent the overall thermodynamic cycle, peptide adsorption step, and ion desorption step, respectively.

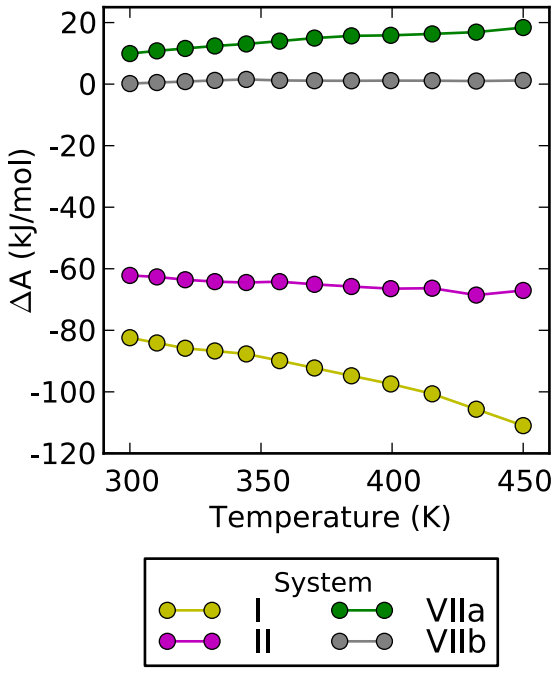

Figure 1.4. Binding free energies calculated for systems I, II, and VIIa/b, corresponding to steps B, A, and C, respectively, of the thermodynamic cycle of adsorption (c.f. Figure 1.3). Binding free energies have been multiplied by -1 for system VIIa to represent ion desorption from the surface.

Figure 1.4 shows that, as expected, the binding free energy of a single water molecule leaving a water bath to adsorb to the carboxylic SAM is essentially 0 kJ/mol and does not change with increasing temperature, validating this original assumption. This is unsurprising, as there is an equal and opposite potential of mean force for a water molecule desorbing into bulk solvent. Results in Figure 1.4 also lend support to the idea that the sodium ions play an important role, as they have an affinity for the charged, COOH/COO surface that is strengthened as the temperature is increased (multiplying the system VIIa values by -1 gives the adsorption values rather than the plotted desorption values). In other words, similar to when peptides adsorb on the charged surface, it is more entropically favorable for the sodium ions to be bound to the charged surface than floating in the bulk. As will be discussed later, the increase in entropy upon adsorption of the peptides and ions to the charged surface is largely due to configurational entropy gains upon release of bound water molecules. Conversely, the step C results for CH_3 (see Figure A1.3 in the SI), show neither water nor sodium ions have an affinity for the surface at any temperature, implying that for a hydrophobic, methyl-terminated SAM step C does not contribute to the mechanism, and thus that steps A and B are energetically the same. Clearly, this is not the case for the COOH/COO SAM in Figure 1.4, which shows very different binding profiles for systems I and II, i.e. steps B and A, respectively. Furthermore, simply adding the results of systems I and VIIa does not give the results of system II, and so a reweighting analysis is needed to determine the value of *x*, or the number of ions desorbed in the steps listed in Figure 1.3. The idea that counter ion release contributes significantly to biomolecular adsorption to a charged surface is well known in the literature, e.g., counter ion release has been shown to drive interactions between proteins and DNA[73], as well as between peptides and charged lipid bilayers;[74] however, as we show below, understanding these contributions on a deeper level allows for greater control over the molecular driving forces in adsorption and thus represents a new step towards rational design of the biomolecule/surface interface.

Reweighting Analysis for the Number of Desorbed Ions. The histogram-based reweighting algorithm by Bonomi et al.[75] was used to generate a 2D free energy surface of the number of ions in solution (defined as at least 4 Å beyond the surface) as a function of the distance between LKα14 and the COOH/COO surface. The ensemble average of the number of ions in solution was then calculated for both the surface-bound and solution states of the peptide. We note that for all reweighting analyses we performed, the portion of the trajectory that corresponds to the MetaD transient was excluded from the analysis. The differences are shown in Figure 1.5 as the number of ions displaced or desorbed from the surface upon peptide adsorption, as a function of temperature. It was originally anticipated that this would only be needed for the MetaD simulation of step A. However, upon performing the reweighting, it was found that it was necessary to also correct step B binding free energies for the hindrance of having to displace ions upon adsorption. This is most likely because the five free surface sites LKa14 should need to adsorb are not always next to each other on the surface. Figure 1.5 shows the average number of ions that are displaced from the surface during peptide adsorption for the different replica temperatures. An average value of 2 that is roughly independent of temperature is found for the peptide adsorption step, whereas the direct simulation of the thermodynamic cycle shows a leveling off in the number of desorbed ions at a value of about 3.5. For the rest of the analysis, the value of *x* is held constant at 3.5 for step A and 2 for step B; this reduces the noise to make it easier to visualize the final results, but does not change the trends or conclusions (see Figure A1.4 for the temperature-dependent results).

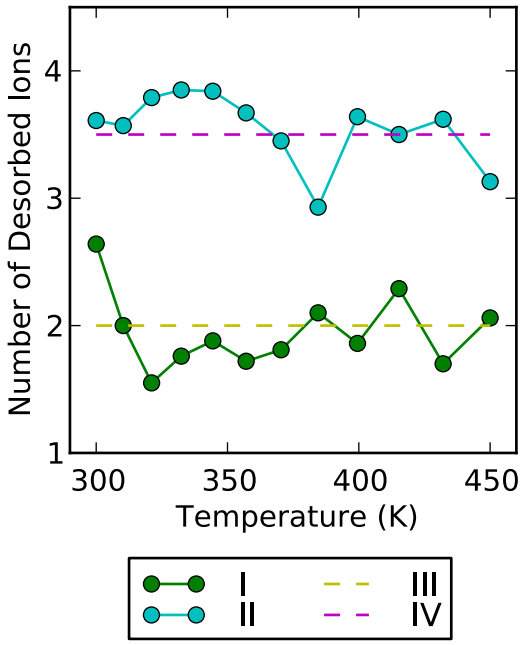

Figure 1.5. Reweighted number of ions displaced from the COOH/COO SAM during LKα14 adsorption as a function of temperature. Lines I and II refer to the peptide adsorption step (step B) and the overall thermodynamic cycle (step A), respectively. Dashed lines refer to the average number of desorbed ions used in each case for later analysis.

Multiplying the value of the binding free energy at each temperature for a single ion (system VIIa in Figure 1.4) by 2 and subtracting these numbers from the binding free energy values for system I in Figure 1.4 gives the binding free energies for the true peptide adsorption step, or step B, of the thermodynamic cycle. Figure 1.6 shows these values, along with the step C results multiplied by x=3.5, and the step A results from simulating system II. The exciting result is seen with the blue line, which represents the addition of the yellow and green lines, or the step B binding free energies plus 3.5 times the step C binding free energies. Thus, the blue line signifies the calculated step A result and can be compared directly to the result from the simulation of step A with system II (i.e. the purple line); the strong agreement between the two lines gives confidence in the mechanism we proposed earlier, and in a broader sense, that the results of separate MetaD simulations can be combined and compared in this manner. The results indicate that peptide/ion binding effects on a negatively charged surface are complementary, i.e. that upon adsorption the peptide displaces ions to preserve charge neutrality at the interface. Supporting these findings, Deshapriya and Kumar found maintaining charge neutralization at the nanobio interface with a charged solid to be a major driving force for adsorption.[72] For further validation of our proposed adsorption mechanism, we look at the case of LKβ15 binding to the COOH/COO SAM.

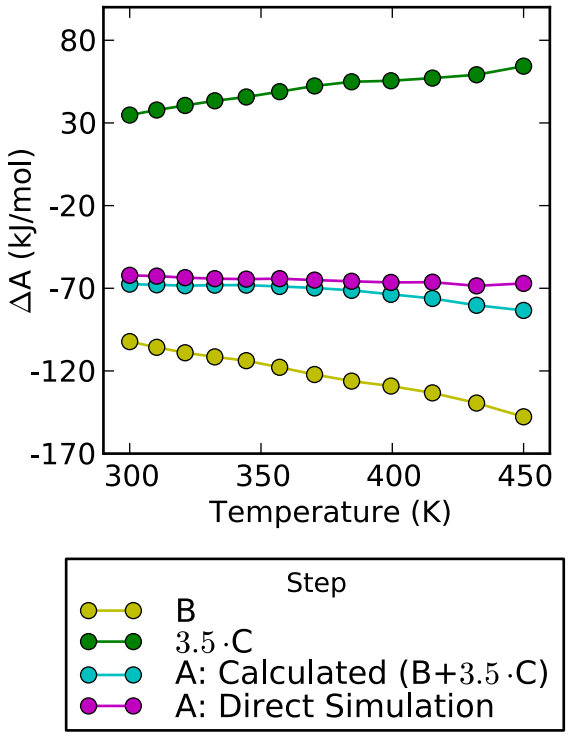

Figure 1.6. Calculated binding free energies for the three steps of the thermodynamic cycle for LKα14 on COOH/COO. Yellow and green lines represent steps B (system I, corrected for ion desorption) and 3.5·C (system VIIa) of the mechanism proposed in Figure 1.3, respectively, whereas purple and blue lines represent the results from the direct simulation of step A (system II) and the calculated results for step A based on addition of the yellow and green lines (B+3.5·C).

Additional Validation with LKβ15. Analogous to Figure 1.5 for LKα14, Figure 1.7 shows the average number of ions displaced from the surface during the adsorption of LKβ15 as a function of temperature. Line I corresponds to the ion desorption results for the peptide adsorption step or system IV in Table 1.1, and line II corresponds to the results for the overall thermodynamic cycle step, or system V. Dashed lines represent the constant number of desorbing ions chosen to correct the step B binding free energies (line III) and represent the value of x in the mechanism of Figure 1.3 (line IV). Figure 1.8 shows the final binding free energy results versus temperature for the system of LKβ15 adsorbing on the COOH/COO SAM. Lines represent the ion-corrected binding free energies for the peptide adsorption step, or step B, of the thermodynamic cycle, as well as the step C results multiplied by x=4, the step A results from simulating system V, and the step A results from adding the step B and 4 times the step C results. Temperature-dependent results (i.e. results calculated with non-constant x values) are shown in Figure A1.5 in the SI. Comparison of the simulated and calculated step A results shows near perfect agreement; considered next to the similar conclusions drawn from the LKα14 on COOH/COO analysis, this leads to a degree of certainty in the proposed adsorption mechanism of Figure 1.3, as least for systems in which electrostatics dominate the interfacial binding energies.

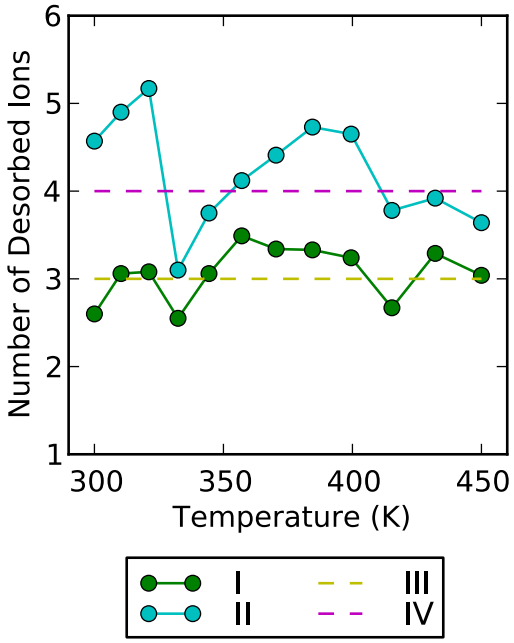

Figure 1.7. Reweighted number of ions displaced from the COOH/COO SAM during LKβ15 adsorption as a function of temperature. Lines I and II refer to the peptide adsorption step (step B) and the overall thermodynamic cycle (step A), respectively. Dashed lines refer to the average number of ions used in each case for later analysis.

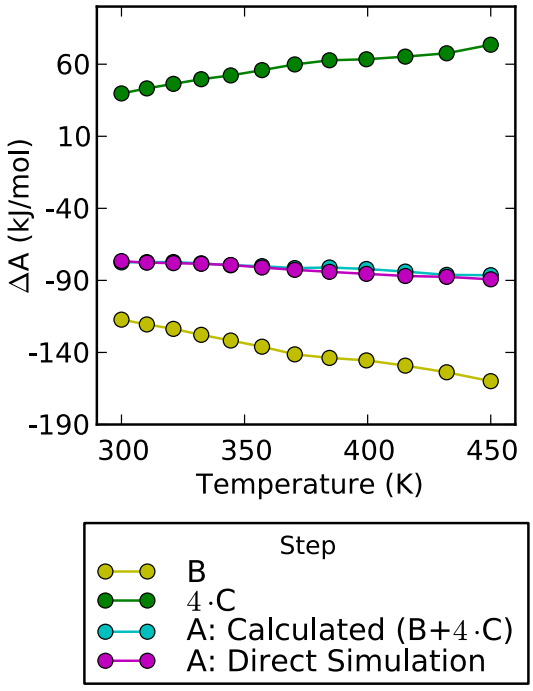

Figure 1.8. Calculated binding free energies for the three steps of the thermodynamic cycle for LKβ15 on COOH/COO. Yellow and green lines represent steps B (system IV, corrected for ion desorption) and 4·C (system VIIa) of the mechanism proposed in Figure 1.3, respectively, whereas purple and blue lines represent the results from the direct simulation of step A (V) and the calculated results for step A based on adding the yellow and green lines (B+4·C).

To reiterate, this analysis was found to be unnecessary for systems with the neutral, hydrophobic CH_3 SAM. This was due to the fact that neither the sodium ions nor water molecules showed any affinity towards the surface, implying the energy of step C is zero and thus that steps A and B are energetically equivalent. The question still remains as to why in the peptide adsorption step the peptides bind less strongly with increasing temperature to the neutral, hydrophobic SAM and more strongly with increasing temperature to the charged, hydrophilic SAM, which is addressed in the next section.

1.3.4 Understanding the Origin of the Sign Change in ΔS

The negative slope in binging free energy versus temperature seen in Figure 1.2 for the peptides on the COOH/COO surface indicates, according to Equation 1, that the change in entropy upon binding is greater than zero, i.e. it is more entropically favorable for the peptides to be on the surface than in the bulk. The opposite trend holds true for systems with the CH_3 surface, that it is more favorable for the peptides to be in the bulk solution than adsorbed to the surface, meaning the change in entropy upon binding is negative. To explain why the entropy of peptides adsorbed to the charged, hydrophilic SAM might overcome the rotational and translational entropic losses that occur on adsorption, we tested whether extended conformations that might increase configurational entropy could be stabilized on the charged, hydrophilic SAM due to strong peptide/surface charge interactions. This was accomplished by reweighting the end-to-end distances between the backbone carbon centers-of-mass of the first and last residues of LKβ15 adsorbed on each surface at both 300 K and ~380 K (384.5 K on the COOH/COO SAM and 381.0 K on the CH_3 SAM). The same reweighting algorithm that was previously used to calculate the number of surface-bound ions was used here to calculate end-to-end distances for surface-bound states.[75] The 2D free energy surfaces are shown below in Figure 1.9, where the top plots correspond to LKβ15 on the COOH/COO SAM surface, and the bottom plots to LKβ15 on the CH_3 SAM surface. Compared to identical solution phase behavior, the ensemble of states on the surface at 300 K clearly shows that the charged, hydrophilic surface stabilizes an extended conformation of LKβ15 (Figure 1.9A, ~3.3 nm end-to-end distance) compared to the neutral, hydrophobic surface, which shows between a 1.5-2 kT increase in free energy upon extension in the adsorbed state to the same end-to-end distance (Figure 1.9C). The multi-state behavior indicated by Figure 1.9A at 300 K is maintained at the higher temperature of 384.5 K (Figure 1.9B), though the extended state appears to be slightly less stable. Similar behavior is observed for the high temperature results on the CH_3 surface (Figure 1.9D) compared to the results at 300 K (Figure 1.9C). Experiments by York et al. investigating the influence of ionic strength on LK peptide adsorption to hydrophilic and hydrophobic surfaces[17] found comparable results: with a charged, hydrophilic surface, the authors found that at high ionic strength, an N-H mode appeared in the SFG spectra that was not present at low ionic strength, and that this mode only appeared when the peptide had a stable secondary structure. In other words, the authors found that at low ionic strength, which correlates to step B in our proposed adsorption mechanism (Figure 1.3), the peptide did not have a stable secondary structure on the charged surface, indicating an extended conformation as seen in Figure 1.9. Also in line with our findings, the authors found no change in the spectra with changes in the solution ionic strength with a neutral, hydrophobic surface present.

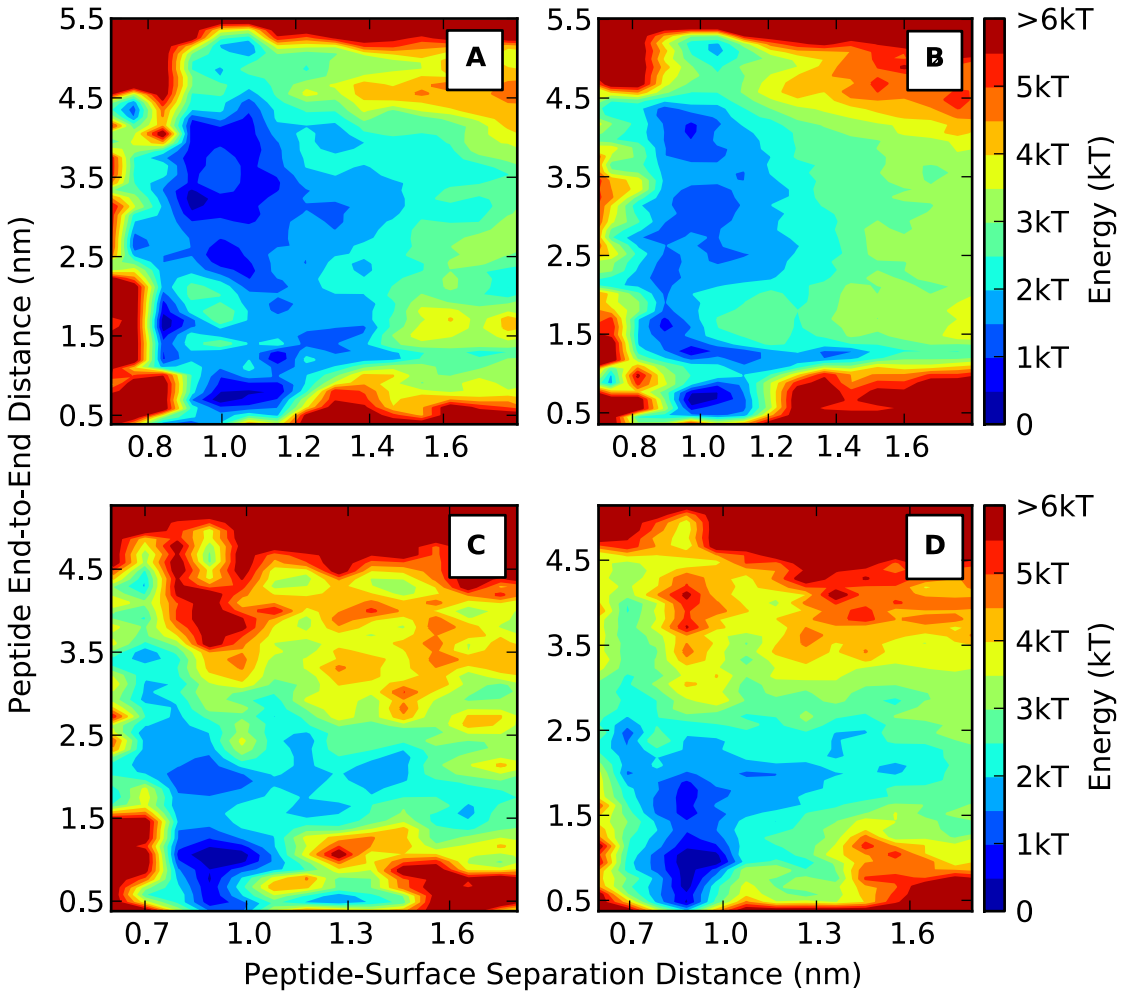

Figure 1.9. 2D free energy profile of the end-to-end distance for LKβ15 versus the separation distance between LKβ15 and A) the COOH/COO-terminated SAM surface (system IV) at 300 K, B) the COOH/COO-terminated SAM surface at 384.5 K, C) the CH_3-terminated SAM surface (system VI) at 300 K, and D) the CH_3-terminated SAM surface at 381 K. Phase space is restricted to include only surface-bound states. Black boxes indicate areas for clustering analyses (each box indicates a separately performed clustering analysis, labeled by the number next to it).

Despite the interesting discovery that the charged surface stabilizes a broader range of peptide configurations, this finding alone is not sufficient to explain the overall increase in entropy that occurs upon adsorption to the charged, hydrophilic surface as seen in Figure 1.2; the possibility that the charged surface only selects a specific set of extended conformations upon adsorption needs to be considered. To examine this, we estimated the configurational entropic effects of adsorption on the two surfaces. Employing protocols similar to those used in our recent past work[76], all surface-bound structures of LKβ15 within kT of the respective energy minima of the surfaces, as well as the solution state structures of system IV, were separately clustered and subsequently reweighted to attain unbiased cluster probabilities. The change in configurational entropy of the peptide upon adsorption to the two surfaces was then estimated via Equation 2:

$$\Delta S_{ads,config} = -k_B \left[\sum p_i \ln p_i \right]_{ads-solv} \quad (1.2)$$

where k_B is Boltzmann's Constant and p_i is the probability of each individual cluster. The results of the analysis (see Table 1.2) indicate the entropic penalty is 1.2 kJ/mol•K larger upon adsorption to the neutral surface than to the charged surface, which translates to a notable difference in entropy of 360 kJ/mol at 300 K (n.b. while we did not explicitly calculate translational or rotational losses upon adsorption, we expect their contributions to the overall loss in entropy upon adsorption to be fairly equal for the two surfaces).

Table 1.2. Clustering analysis results for surface-bound structures of LKβ15.

Surface	# Structures for Clustering Analysis	# Clusters from Clustering Analysis	Configurational Entropy (kJ/mol-K)	$\Delta S_{ads,\ config}$ (kJ/mol-K)
CH_3	48,619	219	6.4	-4.0
COOH/COO	59,419	183	7.6	-2.8
None (solution)	26,864	241	10.4	

However, the fact that the configurational entropy decreases upon adsorption to either surface fails to explain the opposite signs in the overall system adsorption entropy observed for the surfaces in Figure 1.2. Thus, we concluded that solvent entropy effects must play a larger role in determining the peptide-surface binding affinities, which is in line with the work of others.[77] Ten ns, unbiased MD simulations were performed for systems VII and VIII with solvent above the charged and neutral surfaces, respectively, as well as a simulation with triple the amount of sodium (and chlorine) ions above the charged surface. Unidirectional density plots for the water above the surfaces were calculated over the last 5 ns of each simulation (Figure A1.6). The results show nearly a 40% higher density of water above the charged surface than above the neutral surface, which has a density scarcely above that of the bulk water, as well as a second well-defined water layer that exists only above the charged surface. Additionally, the first water layer formed above the surfaces is able to penetrate about 1Å closer to the charged surface than to the neutral surface, indicating a more tightly bound water layer. All of these observations lead to the conclusion that upon peptide adsorption and subsequent release of water to the bulk, the configurational entropy gains would be significantly greater on the charged surface. We believe this result, combined with the results from the peptide configurational entropy analysis, explains the previously ambiguous trends of increased entropy upon binding to the charged surface (and thus the trend of increasing ΔA_{ads} with increasing *T)* and decreased entropy upon binding to the neutral surface.

To summarize: 1) for the charged surface, the large entropic gains incurred from release of bound water molecules upon peptide adsorption heavily dominate over the entropic losses of the peptide itself upon adsorption to the surface, leading to an overall positive change in system adsorption entropy (ΔS_{ads}); and 2) the slight or even negligible entropic gains incurred from release of bound water molecules upon peptide adsorption to the neutral hydrophobic surface are not enough to overcome the configurational entropy losses of the peptide upon adsorption (which we found are notably larger than on the charged surface), leading to an overall negative change in

ΔS_{ads}. Furthermore, Figure A1.6 shows that the addition of excess ions to the simulation interrupts surface/water interactions and significantly decreases the density of the water layers above the charged surface, eliminating the second water layer altogether. This implies less configurational entropy would be gained upon release of bound water molecules during peptide adsorption, explaining the results of Figures 1.6 and 1.8 for systems II and V with excess ions, respectively, where ΔS_{ads}=0, as a balance between water entropy gains and peptide entropy losses.

Conclusions

The enhanced sampling method PTMetaD-WTE was employed to simulate the adsorption of two model peptides, LKα14 and LKβ15, on two self-assembled monolayer (SAM) surfaces, a model hydrophilic SAM consisted of a charged carboxylate/carboxylic acid-terminated head group, and a model hydrophobic SAM consisted of a neutral methyl-terminated head group[29]. Simulations of trpcage miniprotein were included for comparison.[46] Calculated binding free energies plotted as a function of temperature showed all three peptides bound more weakly to the neutral, CH3 surface with increasing temperature and more strongly to the charged, COOH/COO surface with increasing temperature. These results led us to conclude that entropic and energetic driving forces are most likely driven by the choice of surface and by the specific sidechain/surface interactions at play, respectively. A mechanism for the overall thermodynamic cycle of adsorption was proposed to attempt to explain these differences, which required additional MetaD simulations. Reweighting analyses for the average number of ions displaced from the surface during peptide adsorption showed that ions play a crucial role in the thermodynamic cycle of adsorption when the charged, hydrophilic surface is present, but make no contribution to the system's binding thermodynamics when the neutral, hydrophobic surface is present. It was demonstrated for LKα14 and LKβ15 on the COOH/COO SAM that at any temperature, the binding free energy of the peptide adsorbing to the surface (simulation 1) can be directly added to the total binding free energy of the ions being displaced from the surface (simulation 2) to give the overall thermodynamic cycle binding free energies versus temperature (simulation 3). In other words, entropic or configurational effects between ions and peptides are complementary with charged, hydrophilic surfaces, but are not with neutral, hydrophobic surfaces. The idea that free energies from separate MetaD simulations can be combined in a manner similar to classical analyses of complex thermodynamic cycles is an exciting idea, but requires further validation when applied to other types of systems, particularly to protein adsorption, which is a significantly larger challenge than peptide adsorption. Evidently, however, when constructing model systems like those studied here, simulators need to take great care; it is common practice to add only enough counterions to achieve charge neutrality, for instance, but the results of this work show the ion concentration can heavily influence important thermodynamic aspects of a system.

Further analysis on the end-to-end peptide distance of surface-bound states revealed that upon adsorption, the charged surface stabilized extended peptide conformations compared to the neutral surface. Clustering and reweighting techniques were used to estimate the configurational entropy of a peptide bound to either surface, providing a useful way to obtain more from biased simulations. This analysis demonstrated that despite the stabilization of extended states on the charged surface, upon adsorption to either surface the configurational entropy was reduced compared to in the bulk. A solvent density analysis then showed that the trend of increasing peptide

binding strength with increasing temperature on the charged surface could be explained by large entropic gains of bound solvent molecules released upon peptide adsorption dominating over peptide configurational entropy losses. The opposite conclusion was found to explain the trends upon adsorption to the neutral surface. This study thus provides a means to gain valuable atomic level insight about macroscopic ideas, such as the hydrophobic effect as it applies to protein adsorption. An interesting question for future inquiry is whether decreasing the charge of the hydrophilic COOH/COO surface to be neutral leads to a decrease in entropy upon binding as seen with the neutral, hydrophobic CH3 surface, which would indicate electrostatic interactions dominate over hydrophobic interactions.

Perhaps the most important take away from this study is that molecular simulations can provide valuable information about the adsorption process that is not readily measureable or in some cases even possible to measure in the lab. For instance, experiments of protein adsorption are nearly always performed with a water-based salt solution such as PBS to preserve protein conformation and functionality; a study by Xia et al. showed drying caused a significant loss of functionality and altered the conformation of adsorbed proteins.[78] Enhanced sampling simulations circumvent these issues, allowing the free energy of a peptide adsorbing to a surface in the absence of salts to be determined, as we inadvertently investigated with the three peptides adsorbing to the COOH/COO SAM. Tuning the peptide/surface interactions for this step of the thermodynamic cycle of adsorption that is normally hidden from experiments could lead to important strides in the rational design of the biomolecule/surface interface.

Chapter 2

Essential Slow Degrees of Freedom in Protein-Surface Simulations: A Metadynamics Investigation[2]

Introduction

The free energy is an important thermodynamic quantity in biomolecular systems, containing information about relative stability between states, binding affinities, and solubilities[79]. Several enhanced sampling methods can provide reliable estimates of free energies of specific reaction coordinates or free energy differences between stable states from molecular simulations of the system. Recently, the metadynamics family of methods has gained widespread use in such calculations. It allows users to bias several slow degrees of freedom (or collective variables (CVs)) using a history-dependent, dynamic bias. Small biases (usually Gaussian-shaped hills) are deposited in CV phase space to force the system to visit new states, ultimately converging to the underlying free energy surface[80]. Several flavors of metadynamics have since been developed to overcome challenges related to convergence and scaling[81]. Some variations scale the Gaussian height[58] or width[82] according to its location in phase space to improve convergence. Others introduce parallel replicas in phase[83] and temperature[29, 43] space to increase sampling. And finally, some alter the way bias is deposited – by exchanging between replicas[84] or by depositing many low-dimensional bias potentials in parallel[85]– to tackle problems associated with biasing multiple CVs. (For an exhaustive review of metadynamics, the authors direct the readers to some excellent reviews[81, 86]).

Some of these metadynamics methods have been used to calculate free energies of binding of molecules to proteins[87-90] and to surfaces[29, 37, 40, 91-98], especially where strong binding prohibits exhaustive sampling with plain MD. For example, Hughes *et al.* used well-tempered metadynamics (WTM) to exhaustively compute the binding free energies of all amino acids to graphene[99]. Several studies have used parallel tempering metadynamics in the well-tempered

[2]Reproduced in part with permission from K.G. Sprenger, Arushi Prakash, and J. Pfaendtner. Essential slow degrees of freedom in protein-surface simulations: A metadynamics investigation. Biochemical and Biophysical Research Communications, In Press. Copyright 2016 Elsevier.

ensemble (PTMetaD-WTE) to characterize binding[29, 45-46]. Others have used replica exchange with solute tempering augmented with metadynamics (RESTMetaD) to calculate the binding of peptides to various gold crystal facets[92] and to silica[97-98]. Like other dimension-reducing algorithms, the efficacy of metadynamics is influenced by the choice of CV(s). The abovementioned studies used the center-of-mass distance of the molecule from the surface[29, 46, 91-92, 98], helicity of the peptide[97], hydrogen bonds[45-46], or the hydrophobicity of the protein core[46] as CVs. These CVs bias either intrinsic coordinates of the solute or the solute-surface distance.

However, commonly used CVs in biomolecular adsorption systems do not explicitly consider environmental variables that may affect binding, such as solvation[100-101], surface ion content[91], and other local order variables[102]. In fact, Wu et al. showed that ions in the local environment can affect the conformation of adsorbed peptides, stabilize different binding residues, and compete with peptide adsorption[103]. Therefore, it is essential to understand and account for the effect of ions, either from the electrolyte or naturally occurring surface ions in metal oxides like silica, when studying biomolecular adsorption.

To address remaining gaps in knowledge about the best practices for simulating peptide adsorption in relation to charged interfaces and ions in solution, herein we study the interaction of model peptide GGKGG with a silica surface at pH 7.5. Mimicking neutral pH conditions, several surface sites are deprotonated and passivated by sodium ions. We compare the resulting peptide binding free energy profiles of three different variants of metadynamics– (1) well-tempered metadynamics, (2) parallel-tempering metadynamics in the well-tempered ensemble, and (3) parallel-bias metadynamics. We compare the sampling achieved by each of these methods and the results of using two different CVs, namely the center-of-mass and sidechain distance of the peptide from the surface. Finally, we compare peptide-surface binding in the presence of different electrolyte species. From our results, we hope to establish a general protocol that would allow researchers to comprehensively explore conformational phase space and obtain true energetics of inorganic surface-binding peptides. We especially stress thorough exploration of the role of ions in biomolecular adsorption.

Methods

2.2.1 Simulation Setup

All simulations, listed in Table 2.1, were conducted with the GROMACS 5.1.2[104] MD engine. The surface and peptide were modelled with the INTERFACE[105] and CHARMM36[106] force fields, respectively, and the SPC[107] model was used to represent water. The peptide, GGKGG, was generated with the psfgen plugin for Visual Molecular Dynamics (VMD[47]). It was capped with neutral ACE and NME groups to warrant the study of a single binding residue (i.e., positively charged lysine) to silica. To approximate the correct surface ionization state and terminal chemistry for a silica surface at pH 7.5, the structure provided by Emami *et al.*[108] was used. Each simulation box consisted of a ~ 3.5 x 3.4 x 2.0 nm silica unit. This unit was placed next to 3.5 x 3.4 x 4.6 nm of solution containing a single peptide, water, counterions, and electrolyte (optional) (Table 2.1). For production runs, bulk atoms in the silica structure were frozen to prevent extensive deformation of the structure during the simulations. However, atoms within 0.5 nm of the surface

remained free to capture the true interaction of silica with other species[94]. In each simulation, the peptide was placed near the surface in a random starting configuration. One Cl^- ion was added to each simulation box to neutralize the peptide. In some cases, electrolytes in the form of $Ca^{2+}(Cl^-)_2$, Na^+Cl^-, and $(Na^{0.5+})_2Cl^-$ were added to the solution (modelled with the CHARMM36 force field) to model varied experimental conditions. In the case of $(Na^{0.5+})_2Cl^-$, the charges of all sodium ions (including surface ions) were scaled to 0.5+ for consistency, creating a hypothetical micro-environment of like-charged counterions. The system was periodic in x, y, and z dimensions which allowed the peptide, water and ions to interact with both the top and bottom surfaces of silica.

All starting configurations were stabilized using a steepest descent method to minimize the energy of the system. The systems were then equilibrated in the NPT ensemble at 1 bar and 300 K for 1 ns using the Berendsen barostat[109] and Donadio-Bussi-Parrinello thermostat[54]. Finally, production runs were conducted in the NVT ensemble at 300 K, using the same thermostat, for varying simulation times as noted in Table 2.1. All simulations were performed with a 2 fs time step with bonds between hydrogen and heavy atoms constrained using the LINCS[110] algorithm. Short-range Coulomb and Van der Waals forces were explicitly calculated below a cutoff value of 1 nm. Particle Mesh Ewald (PME)[111] was used for calculating electrostatics with the short-range term cutoff at a value of 1 nm.

2.2.2 Biasing Methods

Following NPT equilibration, different enhanced sampling schemes were employed to test their relative efficacy in determining the binding free energy of the peptide to the silica surface. Brief descriptions and relevant parameters for the three chosen flavors of metadynamics are given below. Complete protocols for employing these methods can be found elsewhere (WTM[58], PTMetaD-WTE[44, 55], PBMetaD[85]).

Based on previous sampling procedures for using distance CVs, the initial Gaussian height was set to 2 kJ/mol, the bias factor to 10, and the Gaussian deposition rate to 1 hill/ps for all metadynamics simulations[91]. The σ value was calculated from an unbiased simulation of the system; the peptide-surface distance was tracked with PLUMED[52] and half the standard deviation of the equilibrium fluctuations was used as an estimate[112]. The value used for the simulations was 0.05 nm. During all metadynamics simulations, a harmonic restraint was placed on the distance of the peptide's center-of-mass (COM) from the surface (at ~ 2.2 nm from the top of the simulation box) to promote sampling in the region of interest and use computational time more efficiently. Free energy profiles recovered from the metadynamics simulations were considered converged when they retained their shape over the last 25% of the simulation; while the majority of the simulations converged prior to 1 μs of total sampling time, all simulations were nonetheless carried out to 1 μs for consistency.

2.2.3 Well-Tempered Metadynamics (WTM)

Well-tempered metadynamics (WTM) simulations were performed by biasing either the orthogonal distance between the COM of all peptide atoms and the surface, or the distance between only the sidechain (SC) nitrogen atom of lysine and the surface.

A simple reweighting scheme[113] was used to recover the free energy profile of the SC-surface distance from a simulation where the COM was biased. The weight of each configuration was given by

$$e^{\beta(V(s,t)-c(t))} \tag{2.1}$$

where *V(s,t)* is the instantaneous value of the history-dependent bias potential for the system (*s* is the value of the CV at a given time *t*), and $\beta = 1/k_BT$, where k_B is the Boltzmann constant and T is temperature. The reweighting factor, c(t), is a time-dependent offset bias that was calculated on-the-fly using PLUMED[52] via the following equations,

$$c(t) = \frac{1}{\beta} log \frac{\int ds\, e^{-\beta F(s)}}{\int ds\, e^{-\beta(F(s)+V(s,t))}} \tag{2.2}$$

$$\beta F(s) = -\frac{\gamma V(s,t)}{k_B \Delta T} + log \int ds\, e^{\gamma V(s,t)/k_B \Delta T} \tag{2.3}$$

where ΔT is the so-called hills temperature for the WTM simulation, $\gamma = \frac{T+\Delta T}{T}$ is the bias factor, and Δt is the time between energy depositions of Gaussian hills. These weights were binned according to the SC-surface distance and the free energy was then obtained by taking the natural log of the sum of weights in each bin.

2.2.4 Parallel Tempering Metadynamics in the Well-Tempered Ensemble (PTMetaD-WTE)

Parallel tempering metadynamics in the well-tempered ensemble (PTMetaD-WTE) was employed to assess the use of temperature for overcoming hidden free energy barriers introduced by the presence of an interface. Six configurationally identical replicas were simulated across a temperature range of 300-450 K, spaced per the algorithm developed by Prakash et al.[57] In a 100 ps NVT parallel tempering simulation, the replicas were equilibrated at their respective temperatures with the Donadio-Bussi-Parrinello thermostat[54]. A 10 ns WTM simulation was then performed to establish the well-tempered ensemble (WTE), biasing the potential energy of the system with a bias factor of 30. Gaussian hills were added to the potential energy every ps with a σ value of 150 kJ/mol and an initial height of 2.0 kJ/mol. After 10 ns, a constant exchange success probability of 21% was achieved between replicas.

Following the WTE setup, a second metadynamics simulation was performed with an additional one-dimensional bias potential on the orthogonal distance between the COM of all the atoms in the peptide and the silica surface. As in our previous study[91], the WTE bias continued to be applied to the potential energy, though with a significantly reduced deposition pace of every 5 ps. Reweighting for the free energy as a function of the SC-surface distance was obtained as previously described for WTM simulations.

2.2.5 Parallel Bias Metadynamics (PBMetaD)

Recently, PBMetaD was shown to be a flexible and easy method for sampling multi-dimensional free energy surfaces[85]. Instead of depositing a multi-dimensional bias (which scales poorly with the number of CVs), PBMetaD relies on depositing several low-dimensional biases in parallel in multiple CV spaces.

A 6-replica multiple walker PBMetaD simulation was conducted[83, 85]. The distances of all ions (independently) and the COM of the peptide from the surface were biased, using the same parameters as previously described.

We used the technique of Tiwary *et al.*[113], to make an estimate of the unbiased probability distributions (free energies) of other CVs in PBMetaD (i.e., the SC-surface distance). This resulted in using the same approach described above to calculate weights using the c(t) quantity described above. As a technical note, the current version of PLUMED does not apply the reweighting method within PBMetaD, so we estimated (c(t)) as:

$$e^{\beta c(t)} = \int ds \left[e^{\frac{\gamma V(s,t+\Delta t)}{k\Delta T}} - e^{\frac{\gamma V(s,t)}{k\Delta T}} \right] \tag{2.4}$$

where the various components and variables were described in an earlier section.

Table 2.1. List of metadynamics parameters and collective variables used for different simulations of GGKGG on silica at pH 7.5.

Excess Electrolyte	# of Cations (Water)	Simulation Method	Collective Variable(s)	Total Simulation Time[b]
None[a]	20 (1705)	WTM	SSD (Protein COM)	1,000 ns (1W)
		WTM	SSD (Protein SC)	1,000 ns (1W)
		PBMetaD	SSD (Protein COM/Ions)	1,000 ns (3W)
		PTMetaD-WTE	SSD (Protein COM)	1,000 ns (6W)
$Ca^{2+}(Cl^-)_2$	45 (1575)	WTM	SSD (Protein COM)	1,000 ns (1W)
		PBMetaD	SSD (Protein COM/Ions)	1,000 ns (6W)
Na^+Cl^-	45 (1600)	WTM	SSD (Protein COM)	1,000 ns (1W)
		PBMetaD	SSD (Protein COM/Ions)	1,000 ns (6W)
$(Na^{0.5+})_2Cl^-$	46 (1652)	WTM	SSD (Protein COM)	1,000 ns (1W)
		PBMetaD	SSD (Protein COM/Ions)	1,000 ns (6W)

[a]Refers to silica-neutralizing sodium ions (Na^+ ions) modeled with the INTERFACE force field.
[b]Values in parentheses are the number of identical walkers (W) or replicas used in each simulation.

Results and Discussion

2.3.1 Choice of Collective Variable

The distance between the peptide and the surface can be biased by using the distance calculated from (a) the COM (a function of all atoms of the peptide, in our study), or (b) a specific sidechain atom. Here, both biasing schemes are implemented with WTM to compare their convergence speeds and ability to predict features of the peptide binding free energy profile. The binding free energy was calculated as the difference between the Boltzmann-averaged energy of the adsorbed (0.9 to 1.1 nm) and solvated (2 to 3 nm) states. Here, convergence was defined as the point at which (1) the free energy difference between states stabilized to within 2.5 kJ/mol (k_BT at 300 K) (Figure 2.1, top right), and (2) minimal changes were observed in the free energy profile of the CV over the last 25% of the simulation. For simulations where the COM was biased (scheme (a)), the free energy profile along the SC-surface distance was calculated by reweighting (described in the Methods section).

We first compare the convergence of the SC and COM biasing schemes with WTM (Figure 2.1; top right). The results show the COM bias converges significantly faster than the SC bias. Indeed, the binding free energy of scheme (a) converges to within k_BT of the final value of ~ -8.6 kJ/mol (averaged across the three methods) nearly 400 ns before scheme (b).

In fact, even after 1 μs of total sampling the free energy profiles obtained from both schemes are markedly different (Figure 2.1; top left). While the binding free energy is identical for both sampling schemes, the minima have different features. Two preferred binding modes of the peptide can be resolved using the COM bias – (A) between z-distances of 0.9 and 1.0 nm, and (B) between 1.0 and 1.1 nm. Representative snapshots from the simulation (Figure 2.1; bottom) show that minimum A comprises structures that are flattened on the surface with multiple peptide/surface interaction points. However, minimum B comprises structures with one binding site only. These features are lost by biasing only the SC-surface distance; a single minimum is resolved where the peptide most commonly binds with one or two residues but overall remains at a distance from the surface (minimum C).

From the differences in the binding free energy profiles, it is evident that the profile obtained from biasing the COM incorporates more information about the adsorption process. It can differentiate between binding poses. Importantly, it incorporates changes to the whole peptide structure since the COM bias is a function of the coordinates of all peptide atoms. On biasing the sidechain atom(s) only, we lose information about the binding behavior of neighboring residues that might affect the overall surface-bound conformations of the peptide. We hypothesize that these differences might be more appreciable for solutes with multiple binding residues. These findings suggest that metadynamics users should use CVs that apply bias to all atoms of the solute. On converging the free energy profile (and bias potential) of the original CV, the free energy profile of the true CV of interest can then be obtained via reweighting techniques.

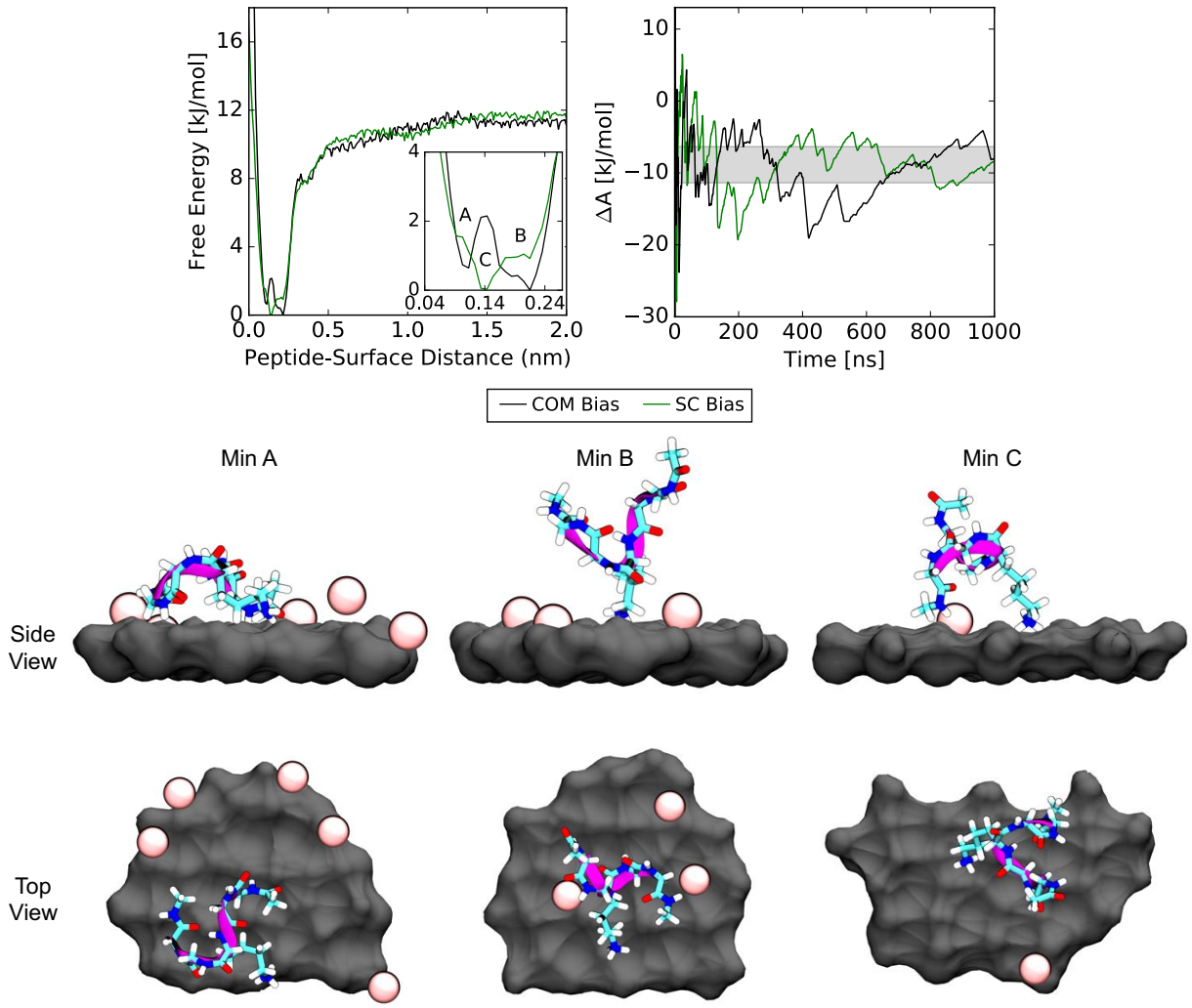

Figure 2.1. (top left) Free energy projected onto the distance between the sidechain (SC) nitrogen atom of lysine and silica for the two biasing schemes: bias added to all atoms in the peptide through the center-of-mass (COM), or bias applied to just the SC nitrogen atom. The black line is the result of reweighting. (top right) Change in Helmholtz energy between the solvated and adsorbed state of GGKGG on silica for the two different biasing schemes. A gray box highlights the "convergence region" of +/- k_BT of the average of the final ΔA values of the three methods of ~ -8.6 kJ/mol. (bottom) Snapshots from each of the labeled minima in the top left plot. The surface and sodium ions are shown in gray and pink, respectively, and are restricted to within 1 nm of the peptide. The peptide backbone is shown in purple, and carbon, hydrogen, oxygen, and nitrogen atoms are shown in cyan, white, red, and blue, respectively. Water is not pictured for clarity.

2.3.2 Comparison of Sampling Schemes

From the above simulations, it was concluded that applying bias to all atoms in a biomolecule results in faster calculations and provides greater information about the system. Another key issue that metadynamics users face is selecting the appropriate biasing scheme to calculate a variable of interest. For the calculation of peptide binding free energy, we compare

three biasing methods - (1) WTM, (2) PTMetaD-WTE, and (3) PBMetaD. WTM is a relatively simple metadynamics sampling scheme where applying bias to more than 2 or 3 CVs becomes computationally expensive. In the PTMetaD-WTE framework, sampling is enhanced by a parallel tempering scheme that allows the system to sample several temperatures (by biasing the potential energy). However, on increasing the number of CVs beyond 2 or 3, this method also becomes computationally prohibitive as the scheme for building the bias potential is identical to WTM. In contrast, in the PBMetaD biasing scheme low-dimensional biases are deposited in parallel, instead of higher dimensional biases, making computations more efficient and allowing for the possibility of biasing many CVs. Thus, with PBMetaD the distance of each ion from the surface can be biased in addition to the peptide-surface distance. The free energy profile recovered from PBMetaD simulations is considered to be the most accurate among the three methods since it accounts for the most degrees of freedom in the system. The peptide COM distance from the surface is biased in all schemes, and the SC distance is reweighted for as described earlier.

The results (Figure 2.2) show that within the resolution of the calculations (i.e., σ = 0.05 nm and k_BT = 2.5 kJ/mol), PBMetaD and WTM recover nearly identical free energy profiles with two free energy minima (A and B) resolved for the adsorbed state. These minima correspond to the minima and snapshots shown in Figure 2.1 for the WTM simulation.

However, the PTMetaD-WTE simulation does not recover the minimum closest to the surface, minimum A. We hypothesize that this discrepancy is either due to: 1) the behavior of ions at high temperature, since PTMetaD-WTE allows the system to visit temperatures from 300-450 K, or 2) incomplete sampling. In regards to the former hypothesis, it has been predicted that the binding affinity of sodium ions to a hydrophilic surface increases with increasing temperature[91]. Therefore, we expect most replicas at high temperatures to sample configurations with ions bound to the surface, rather than free in solution. As noted before, PTMetaD-WTE relies on the exchange of these high temperature configurations so that they eventually reach the lowest-temperature replica at 300 K. However, at 300 K, these ions are tightly bound beyond k_BT (Figure 2.3A), and the impact of the temperature alone on sampling the ion degrees of freedom is unclear[16]. Thus, it is possible that the 300 K trajectory primarily samples a surface with charged sites predominantly bound to ions and the peptide cannot adsorb at the free energy minimum located 0.1 nm from the surface (minimum A). Essentially, we propose that the temperature benefits of the enhanced sampling algorithm are reduced in this case. Examining the behavior of the MetaD transient throughout the PBMetaD simulation (results not shown), does show that the location of the minima fluctuates modestly in terms of location and energetics, albeit within the standard deviation of the CV (0.1 nm) and thermal noise (2.5 kJ/mol). In light of these results, we set out to explore the role of ions, ion binding, and the capability of a sampling algorithm to effectively tackle various ion-specific effects.

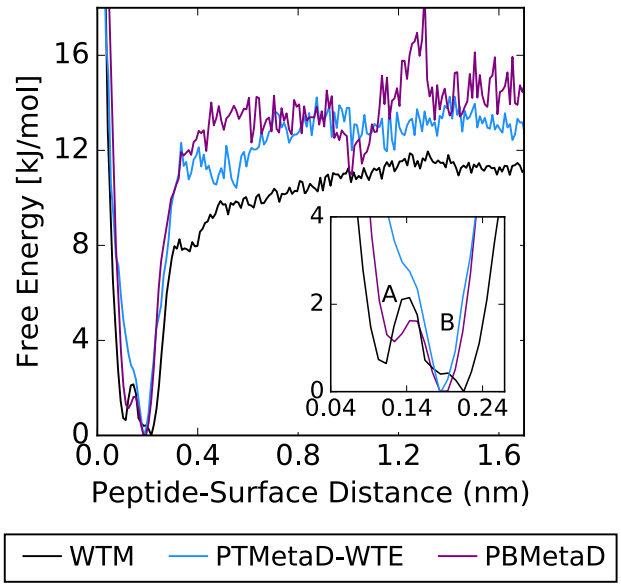

Figure 2.2. Free energy (kJ/mol) along the distance of the nitrogen atom of the lysine sidechain from the surface using enhanced sampling method WTM (black), PTMetaD-WTE (blue), and PBMetaD (purple).

2.3.3 Ion-Surface Distance as a Collective Variable

The simulations discussed above were carried out at low ionic strengths. However, many proteins adsorb in conditions of high ionic strength where ions can promote aggregation or "salting-out" of proteins[114] or screening of protein-protein interactions or protein-surface interactions [115]. To understand the influence of ions on the adsorption behavior of our model peptide, additional metadynamics simulations were performed with different electrolytic conditions (i.e., monovalent and divalent cations). Simulations were carried out with additional cations (25 Ca^{2+}, 25 Na^{+}, or 26 $Na^{0.5+}$) and with requisite numbers of anions (Cl^{-}) to maintain system charge neutrality (Table 2.1). In all simulations, the added electrolytes were initially evenly distributed in the solution and allowed to adsorb spontaneously onto the surface. Both WTM and PBMetaD simulations were performed in each case, and the results (Figure 2.3) are discussed in detail below.

Protein Adsorption Without Excess Ions

Without excess ions in the solution, the previous simulations represented the limit of a dilute electrolytic solution in contact with a silica surface. In those cases, the WTM and PBMetaD peptide binding free energies were shown to be identical (Figure 2.2, Figure 2.3A). Additionally, Figure 2.3A shows the binding free energy of the sodium ions, calculated from the PBMetaD simulation, was equal to that of the peptide. Thus, it is likely the ions competed with the peptides for binding sites; however, in the limit of a dilute electrolytic solution, 1 μs of total simulation time was found to be more than enough to converge to the correct peptide binding free energy profile with either method.

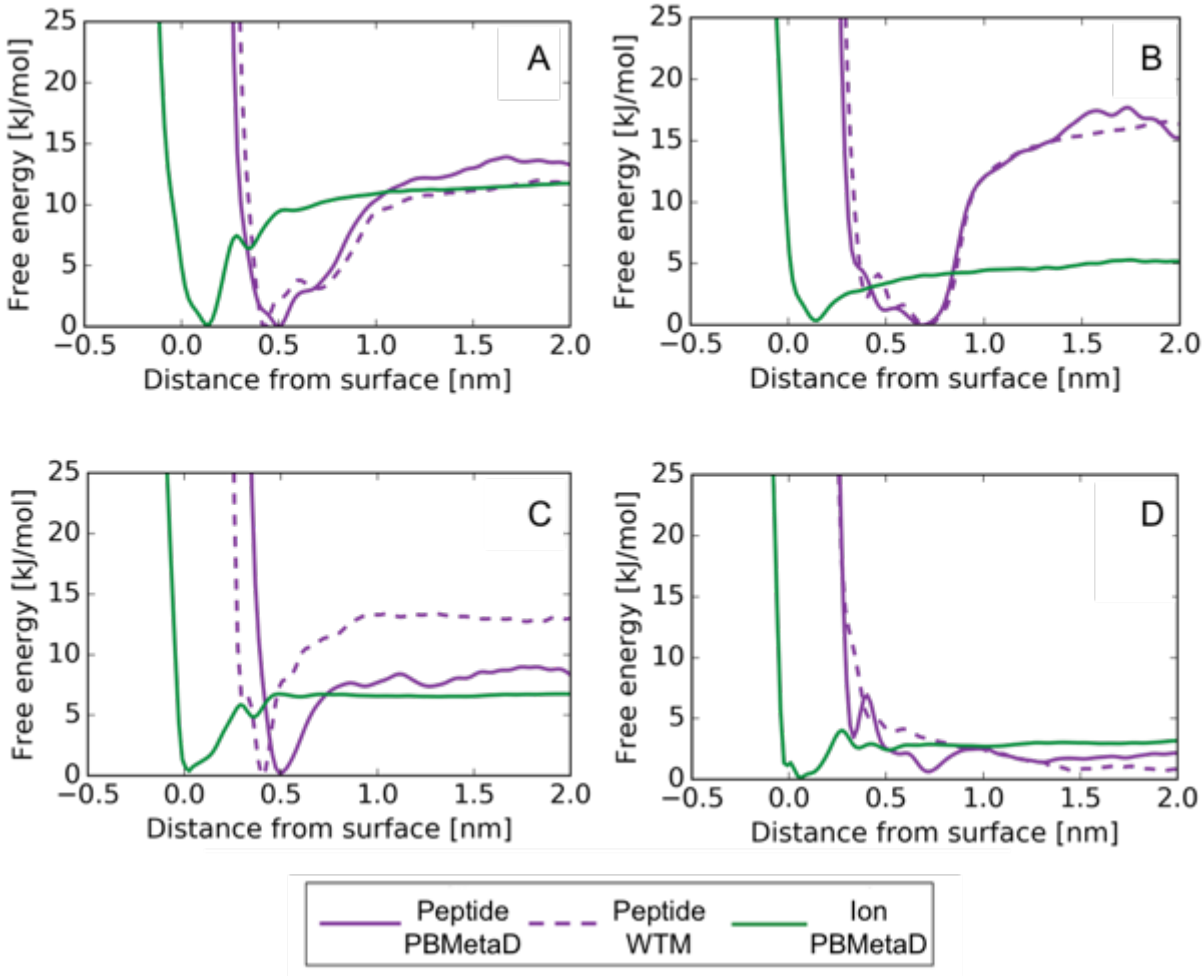

Figure 2.3. Free energy projected onto the COM-surface distance with: (A) no electrolyte, (B) excess $Na^{0.5+}$ ions, (C) excess Na^{+} ions, and (D) excess Ca^{2+} ions. Green lines indicate thermally-averaged ion binding profiles from PBMetaD simulations. Solid and dotted purple lines indicate peptide binding profiles from PBMetaD and WTM simulations, respectively.

Effect of Excess $Na^{0.5+}$ Ions on Peptide Adsorption

On investigating peptide adsorption in the presence of excess $Na^{0.5+}$ ions from the PBMetaD simulation (Figure 2.3B), it is evident that the peptide binds much stronger to the surface (~ 20 kJ/mol) compared to the ions (~ 5 kJ/mol). This is a departure from the binding free energy predicted in previous simulations that did not have excess ions (~ 13 kJ/mol for both peptide and ions), but had the requisite monovalent sodium ions to neutralize the surface. When $Na^{0.5+}$ ions adsorb to the surface, they cannot completely compensate for the -1 charge of the surface sites. Thus, an ion adsorbed on a surface site fails to completely screen the charge of the surface, promoting other solutes that can fully compensate for the charge, like the +1-charged lysine sidechain, to adsorb in its place. Additionally, the difference in charge (+0.5 versus +1) provides the lysine sidechain with a greater electrostatic driving force to adsorb on a surface site. These

factors lead to an increase in the binding free energy of the peptide in the presence of $Na^{0.5+}$ ions compared to in the presence of Na^{+} ions.

Notably, both WTM and PBMetaD simulations again converge to the same free energy profile for the peptide SC-surface distance, even though ions were not explicitly biased in the WTM simulation. This results from the higher binding free energy of the peptide to the surface compared to the ions. Thus, a WTM simulation biasing a single CV to study peptide adsorption does not face sampling limitations from surface-bound ions. From these findings, we conclude that more intensive sampling methods like PBMetaD and PTMetaD-WTE are not required in this case to calculate the binding free energy of the peptide.

Effect of Excess Ca^{2+} Ions on Peptide Adsorption

With 25 additional Ca^{2+} ions and 50 Cl^{-} ions added to neutralize the system, this represents an electrolyte of high ionic strength with cations that have twice the charge of the peptide's lysine sidechain. The results of the PBMetaD simulation (Figure 2.3D) show that a driving force (~ 4 kJ/mol) exists for dissolved Ca^{2+} to adsorb to the surface. Conversely, both WTM and PBMetaD simulations predict that as the peptide approaches the surface, the peptide is repelled from it and thus prefers to remain ~0.7 nm away from the surface. Since Ca^{2+} ions preferentially bind to the surface adsorption sites and more than compensates for the surface charge of each site, the protein effectively feels a positive surface charge and is repelled. Visual analysis of the trajectory in VMD[47] confirms this; whenever the protein approaches the surface to bind (mainly driven by the influence of the bias potential), it is promptly repelled and remains in solution (Figure 2.4). We note that while a small free energy minimum does appear for the PBMetaD simulation at a peptide-surface distance of ~0.4 nm, the peptide must cross a free energy barrier to fall into this minimum, presumably to displace the bound Ca^{2+}. We recommend future work to investigate this effect, perhaps performing simulations to explore the surface charge effects (e.g., number of charge sites and magnitude of charges) as well as extending the study to peptides with multiple charged side chains to promote different binding mechanisms and thermodynamics.

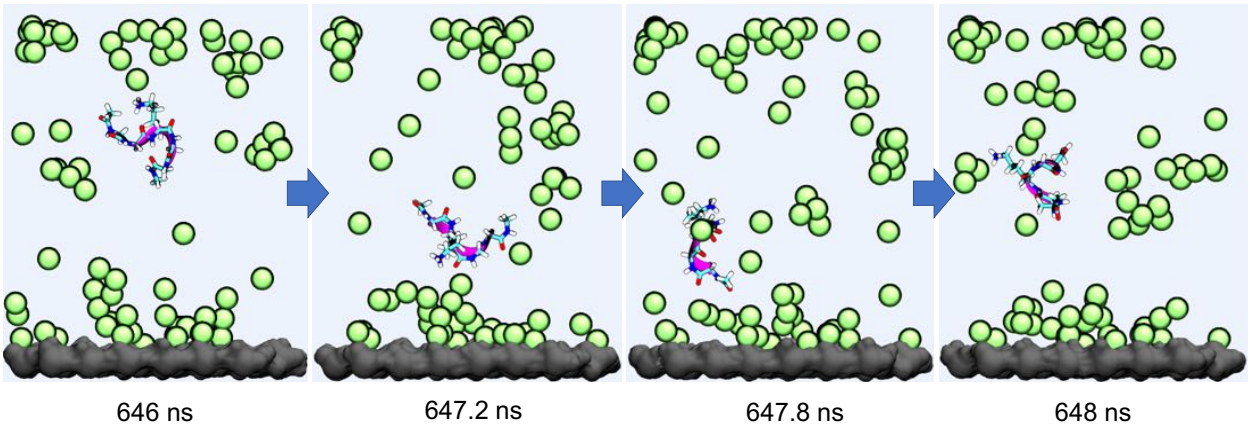

Figure 2.4. Snapshots from a WTM simulation with added electrolyte $Ca^{2+}(Cl^{-})_2$. The surface and peptide are colored as previously described (Figure 2.1). Ca^{2+} ions are shown in green. Explicit water molecules and Cl^{-} ions are not pictured for clarity.

Effect of Excess Na^+ Ions on Peptide Adsorption

Finally, in the case where the electrolyte contains excess Na^+ ions, a difference of ~ $2k_BT$ is observed in the peptide binding free energies of the WTM and PBMetaD simulations (Figure 2.3C). This is noteworthy, given that WTM and PBMetaD converged to the same peptide binding free energy in the absence of excess Na^+ ions (Figure 2.3A). In both cases (i.e., with or without excess Na^+ ions), the binding free energy of the peptide and ions are predicted to be the same with PBMetaD.

The excess Na^+ ions provide a crowded microenvironment where the peptide diffuses slowly in solution and ions quickly replace other ions on the surface. Consequently, when an ion leaves a binding site, it is easier for another ion in the vicinity to take its place. Once an ion is bound to a surface site, it is a "rare event" for it to desorb spontaneously (in a 300 K simulation) and allow the peptide to adsorb in its place. Thus, in situations where an electrolyte and biomolecule have similar surface affinities, WTM simulations, which only provide energy to the degrees of freedom of the solute, are severely limited by the availability of surface binding sites when the peptide approaches the surface.

On the other hand, in PBMetaD bias is also provided to the ions to adsorb/desorb multiple times during the simulation. This leads to many opportunities for the peptide to adsorb to free surface sites and thus for the system to collect information about solute binding/unbinding. This is also evident in Figure 2.5, where we track the position of the lysine sidechain with respect to the surface. Although the WTM simulation (Figure 2.5; left) samples a few binding events, the peptide spends most of the time in solution. It draws near the surface several times, but likely due to inaccessible surface sites, quickly moves back into solution. In the PBMetaD simulation (Figure 2.5; right), the peptide spends time on the surface and in solution almost equally. Additionally, most approaches of the peptide to the surface result in a successful binding event. Thus, the use of PBMetaD results in many more binding/unbinding events collected during the simulation and converges much more rapidly than WTM. We hypothesize that the WTM simulation would require several more microseconds to sample enough peptide binding/unbinding events to converge to the same free energy profile as PBMetaD.

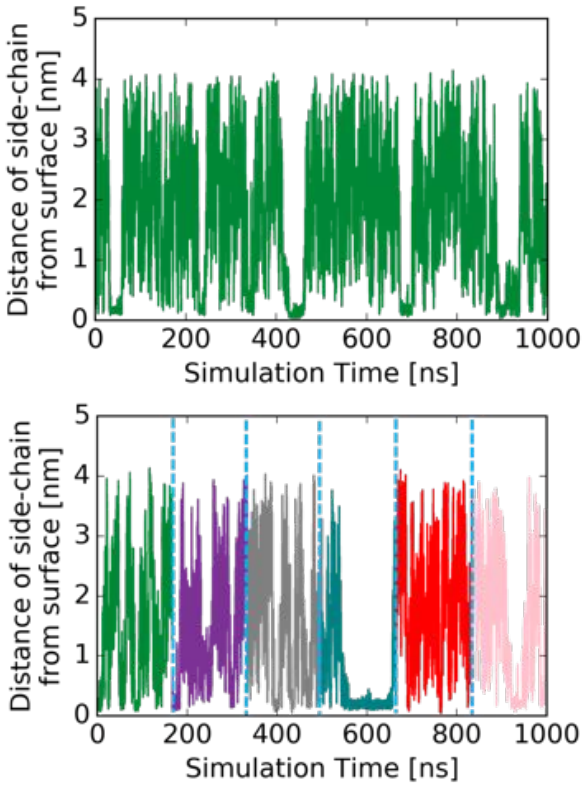

Figure 2.5. Distance of the sidechain nitrogen of lysine from the surface, as sampled by (left) WTM and (top) PBMetaD (bottom). Each walker in the PBMetaD simulation is colored separately, with the x-axis denoting cumulative simulation time. Each simulation starts from independent starting structures and evolves sharing the same bias potential for 166.67 ns each.

Conclusions

Three different metadynamics methods, namely WTM, PTMetaD-WTE, and PBMetaD, were employed to simulate the adsorption of model peptide GGKGG to silica. These methods were evaluated for their ability to converge to the "correct" peptide binding free energy profile in the presence of various electrolytic solutions. Here, PBMetaD was expected to provide the most accurate results due to its capacity to bias both solute (peptide) and solvent (ion) degrees of freedom. It was determined that competing slow degrees of freedom from solvent constituents can, in some cases, have a large impact on the resulting binding free energy of the solute. Thus, it is important to consider how surface-bound and electrolytic ions affect biomolecular adsorption. Indeed, there are several cases where ions dictate how proteins interact with surfaces, for example in the polyamine-directed precipitation of silica nanospheres that only occurs in the presence of multivalent anions such as phosphate [116].

Based on our findings, we recommend the following biasing protocols to recover "correct" binding free energy profiles when simulating biomolecular adsorption processes:

1) Dilute solutions – A simple metadynamics sampling scheme like WTM can be used. While competition for surface sites may exist between the solute and ions, the dilute electrolytic conditions allow for many solute binding/unbinding events to occur throughout the simulation. Thus, convergence to the correct solute binding free energy profile can be achieved in a shorter simulation timescale (within 1 μs).
2) Excess, competing electrolyte - It is necessary to use a method like PBMetaD that can bias both the solute and ion degrees of freedom. If the binding affinities of the constituents are not known *a priori*, it can be roughly assumed that ions with a charge equal to the charge of binding sidechain residues will have a similar binding affinity to the surface as the solute. We strongly emphasize the need to explicitly bias the ion degrees of freedom in these situations; it was determined that a method like PTMetaD-WTE that only indirectly biases the solvent degrees of freedom may result in an incorrect binding free energy profile of the solute.
3) Excess, weak-binding electrolyte - It is sufficient to use a sampling scheme like WTM. The protein sidechain can displace the surface ions upon adsorption, allowing for adequate sampling of binding/unbinding events to converge to the correct binding free energy profile of the solute.
4) Excess, strong-binding electrolyte – Since the ions bind considerably stronger, it leads to extremely long sampling times required to obtain the correct solute binding free energy. Thus, we recommend using a method like PBMetaD to speed up the time to convergence, and a larger bias factor for CVs.
5) Choice of solute CV - Conformational CVs should be chosen that apply bias to all atoms of the solute to obtain quicker convergence and maximum information about the solute conformation. On converging this free energy profile, the free energy along a different reaction coordinate can be calculated through simple reweighting procedures.

Chapter 3

Probing How Defects in Self-Assembled Monolayers Affect Peptide Adsorption with Molecular Simulation[3]

Introduction

The formation and characterization of self-assembled monolayers (SAMs) on solid surfaces has been extensively studied for several decades. The easy preparation of SAMs with different terminal chemical functionalities has made them convenient for far-reaching and numerous applications, including bio-related technologies such as biosensors and medical implants, nano- and microfabrication, nanodevices, and corrosion protection. Experimental microscopy studies have long shown that SAMs have high concentrations of defects[117-119]; in some cases, as with the nanofabrication method of microcontact printing, naturally-occurring imperfections in the SAMs were shown to play a beneficial role in the process[120]. In most cases, however, defects in the monolayers can have unexpected and perhaps undesirable consequences. Two commonly occurring defects arise from imperfections in the substrate (leading to increased surface roughness after self-assembly) and imperfections in the self-assembly process (i.e., so-called film defects).

Though molecular simulation can offer unique insights into the consequences of SAM structural imperfections, it has only rarely been done[121-125]; limitations of small simulation cell sizes and/or insufficient sampling times have prevented the explicit exploration of defects in typical SAM modeling studies.[120] We have employed the enhanced sampling method parallel tempering metadynamics using the well-tempered ensemble (PTMetaD-WTE), which we have used successfully in several prior studies to study peptide/protein adsorption at interfaces.[29, 46, 55] A description of other simulation approaches to studying these types of problems can be found elsewhere.[29]

In this work we build on our prior simulations[29] of the model peptide LKα14[11] adsorbing onto an ideal SAM. Past work focused on obtaining structural and thermodynamic information of

[3]Reproduced in part with permission from K.G. Sprenger, Y. He, and J. Pfaendtner. Probing How Defects in Self-Assembled Monolayers Affect Peptide Adsorption with Molecular Simulation. Foundations of Molecular Modeling and Simulation, 21-35, 2016. Copyright 2016 Springer Singapore.

adsorbed peptides, with a specific emphasis on quantitative comparison to experimental measurements of sidechain orientation. However, the systems studied were very idealized due to their lack of SAM structural imperfections. In this work, we take the logical next step by studying the impact of incorporating surface defects, and provide new insights into the consequences of SAM imperfections on the structure and binding thermodynamics of adsorbed biomolecules. Herein we have performed a series of molecular dynamics (MD) simulations with PTMetaD-WTE of LKα14 adsorbing onto a carboxyl-terminated alkanethiol SAM with both substrate and film naturally-occurring defects incorporated to mimic experimental observations. In addition to the simplicity of the peptide (the alpha helix organizes the sidechains into a hydrophobic and charged, hydrophilic face with sequence LKKLLKLLKKLLKL), this combination of surface and peptide was chosen owing to the ease with which future experiments could be performed related to further understanding defects in SAMs. With an idealized SAM as a control, two types of defects are introduced, namely a gold depression that creates shortened alkyl chain lengths to mimic a characteristic defect in the underlying gold substrate, and a characteristic film defect arising from faulty packing of the SAM (i.e., chains pointed towards and away from each other), creating domain boundary effects. We also used an advanced clustering analysis and reweighting technique to reveal large differences in surface-bound peptide conformations caused by the presence and type of incorporated SAM defect. As we discuss, this analysis is quite general and can be applied to any type of biased protein/surface simulation.

Methods

3.2.1 System Setup

System specifications are reported in Table 3.1, including information from a control simulation without defects from our past work[29]. Systems consisted of one LKα14 peptide, a SAM surface functionalized with a carboxylic acid/carboxylate head group, explicit TIP3P waters, and sodium ions to achieve system charge neutrality. The LKα14 peptide structure was generated with the VMD psfgen plugin[47] and the defect-free SAM surface was based on our prior studies. LKα14 was capped with a deprotonated carboxylate group to match experimental conditions[12-19, 23], imparting it an overall peptide charge of +5. Two types of defects were introduced into the SAM surfaces. The first type of defect mimics an experimentally observed defect in the underlying gold substrate where depressions in the gold layer lead to shortened alkyl chain lengths (hereafter referred to as a "Type I" defect, see Figure 3.1).

The original surface consisted of 100 randomly alternating protonated and deprotonated chains in a 1:1 ratio to mimic a bulk pH of 7.4[49]. Fifty chains were randomly mutated to have reduced alkyl chain lengths from 12 to 8 carbons. The same force field parameters were used for the head groups of both the healthy and mutated chains, leaving the overall surface charge of -50 unaffected. Force field parameters were taken from the AMBER99SB-ILDN force field[51] (i.e. COOH/COO from glutamic acid/glutamate). Triplicate systems were set up in this manner; distributions of the healthy/mutated chains for the 3 systems are shown in Figure 3.2.

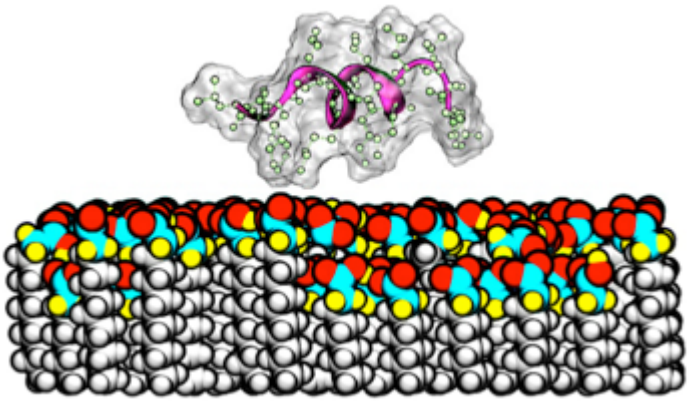

Figure 3.1. Side view of LKα14 (sidechains shown in space-filling representation and hydrogen not included) on a SAM surface with a Type I substrate defect causing areas of shortened alkyl chain lengths. The +z direction is orthogonal to the SAM surface and the +x direction is out of the plane of the page. Chains are colored to highlight frozen atoms (silver = frozen CH_2 atoms) and head group atoms allowed to remain free during MD simulation (teal = carbon, yellow = hydrogen, and red = oxygen).

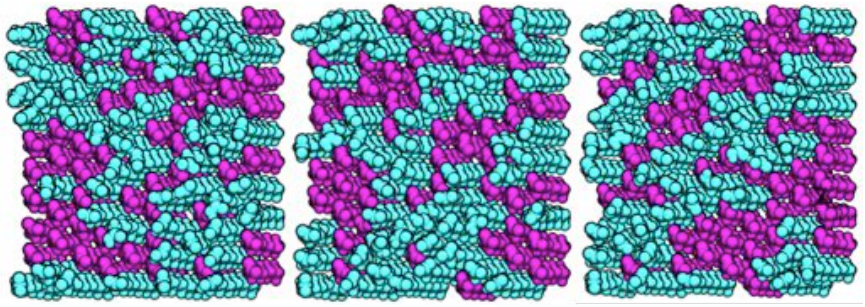

Figure 3.2. Distribution of healthy to defective (i.e. short) chains for the three Type I defect simulation trials. The +z direction is out of the plane of the page. Cyan and magenta represent healthy and defective chains, respectively.

The second type of defect mimics a characteristic film defect that occurs during SAM self-assembly, where alkyl chains pointing in opposite directions leads to domain boundary effects (hereafter, "Type II" defect, see Figure 3.3). To introduce this defect while still maintaining the original R3 geometry and 30° normal tilt angle of the SAM chains[48], it was necessary to remove 30 of the original 100 chains. A portion of the remaining chains was then rotated about the chains' centers of mass (minus the head groups), creating both the outwards and inwards defects shown from left to right in Figure 3.3. To prevent spurious interactions with the thiol group exposed at the base of the inward boundary defect, thiol groups were removed from the original surface. As all simulations used periodic boundary conditions in the x, y, and z dimensions to allow for electrostatic calculations with the particle mesh Ewald (PME) method[53], the peptide could interact with water in the triangular regions marked in blue in Figure 3.3.

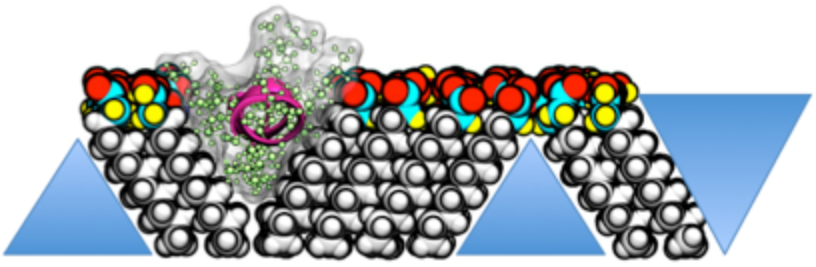

Figure 3.3. Side view of LKα14 (sidechains shown in space-filling representation and hydrogen not included) on a SAM surface with a Type II film self-assembly defect causing inwards and outwards boundary effects. The +z direction is orthogonal to the SAM surface and the +y direction is out of the plane of the page. Chains are colored to highlight frozen atoms (silver = frozen CH_2 atoms) and head group atoms allowed to remain free during MD simulation (teal = carbon, yellow = hydrogen, and red = oxygen).

Simulations used the GROMACS 4.6.5 MD engine[50] with the AMBER99SB-ILDN force field[51] and the PLUMED 2.0 plugin[52]. Box heights were chosen to permit diffusion of the peptide beyond the short-range van der Waals and Coulombic cutoff distances of 1.0 nm to experience a bulk-like state. The peptide was prevented from interacting with the image of the surface by placing a harmonic restraint on its center of mass that began acting on the peptide at a z-distance of 4.5 nm from the top of the surface. Energy minimization was performed on all surfaces with a steepest descent algorithm for 40,000 steps, followed by minimization of the solvated peptide/surface systems where the first 6 and 10 CH_2 groups were frozen for the mutated and healthy SAM chains, respectively. Chains were frozen to prevent diffusion or melting at high temperatures and remained frozen in all ensuing simulations while movement of the head groups was unrestricted.

Table 3.1. Setup of PTMetaD-WTE simulations.

Defect Type	Trial	Total Number of SAM Chains	COO/ COOH Chain Ratio	Mutated COO Chains	Mutated COOH Chains	Na^+	Waters	Box Lengths (nm^3)
I	I	100	1:1	24	26	45	4334	4x5x8
I	II	100	1:1	27	23	45	4339	4x5x8
I	III	100	1:1	23	27	45	4334	4x5x8
II	N/A	70	3:4	16	24	25	4490	4x5x8
None[29]	N/A	100	1:1	0	0	45	3957	4x5x8

3.2.2 Enhanced Sampling

Due to the strong binding forces that exist between the peptide and surfaces, the use of a multiscale modeling algorithm to overcome sampling challenges is essential. This type of algorithm, as applied to protein adsorption, should 1) have strong atomistic detail (e.g., be based

on MD or other molecular techniques), 2) be scalable to systems of practical size, and 3) allow for quantitative comparison with experiments (e.g. in resolving the conformation and orientation of adsorbed proteins for comparison with, for example, SFG results). A method that can address all of these challenges is metadynamics (MetaD)[42, 126], which works by applying a history-dependent bias to one or more collective variables (CVs) that describe the underlying changes in a system (e.g., interfacial versus solution state structure of biomolecules in an adsorption process) in reduced dimension:

$$V(\mathbf{s}(\mathbf{r}),t) = W \sum_{t'=\tau_G,2\tau_G,}^{t'<t} \prod_{i=1}^{N_{CV}} \exp\left[\frac{-\left(\mathbf{s_i}(\mathbf{r}) - \mathbf{s_i}(\mathbf{r}(\mathbf{t}'))\right)^2}{2\sigma_i^2}\right] \tag{3.1}$$

The added bias potential, *V(s,t),* is added to the overall potential energy and is repulsive, Gaussian-shaped, and centered on the CV at the time of addition. This results in a net force that prevents the system from exploring previously visited states and instead encourages it to explore new regions of the CVs. To achieve smooth convergence of the bias potential we use the well-tempered variant of metadynamics (WTM)[58]:

$$W' = \omega * \exp\left[-\frac{V(s,t)}{k_B \Delta T}\right] \tag{3.2}$$

In Eqn. 1, the number of CVs is given by N_{CV}, the values of which are defined by a functional mapping that relates the CV to the system's geometry, or **s(r)**. Gaussian "hills" are added every τ_G time steps with characteristic height W and width σ. WTM leads to an exponential decrease in the amount of bias added to previously explored regions of phase space (Eqn. 2). The instantaneous hill height, *W'*, is also controlled by an adjustable parameter ΔT that is related to the characteristic barrier heights in the system. In a post-processing manner the cumulative bias from the simulation can be inverted to obtain the underlying free energy surface (FES) as projected onto the CVs[80].

Despite its capacity to greatly enhance conformational sampling, MetaD suffers from the ability of the chosen CVs to overcome hidden degrees of freedom in the system. This can be addressed with the use of parallel tempering (PT)[56, 127], which manipulates some or all degrees of freedom in a more general way (e.g. by increasing the temperature of the system). PT works by requiring many parallel simulations or "replicas" of the system that span a wide temperature range and exchange configurations periodically according to the Metropolis criterion. In this way PT can be combined with metadynamics (PTMetaD[43]) to both increase the exploration of CV space and overcome hidden energy barriers.

The addition of sampling in the well-tempered ensemble (WTE)[44] provides an efficiency boost to the method, which has been discussed elsewhere[55]. The WTE algorithm works by using the potential energy itself as a CV and amplifying energy fluctuations (while leaving average energies of the original ensemble untouched) to increase overlap in the energy distributions of adjacent temperatures. This in turn increases the frequency of exchange between replicas and thus increases the overall efficiency of the method. The degree of amplification of the energy fluctuations is controlled via the same adjustable parameter ΔT. However, the WTE bias of the

simulation is generally constructed with a different value of this parameter. Commonly, ΔT is rewritten as γ, called the bias factor, where $\gamma = (\Delta T+T)/T$.[126]

PTMetaD-WTE was used with the same procedure described in past work[29], including the use of a new functionality in PLUMED2[52] to provide a slight improvement to the method. Spanning a range of 300-450 K, 12 configurationally identical replicas were simulated in a short, 1 ns NVT PT simulation to equilibrate each replica at its respective temperature. A 10 ns WTM simulation biasing the potential energy was then performed to establish the WTE to increase sampling efficiency through increasing the spread in the system's potential energy. A bias factor of 20 was used in all WTM simulations with Gaussian hills added every ps with a width of 200 kJ/mol at an initial height of 2.0 kJ/mol.

Production runs biased two CVs for LKα14 with an additional two-dimensional MetaD bias potential. As with past work[29], the first CV biased the distance between LKα14's center-of-mass (COM) and the surface, whereas a second conformational CV biased the number of backbone α-helical hydrogen bond contacts. A switching function with a reference bond length of 0.25 nm was used to define the degree of the contact, which was defined between α-helical hydrogen bond donor/acceptor pairs along the peptide backbone (i.e., i, i+4 pairs). The distance and conformational CVs were biased with Gaussian hill widths of 0.05 and 0.1 nm, respectively. A bias factor of 10 was used in all PTMetaD-WTE simulations with Gaussian hills added every ps at an initial bias deposition rate of 3.0 kJ/mol/ps.

Results and Discussion

3.3.1 Convergence of MetaD Simulations

To assess convergence of the PTMetaD-WTE simulations, the free energy difference between the adsorbed and solvated states was calculated as a function of time. Convergence was established when the change in the free energy difference became negligible with time. Figure 3.4 shows the change in the Helmholtz binding energy as a function of simulation time for each of the systems listed in Table 3.1. All simulations were initially run for 200 ns per replica, and all Type I defect simulations were deemed converged by the end of that time period. The Type II defect simulation was extended by 50 ns per replica to achieve convergence. Figure 3.4 shows both the type of defect (i.e. Type I vs. Type II) and the distribution of the defects (i.e. Type I, trials I-III) impact the final value of the free energy change upon binding as compared to the control simulation.

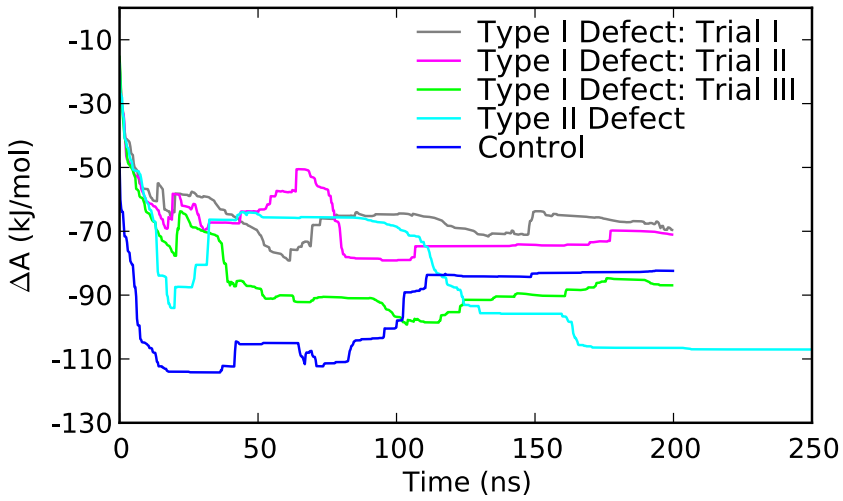

Figure 3.4. Convergence of free energy differences between solvated and adsorbed states for PTMetaD-WTE simulations at 300 K. The negative value implies a decrease in free energy upon adsorption.

3.3.2 Clustering of Surface-Bound Structures

Figure 3.5 shows the Helmholtz energy as a function of distance between LKα14 (Cα centers-of-mass (COM)) and the surface (frozen C10 atom) for each simulation listed in Table 3.1. Figure 3.5c shows the minimum peptide/surface distance for the control simulation is approximately 1 nm, therefore any minima in Figure 3.5a,b below 1 nm represents binding to defective areas of the SAM surfaces.

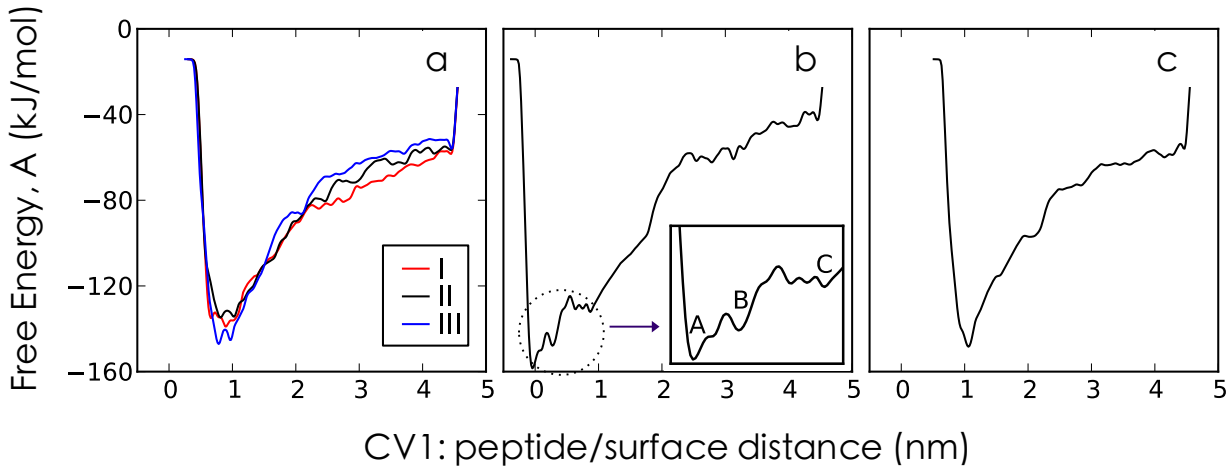

Figure 3.5. Helmholtz free energy as a function of LKα14/SAM distance for PTMetaD-WTE simulations at 300 K: a) Type I defect simulations, trials I-III; b) Type II defect simulation, energy minima highlighted in inset; and c) control simulation. Note that the relative energy scale is arbitrary owing to the trivial constant introduced in the estimation of the free energy from the MetaD bias potential.

To determine the effect of the defects on peptide adsorption, an RMSD-based clustering algorithm[128] was used to extract the most dominant structures in each of the wells in Figure 3.5. The algorithm works by first removing external translational and rotational motions so that only the internal structural fluctuations can be characterized. A least-squares alignment between all unique pairs of structures is then performed and an RMSD value is calculated for each pair. For each structure, other structures that fall below a given cutoff value in RMSD are assigned as "neighbors". The structure with the largest number of neighbors and all of its assigned neighbors are assigned a cluster number and removed from the pool of clusters. The process is then repeated for all remaining structures until each is assigned a cluster value.

An important point should not be overlooked. The clusters obtained in the manner described above are obtained from biased MD trajectories. Therefore, it is impossible to directly compute relative cluster weights or probabilities only using the output of a clustering analysis. Instead, we employed a previously demonstrated reweighting technique[82] that makes use of the classic Torrie-Valleau umbrella sampling reweighting approach[129] with statistical weights calculated according to Eqn. 3:

$$w = \exp(-V_{bias}\beta) \tag{3.3}$$

where the bias potential in this case is obtained by using the final MetaD bias potential treated as a static biasing potential. We note for interested readers that this analysis is trivially performed within PLUMED/GROMACS by using the "-rerun" functionality of the MD engine along with the final MetaD bias (e.g., the 'HILLS' file) and the MD trajectory (i.e., in this case the 300 K replica trajectory from the sampling scheme). Care should be taken to avoid using the portion of the trajectory that corresponds to the MetaD transient. However, in this case this is not an issue as we only clustered the 2nd half of the trajectories – far beyond the end of the transient period. With the proper statistical weights in hand for the trajectory of surface-bound structures, the final probability of each cluster is trivially calculated by normalizing and summing the individual weights (calculated via Eqn. 3) for each member in each cluster.

The analysis was first performed on the trial III Type I defect simulation; since Figure 3.5a shows similar free energy profiles for the three trials, we deemed analysis of a single trial to be sufficient. Skipping every second frame to reduce computation time, surface-bound structures (defined as peptide/surface distances below 1.2 nm) were clustered with an RMSD cutoff value of 0.2 nm. As noted above, we used only the second half of the trajectory for the clustering analysis to eliminate the transient part of the MetaD bias potential. Among 39,696 structures, 78 clusters were determined. The control simulation was analyzed in a similar manner, resulting in 29 clusters from 29,848 surface-bound structures. The central conformation of each cluster, the so-called 'cluster centers', for the top three weighted clusters for each of these simulations, along with their respective weights, is shown in Figure 3.6. Both top and side views are included for the Type I defect simulation to highlight binding to either normal or shortened alkyl chain lengths.

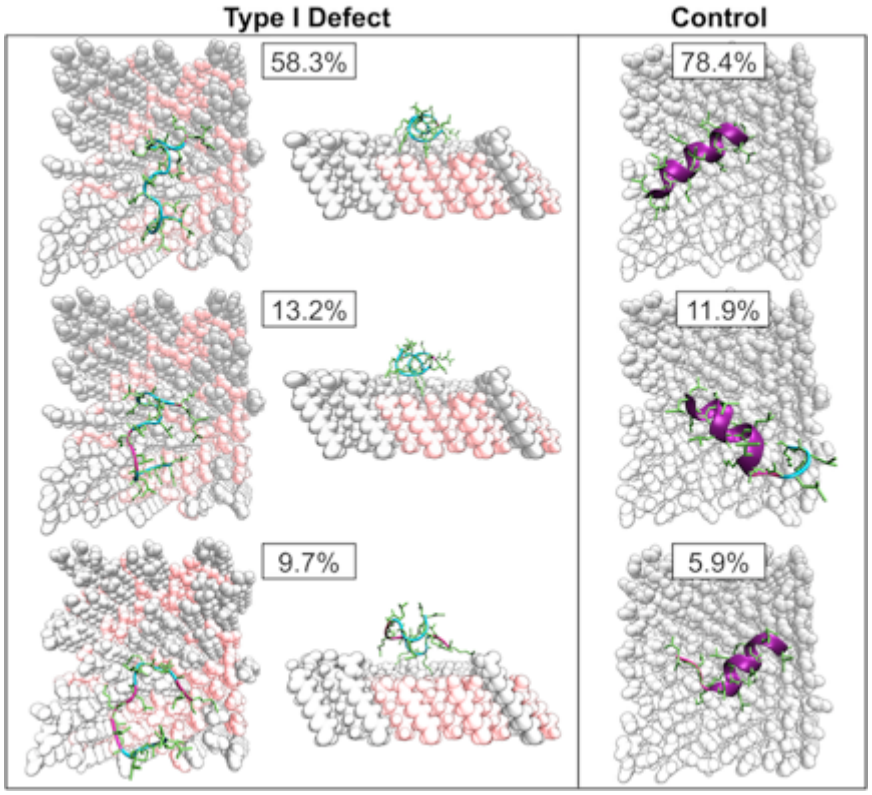

Figure 3.6. Top 3 surface-bound cluster center conformations from a clustering analysis of the Type I, trial III defect simulation compared to the control simulation with no chain defects. Secondary structure is indicated by peptide backbone color: purple designates an α-helix, magenta a turn, and cyan a random coil. Silver and pink represent healthy and defective chains, respectively.

The first thing to note is the difference in cluster distribution between the defect and control simulations: conformations in the top three clusters of the defect simulation make up about 81% of the total probability of surface-bound states, whereas conformations in the first cluster alone in the control simulation have a similar probability of existing on the surface of just over 78%. As Figure 3.6 shows, this is because areas of shortened alkyl chain lengths caused by depressions in the gold substrate below the SAM surface dramatically disrupt the helical structure that LKα14 normally adopts at interfaces, leading to a wide array of unfolded structures. Nearly all secondary structure, indicated by the color of the peptide's backbone (i.e. magenta, cyan, and purple indicate turns, coils, and alpha helical residues, respectively), is lost with the addition of the surface defects. Unlike the central cluster conformations from the control simulation, those from the defect simulation appear to have little in common apart from a tendency towards unstructured coils, which makes sense as defective chains are randomly distributed across the surface.

The same analysis was performed on the Type II defect simulation for each of the three energy minima highlighted in Figure 3.5b (i.e. A, B, and C). These minima are related to the presence of the outwards boundary defect (see Figure 3.3); the inwards boundary defect appears to have little influence on binding. Within +/- sigma of each minimum, all structures below an RMSD cutoff of 0.2 nm were clustered. This resulted in 9,885 structures in 11 clusters for minimum A, 41,203 structures in 23 clusters for minimum B, and 14,710 structures in 9 clusters

for minimum C. The central cluster conformations of the clusters with the top 3 weights calculated for each of the minima are shown in Figure 3.7.

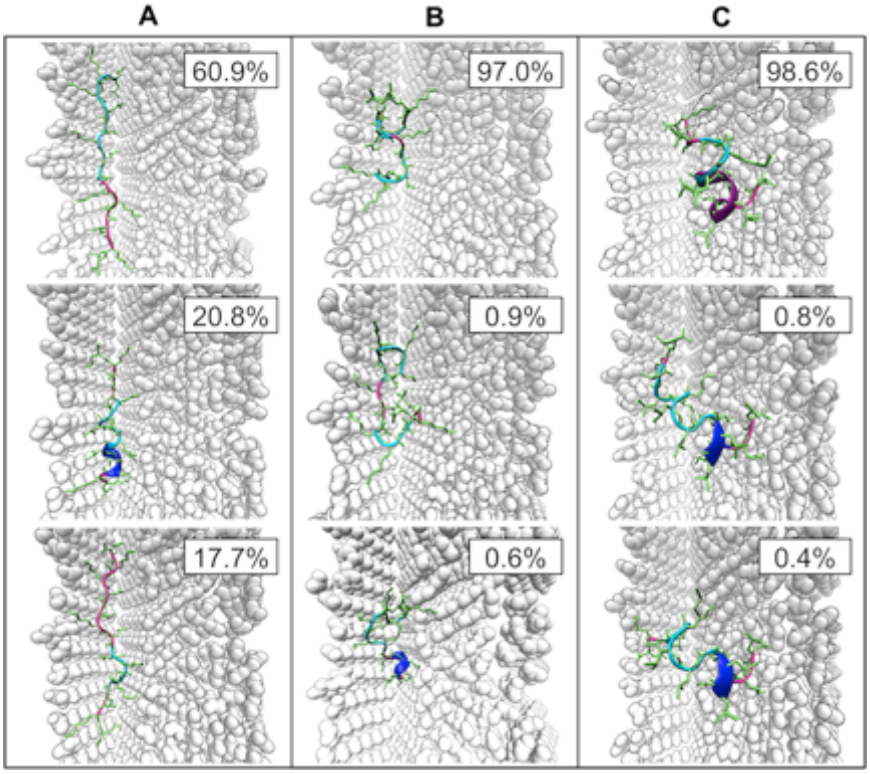

Figure 3.7. Top 3 surface-bound cluster center conformations from a clustering analysis of the Type II defect simulation for each energy minima highlighted in Figure 3.5b. Secondary structure is indicated by peptide backbone color: purple designates an α-helix, blue a 3_{10}-helix, magenta a turn, and cyan a random coil.

Similar to the Type I defect results, conformations in the first cluster of energy minimum A make up about 60% of all surface-bound states. As the distance between the peptide and surface increases to correspond to energy minima B and C, however, the cluster distributions become tighter (i.e. over 95% of all surface-bound structures reside in the top weighted cluster), similar to what was observed with the control simulation. The trends make sense given that the results for energy minimum C should most closely represent those of the control simulation due to the particular peptide/surface distance.

Deep in the hydrophobic cleft (i.e., minimum A) highly extended conformations of LKα14 are stabilized compared to structures in the control simulation, which we believe is due to the shape of the defect. Figure 3.5b shows binding in the pocket of minimum A is stronger than that for minimum B and much stronger than that for minimum C on top of the surface, which, as mentioned earlier, should most closely resemble the control simulation. Some α-helicity is retained on top of the surface (i.e., minimum C), as indicated by the purple color of the peptide's backbone in the cluster center conformations. However, even the mere presence of the defect causes the peptide to

extend over the edge of the surface into the cavity, thereby affecting the normally helical structure of LKα14.

Conclusions

The enhanced sampling method PTMetaD-WTE was employed to simulate the adsorption of LKα14 to a model hydrophilic SAM with a carboxylate/carboxylic acid-terminated head group and two types of induced surface defects. Naturally occurring defects were chosen to best mimic what has been observed experimentally and included both a substrate defect and a characteristic SAM film defect. Results of free energy versus peptide/surface distance showed a difference in the location of the free energy minima for the surfaces with defects compared to a control surface with no defects. The results also indicated binding to the surface with the characteristic film defect ("Type II" defect) is much stronger than binding to the control surface, which we hypothesized is due to the specific shape of the hydrophobic cleft defect.

A clustering analysis was performed to elucidate structural differences in the bound peptide caused by the surface defects. Results showed the presence of either type of defect heavily disrupts the helical structure that LKα14 normally adopts at interfaces. In performing this analysis, peptide structures were extracted from basins, aligned, and clustered, and thus orientation of the peptides with respect to the surface was not taken into account, only the conformation. In this case, it was not important to distinguish between orientations because charged or hydrophobic sidechains dominate the surface-bound orientations. However, prior to reweighting it would be trivial to extend the clustering analysis to distinguish between orientations by subdividing further to, for example, distinguish between hydrophobic/hydrophilic patches on a peptide or protein or using other directional descriptors to account for protein orientation in conjunction with the conformational clusters.

This work will also have implications for future experimental work. Surface-guided self-assembly of proteins is growing in interest; the observed effects on peptide structure from relatively small changes in surface roughness suggest careful design of the electrostatic and van der Waals interactions at the protein/surface interface may be required. Additionally, this method could be used as a means to reverse engineer protein structure by designing and incorporating specific surface defects to control the structure of biomolecules upon adsorption.

Finally, we note that the predictions from these simulations could be directly probed with surface spectroscopies such as sum frequency generation (SFG) spectroscopy[14]. Provided self-assembly of SAMs of different chain lengths were possible, adsorption of LKα14, we predict, would reveal no appreciable SFG signal compared to neat SAMs, which reveal the expected helical structures. Likewise, using a combination of techniques like surface plasmon resonance (SPR) and atomic force microscopy (AFM)[26], we propose it would be possible to study the expected increases in binding energy due to the film formation defects. Of course, this would depend on being able to synthesize in a controlled way the film type defects.

Chapter 4

Using Molecular Simulation to Study Biocatalysis in Ionic Liquids[4]

Introduction

4.1.1 Motivation for Studying Biomolecules in Ionic Liquids

The study of biomolecules in room temperature ionic liquids (sometimes referred to as RTILs; here called simply ILs) has applications in biochemistry, biomedicine and nanotechnology.[130] Early examples in the area of biocatalysis in ILs included new biomass pretreatment methods to facilitate the breakdown and conversion of cellulose to green fuels and chemicals [131], and the enzymatic breakdown of waste food oils [132-133]. Compared to traditional organic solvents, ILs have many beneficial properties that could support novel uses of enzymes in industry (e.g., negligible vapor pressures, low flammability, high recoverability, and extremely tunable solvent properties).[134] Although there are many examples of combing ILs with the full spectrum of various biomolecules, in this article we are focusing on enzymes due to their unique ability to catalyze chemical transformations. Both computational and experimental efforts have provided valuable insights into the microscopic interactions of enzymes with ILs. However, this book chapter is focused on simulations of biocatalysis in ILs, and thus experimental studies will be highlighted only briefly. A very recent review article by Benedetto and Ballone paints a more detailed picture of the current status of experiments related to biomolecules in ILs.[130] Additionally, we emphasize that this chapter is not meant as a comprehensive review of all enzyme/IL simulations to date. Rather, our aim is to provide a guide for researchers looking to start performing their own enzyme/IL simulations and focus on best simulation practices, potential pitfalls and limitations, and common analysis techniques.

[4]Reproduced in part with permission from K.G. Sprenger and J. Pfaendtner. Using Molecular Simulation to Study Biocatalysis in Ionic Liquids. Methods in enzymology, 577: 419-441, 2016. Copyright 2016 Academic Press.

4.1.2 Experiments of Biocatalysis in Ionic Liquids

Experiments of enzymes in ILs have been on the rise over the past decade or so, though, like its computational counterpart, the field is still very new compared to the study of enzymes in native or native-like environments. Difficulties arising from the use of non-aqueous solvents have prevented the use of many of the typical analysis techniques used to study enzymes. Many papers that have been published thus far that deal specifically with enzymes in ILs utilize techniques such as circular dichroism (CD) [135-136], differential scanning calorimetry [137], and fluorescence spectroscopy techniques [136-139] to observe dynamics and conformational changes upon insertion in the IL. Dynamic light scattering has been used to observe enzyme aggregation or large-scale denaturation in ILs.[140] Additionally, Kaar and coworkers have recently demonstrated the use of 2D NMR methods as a means to understand and improve enzyme stability in ILs.[141] Common activity assays have also been used including colorimetry [135], activity and conformational stability assays [142-143], and fluorescence quenching assays to find the extent of ion binding to the enzyme surface [143]. Kinetic studies using ultraviolet-visible spectroscopy or other techniques have yielded Michaelis-Menten constants [135-136, 139], and quantitative structure-activity relationship (QSAR) models have been developed to predict enzyme performance in ILs [144]. To determine IL effects on enzyme activity, experiments by Kim and coworkers in 2014 investigated the lipase-catalyzed transesterification reaction of butyl alcohol with vinyl acetate in ILs.[145]

Directed evolution techniques have also recently been applied to enzyme/IL systems as a means of optimizing enzymes with respect to desired properties.[146-148] For example, Carter and coworkers mutated *Candida boidinii* with error-prone PCR to improve the enzyme's thermostability and tolerance to ILs.[149] Wolski et al. evolved variants of cellulase enzyme Cel7A via biased clique shuffling, based on the standard DNA shuffling, to have increased stability in aqueous ILs to improve the biomass dissolution process for biofuels production.[150] Iterative saturation mutagenesis is another technique to direct the evolution of enzymes. Ultimately, the choice of technique depends on the desired enzyme property to be optimized.[149]

4.1.3 Simulations of Biocatalysis in Ionic Liquids

To our knowledge, the first paper of a molecular simulation study of an enzyme in an ionic liquid was published in 2008 [151] followed by subsequent work in in 2011 from other groups [152-153]. Since that time, the number of classes of enzymes and types of ILs studied has grown significantly. For reference, Table 4.1 shows the broad range of combinations of enzymes and ILs, both aqueous and nonaqueous, studied up till now (n.b., this list may not be exhaustive but is meant to show range and depth both in the choice of enzyme and IL). The full names and chemical formulas of the ILs listed in Table 4.1 are given in Table 4.2. As Table 4.1 shows, although a wide variety of biomolecules and ILs have been studied, the large amount of prior experimental work has led to more common uses of lipases such as *Candida antarctica lipase B* (CALB) and *Candida rugosa lipase* (CRL), and imidazolium-based ILs (i.e. [RMIM], where R is the alkyl side chain defined in Table 4.2).

The majority of the studies published thus far have focused on characterizing biomolecular structure and dynamics in ILs with water as a reference to measure and learn how to control the stability and retained activity, respectively, of enzymes in nonaqueous media. Recently, examples of enzyme/IL simulations have begun to explicitly mimic advanced enzyme design procedures.

Burney et al. [154] following the experimental protocol developed by the Kaar group [142-143], simulated the surface charge modification procedure and studied the effect of controlled surface charge modification on the long-range structure and properties of the ionic liquid near the enzyme surface. In many cases, as Table 4.1 shows, the role of water content has been explicitly examined by studying solute behavior in varying concentrations of aqueous IL solutions. Common practices of simulating enzymes in IL solutions, even in pure IL solvents, include retaining the crystallographic water molecules that surround and at times penetrate the structure of the enzyme, primarily because of their important role in upholding enzyme stability [153]. A common method of analysis to gain mechanistic insight into enzyme stability in ILs is to track the diffusion of water from the enzyme surface over time in IL solutions compared to in pure water (Micaêlo 2008, Burney Pfaendtner 2013, latif 2014). However, some studies have taken the approach of removing the surface crystallographic waters while retaining just the water molecules buried within the enzyme [152] or removing the structural waters altogether [145], citing the activation of the enzyme under anhydrous conditions.

Later in this chapter we provide more detailed descriptions of common analysis tools and techniques to study enzymatic behavior in ILs with specific tips and suggestions to get started. However, to highlight the wide range of examples and interesting analyses that atomistic molecular simulations uniquely enable, we describe a few unique examples in the remainder of this section. Some additional analysis methods to note include the use of principal component analysis to gain insights into the representative slow modes of enzymes in water and ILs [140], which suggested that the ILs can disrupt relative ordering and strength of various slow, correlated motions. Building on this was the observation that in glycoside hydrolases, transient behavior of hydrogen bonds and salt-bridges could be linked to enzyme stability. [155]

Finally, there are growing examples of studying how ILs affect the structure and dynamics of substrate binding pockets in relation to enzyme activity. Kim and coworkers analyzed the conformational changes of the active site of CALB in various ILs and found a direct relationship with enzymatic activity.[145] Depth profiles from the interior of the catalytic cavity to the enzyme surface changed in response to the IL type and differed notably from one solvent to the next. These findings were also verified with follow-up experiments. In another example of computational predictions that were validated for the case of biocatalysis in ILs, the prediction by Jaeger and Pfaendtner [140] of the role of ILs as competitive inhibitors to enzyme substrates was later independently confirmed [156]. As the field of computational biocatalysis in ILs is still relatively new, many testing and validation studies are still needed to understand the accuracy of the force fields used. As others have noted, the tolerance of a given enzyme to any particular IL is generally unpredictable at this time.[130]. Thus, it is an exciting step forward whenever computational predictions are later proved to be correct through experiments.

Table 4.1. Overview of enzyme/IL systems studied to date with molecular simulations.

Biomolecule	IL	% or Conc. in Water	Ref
Serine protease cutinase	[BMIM][NO_3] [BMIM] [PF_6]	2.5,5,7.5,10,15,25, 35,50, 60,75	151
CALB	[BMIM][NO_3] [BMIM][BF_4] [BMIM][PF_6] [MOEMIM][BF_4] [BAGUA][BF_4] [BCGUA][BF_4] [MCGUA][NO_3] [DCGUA] [NO_3]	100	152-153
Xylanase	[EMIM][ACE] [EMIM] [$EtSO_4$]	10,20,50	140
CRL	[BMIM][PF_6] [BMIM] [NO_3]	100	157
Adenosine deaminase	[AMIM][Cl] [OMIM][Cl]	100	139
CALB,CRL	[BMIM][PF_6] [BMIM] [BF_4] [BMIM][Cl] [BMIM] [TfO] [BMIM][Tf_2N]	5,10,15,20,50	158
CALB	[BMIM][TfO] [BMIM] [Cl]	100	145
Cellulases*	[EMIM][ACE]	15,50	155
CRL, bos taurus a-chymotrypsin	[BMIM][Cl] [EMIM] [$EtSO_4$]	20	154
Hen egg white lysozyme	[PMIM][Br]	1.5M	138
Renilla luciferase	[BMIM][PF_6] [BMIM] [BF_4]	14mM,16mM**	136

*Trichoderma viride, thermogata maritime, pyrococcus horikoshii
**Pairwise for ILs

Table 4.2. Abbreviations and formulas for IL cations and anions listed in Table 4.1.

Cation			Anion		
Abbrev	Name	Formula	Abbrev	Name	Formula
[EMIM]	1-ethyl-3-methylimidazolium	$C_6H_{11}N_2^+$	[PF_6]	Hexafluorophosphate	PF_6^-
[BMIM]	1-butyl-3-methylimidazolium	$C_8H_{15}N_2^+$	[NO_3]	Nitrate	NO_3^-
[OMIM]	1-octyl-3-methylimidzolium	$C_{12}H_{23}N_2^+$	[BF_4]	Tetrafluoroborate	BF_4^-
[AMIM]	1-allyl-3-methylimidzolium	$C_7H_{11}N_2^+$	[ACE]	Acetate	$CH_3CO_2^-$
[PMIM]	1-methyl-3-pentylimidazolium	$C_9H_{17}N_2^+$	[$EtSO_4$]	Ethyl Sulfate	$C_2H_6O_4S^-$
[MOE-MIM]	1-methoxyethyl-3-methylimidazolium	$C_7H_{13}N_2O^+$	[TfO]	Trifluoromethane-sulfonate	$CF_3SO_3^-$
[BAGUA]	acyclic butylpentamethyl-guanidinium	$C_{10}H_{24}N_3^+$	[Tf_2N]	bis(trifluoromethane-sulfonyl)imide	$C_2F_6NO_4S_2^-$
[MCGUA]	cyclic tetramethyl-guanidinium	$C_6H_{13}N_4^+$	[Cl]	Chloride	Cl^-
[BCGUA]	cyclic butyltrimethyl-guanidinium	$C_9H_{19}N_4^+$	[Br]	Bromide	Br^-
[DCGUA]	cyclic decyltrimethyl-guanidinium	$C_{15}H_{31}N_4^+$			

Methods for Simulating Biomolecules in Ionic Liquids

4.2.1 Choice of Force Field/Parameterization Process

Many different approaches exist for setting up simulations of enzymes in ionic liquids. There are three main parts to the setup process – choice of force field to simulate the enzyme, choice of force field to simulate water, and parameterization of the force field for the IL. Whereas there are many choices for the former two, there are widely varying approaches to establishing the IL force field. Due to length considerations, we will not discuss the choice of biomolecular force field here beyond noting the general need for the class of the force field to have compatibility with both the water model and force field for the IL. Regarding the choice of water model, a wide range of force fields have been used to simulate both the water molecules present in aqueous IL solutions or the crystallographic water surrounding the enzyme in pure IL solvents. The most common choice of model in studies published thus far are 3-site water models like SPC [151], SPC/E [138], and

TIP3P.[140, 154-155, 157]. There has been a single study with a 4-site water model [158]. While more computationally expensive than 3-site or 4-site water models, studies by Klahn et al. used the TIP5P water model [152-153], citing their previous experiments that found TIP3P overestimated the strength of water/IL interactions and led to incorrect mixing of [BMIM][PF_6], known to be a water-immiscible IL, with water; TIP5P correctly predicted no mixing of [BMIM][PF_6] with water in the simulations [159].

The greatest variation in approach among the published studies on biocatalysis in ILs lies in the parameterization of the IL cations and anions. Generally, approaches fall into one of three categories: 1) use or manually create a customized force field specifically refined for a particular IL; 2) use a generic or universal force field combined with customized partial atomic charges via quantum mechanics calculations; or 3) use a self-contained computer program to generate force field parameters for the IL(s) of interest. Guiding the choice of how to proceed with IL parameterization are the major considerations of scope and scale of the researcher's study. If the goal is to test the stability of a particular enzyme in many ILs from different families (i.e. no common cation or anion to all ILs) in a high-throughput manner, the parameterization method should be general and easily applied across the range of ILs tested, like in approach 2. Conversely, if the goal of the study is, for example, to reproduce experimental results of the behavior of an enzyme in a particular IL or family of ILs, an IL-specific (and thus more accurate) parameterization method would be appropriate, like in approach 1. A few studies thus far have utilized a united-atom model for the ILs: in the spirit of the first type of approach, Micaêlo et al. developed a united-atom based model for two imidazolium-based ILs in the framework of the GROMOS force field [151], whereas Ghosh et al. and Ajloo et al. followed the third type of approach and generated united-atom IL force field parameters from software such as the Automated Force Field Builder and Repository (ATB and Repository) [160], also based on the GROMOS force field, and PRODRG2 [161], respectively [138-139]. OPLS-based force fields have also been common: Latif et al. [158] and Klahn et al. [152-153] used the IL family-transferable force fields developed by Canongia Lopes et al. [162-164] to model the ILs in their studies, and Kim et al. [145] used the force field parameters developed for 68 different ILs by Sambasivarao et al. [165]. Our own group, seeking to leverage the high quality AMBER family force fields for proteins and carbohydrates, adopts the second approach and makes use of the fact that the general AMBER force field (GAFF) can reproduce many properties of neat ILs [166]. This has permitted seamless combining of GAFF-parameterized ILs with modern AMBER protein force fields like AMBER ff99SB [167] and GLYCAM [168], the AMBER carbohydrate force field, to study many different biomolecule/IL systems [140, 154-155, 157, 169]. For many ILs, GAFF already includes all the necessary bonded and non-bonded interactions with the exception of electrostatic point charges, which are obtained in a customized manner via a standard procedure. The developers of the CHARMM family of force fields recently published a generic force field, CGenFF [170-171], which is an ideal candidate to further test the ability of a general force field to have widespread predictive power.

A common need, irrespective of the various parameterization methods, is accounting in some way for the slow dynamics of room temperature ILs, which if ignored may lead to poor configurational sampling and increase simulation convergence times [151, 158]. The origin of this problem stems from the high viscosity of many ILs that can artificially dampen system dynamics if not captured and corrected for by the IL force field, and can lead to wildly inaccurate predictions of IL transport properties. Micaêlo et al. found it helpful to scale the masses of heavy atoms in their system in order to bring all of the characteristic vibrational frequencies more in line with each other.[151] They acknowledged this had an effect on the overall dynamics of the system, but it left

the equilibrium thermodynamic properties of the system unchanged. Another method to overcome the sluggish dynamics of ILs is to simply scale down the partial charge of each cation/anion. Studies have shown that classical force fields generally do a poor job of estimating the transport properties of ILs (e.g. shear viscosity or self-diffusivity) when full charges are assigned to the ions.[172] This is best explained by the fact that using full charges leads to a poor approximation of liquid-like behavior [172-174], i.e. charge screening by neighboring molecules and electron density effects by neighboring molecules of opposite charge are not well represented. It has been shown that scaled partial charges lead to significant improvements in the resulting transport properties of ILs.[172] The use of GAFF combined with the Antechamber/RESP charge generation method and scaled IL partial charges by a factor of 0.8 has been used in some more recent biocatalysis/IL studies.[154-155] The steps to carry out this particular parameterization method will be described in the next section, followed by a small case study to illustrate the effects of charge scaling in ILs on the conformational transitions of biomolecules. It should be noted that this method of performing quantum calculations on isolated ions or pairs of ions and then uniformly scaling the charges does sacrifice some accuracy; a much more accurate but computationally exhaustive approach is to use a QM/MM description of the liquid to derive the parameters for the Coulombic potential using the actual charge distributions in the liquid. This method has been employed by Klahn and coworkers to simulate CALB in a number of IL solvents.[152-153]

4.2.2 System Setup Example Using GAFF/Antechamber Tools

There are many freely available programs to establish a workflow for creating the required force fields and starting structures for simulating enzymes in ILs. Figure 4.1 shows a general flowchart for setting up an enzyme/IL system for use in a molecular dynamics (MD) simulation carried out with the GROMACS MD engine.[50] We note that for any individual step, a number of different programs could be substituted, and interested readers are suggested to review the literature listed in Table 4.1 for alternative methods. Steps 1-5 are performed consecutively first for one ion, then for the other, but will be discussed here in tandem to avoid redundancy. The point charges for the cation/anion pair are created by drawing the chemical structures of the ions (step 1, completed in GaussView or a free program like Molden or Avogadro) and subsequent geometry optimization and calculation of the electrostatic potential around the molecule (in step 2, Gaussian [175] conveniently works with the AmberTools program in our workflow, but there are many other options for completing this step). Fitting the electrostatic potential leads to the determination of the atomic point charge on each atom (in step 3, Antechamber [176] is a convenient program because it also assigns GAFF atom types). Combining the generated structural and electrostatic parameters with the equation of GAFF to describe the potential energies and their force derivatives creates force field files for the ions (a program like Parmchk can be used to identify missing parameters (step 4) that can be manually parameterized and incorporated in LEaP [177] in step 5). After constructing the molecular structure file for the full system (step 6, completed in a program like Packmol [178]), force field files for the system are created as before with the ions (step 7, completed in LEaP). Lastly, input files are generated to perform an MD simulation (in step 8, ACPYPE [179] is used to convert between the AMBER and GROMACS file formats).

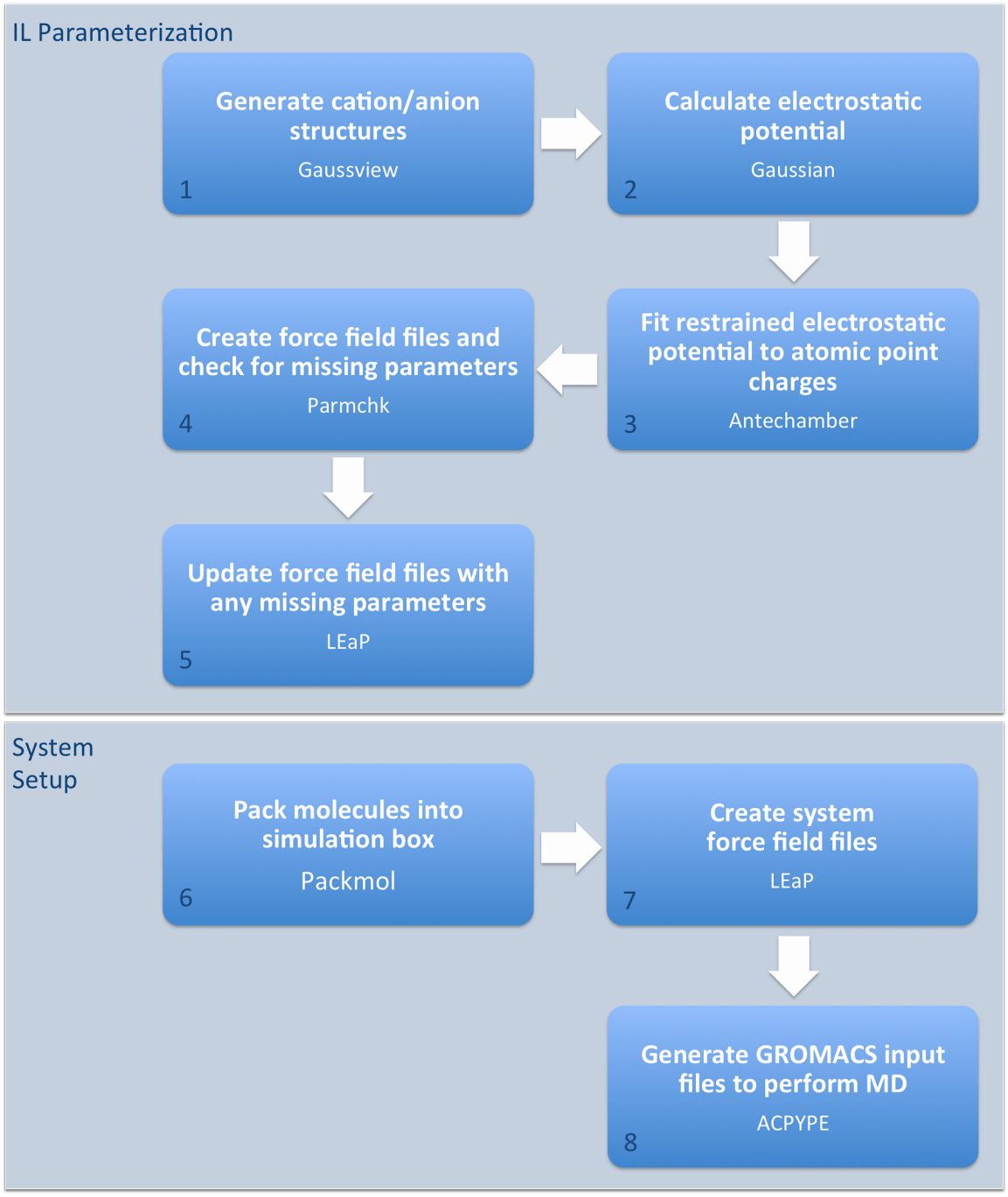

Figure 4.1. Flow chart for developing MD simulation files for an enzyme/IL system to be simulated in GROMACS. This workflow is general and could be used with any MD engine compatible with modern force fields.

4.2.3 Electrostatics and Charge Scaling Considerations

As noted above, many different approaches have been applied to setup simulations of biocatalysts in ILs. One aspect of the parameterization process that deserves further attention is

the scaling of electrostatic point charges of IL ions. In IL systems, the prevalence of large numbers of significant point charges means that electrostatic interactions can contribute to molecular scale fluctuations much more than in purely aqueous systems. To illustrate the potential impact of charge scaling on the properties of an MD simulation of an enzyme, we have performed some simple calculations of alanine dipeptide in solutions of pure 1-butyl-3-methylimidazolium chloride ([BMIM][Cl]) with ion partial charges scaled to values of 1.0 (unscaled), 0.9, and 0.8. Additionally, a simulation is done in water for comparison. The AMBER03 force field [180] was used to model the peptide, TIP3P [181] to model the water, and GAFF [182] to model the organic molecules with RESP partial charges generated via Antechamber. Following a steepest-descent energy minimization of 10,000 steps, a short 1 ns NPT simulation at 300K was run to equilibrate the system pressure, using a global stochastic thermostat [54] and the Berendsen barostat [109]. MD combined with the well-tempered version of metadynamics (MetaD) [42, 58, 86] was used to construct free energy surfaces for each of the four simulations, shown in Figure 4.2. The free energy has been projected onto the phi and psi dihedral collective variables (CVs), which were biased using PLUMED [52] with Gaussian hill widths (σ) of 0.2 radians and hills added every 4 ps (τ) at an initial bias deposition rate of 0.075 kcal/mol/ps. A biasfactor (γ) of 9 was used in all simulations. Figure 4.2 shows the free energy profile of alanine dipeptide in water (A) is very similar to that in [BMIM][Cl] with partial charges scaled by a factor of 0.8 (B). Scaling the charges by a factor of 0.9 (C) and by a factor of 1.0 (D) leads to increasingly higher barrier heights, hinting that longer transition times between states might exist for systems with less scaling of IL partial charges and that the main pathway(s) for changing states might differ as well. Despite these apparent differences in the free energy landscapes, all free energy surfaces show a two-state system with the same general locations of the low free energy wells, α (indicated with an asterisk) and β, across all simulations. It is interesting to note, however, that while the two-dimensional shape of the α well remains very similar across the four simulations, the shape of the β well changes considerably as the partial charges are less scaled, or essentially as the viscosity of the system increases. We also note that convergence was based on the free energy difference between the two wells over time, with the slow dynamics in the IL systems leading to longer convergence times.

To assess the rate of conformational transition in the different systems, we also initiated a series of simulations with the infrequent metadynamics approach [183] to estimate the rate of reaction of the $\alpha \rightarrow \beta$ transition of alanine dipeptide in the different solvents (n.b., a white asterisk in each plot in Figure 4.2 represents the point in CV phase space at which state α is designated). These simulations used well-tempered MetaD simulations with hills now added every 80 ps at an initial bias deposition rate of 0.00375 kcal/mol/ps and post-processed with the infrequent metadynamics algorithm. Calculation of an individual transition time over a barrier is calculated via Equation 1,

$$t^{eff} = \sum_{i=0}^{N} \Delta t_i^{MetaD} e^{\beta V_{bias}(s,t)} \tag{4.1}$$

where Δt is the MetaD time step and V_{bias} is the instantaneous value of the MetaD bias potential at the ith point in the transition trajectory. The point in time at which alanine dipeptide leaves the α basin defines the interval of time in the so-called "transition trajectory". Details about the p value analysis can be found in the work of Salvalaglio, who suggested the approach for checking the validity of the calculations [184]. We also note our analysis and bootstrapping process followed our recent use of infrequent metadynamics to study chemical reaction rates [185].

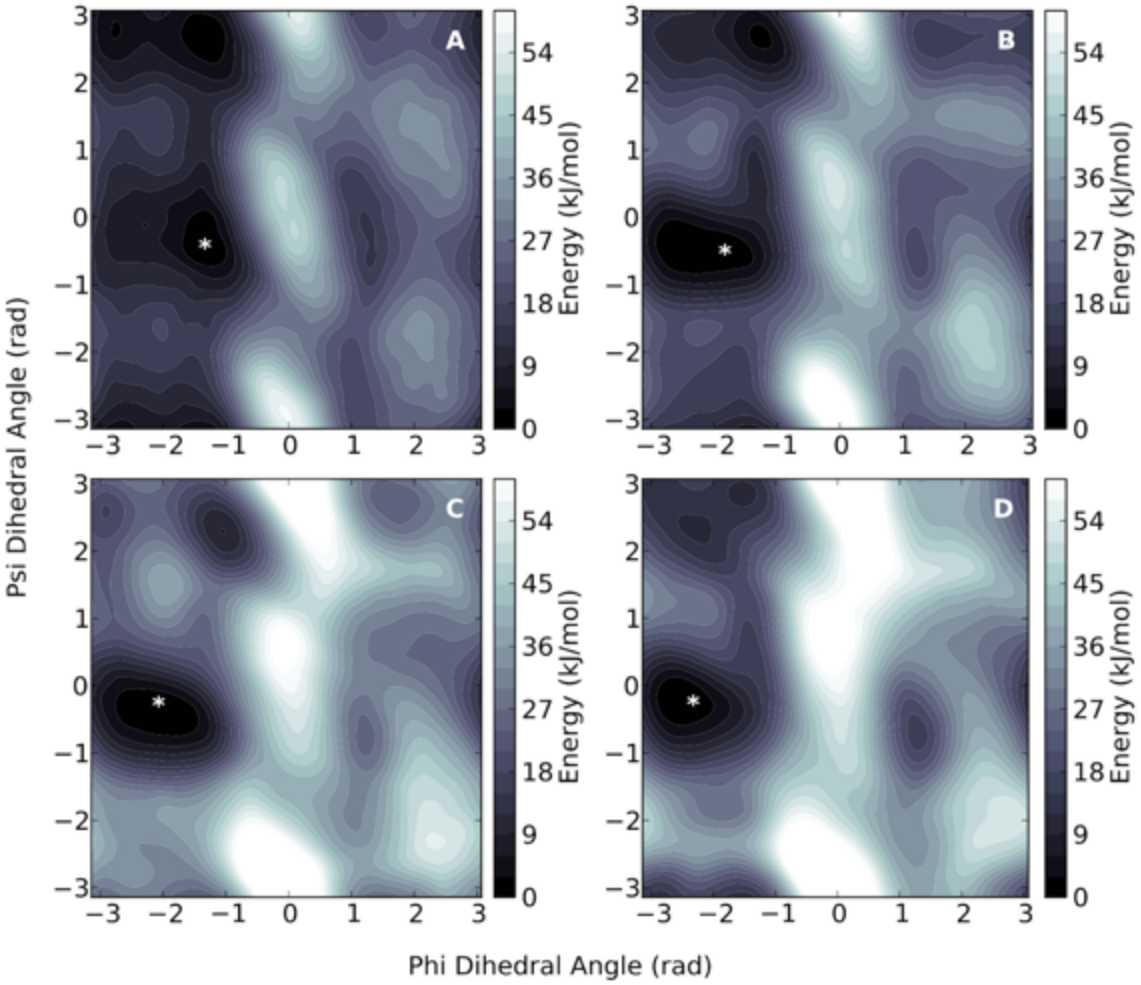

Figure 4.2. Free energy surfaces for alanine dipeptide in: (A) water; (B) [BMIM][Cl] with partial charges scaled by multiplying by 0.8; (C) [BMIM][Cl] with partial charges scaled by multiplying by 0.9; and (D) [BMIM][Cl] with partial charges scaled by multiplying by 1.0 (i.e. not scaled).

Table 4.3 shows that all *p* values are above the recommend threshold of 0.05, thus we can conclude the transition times are uncorrelated and follow a Poisson distribution. The results show drastically different mean transition times, beyond an order of magnitude, are calculated depending on the degree of scaling of the ion partial charges. As predicted in the earlier discussion of the free energy profiles, the mean transition time generally increases as the degree of scaling is decreased from 0.8 to 1.0. In other words, the viscosity has a significant effect on the conformational transitions of biomolecules and should be treated with extreme care when constructing simulations of enzymes in IL solvents.

Table 4.3. Mean escape time data for the alanine dipeptide $\alpha \rightarrow \beta$ transition.

Solvent	Charge scaling	t^{eff} (ns)[a]	p value[b]	Number of events	τ[c]
Water	N/A	0.136(0.016)	0.78	100	80
[BMIM][Cl]	0.8	9.00(1.2)	0.99	100	80
[BMIM][Cl]	0.9	138(17.9)	0.92	100	80
[BMIM][Cl]	1.0	149(21.3)	0.08	100	80

[a]Mean transition times of 100 rare events, with standard deviations in parentheses.
[b]Kolmogorov-Smirnov significance (p) values.
[c]Time between successive Gaussian hills deposition in MetaD simulations.

4.2.4 IL-Specific Aspects of Enzyme MD Simulations

Determining the force field, topology, and all input files for simulations of enzymes in ILs often proves to be the most time-consuming aspect (from the point of view of human effort) of the whole process. Once this is completed, the simulations can be performed with any parallel and efficient MD engine. Beyond the initial setup, there are a few potential considerations that merit further attention.

First, there is a common practice in the enzyme/protein simulation community to use a slow annealing from a low temperature near 10K to the desired simulation temperature. Simulations with a barostat to mechanically equilibrate the system at a fixed pressure often follow this heating step. The IL systems are highly viscous and initial random configurations may lead to strong electrostatic interactions that could artificially pull the system far from equilibrium. To address this we recommend a twofold remedy. First, simulations should be performed with multiple instances of generating the packed box of ILs (and water if desired). Second, in each system we suggest freezing all protein atoms and holding the system at a very high temperature (above 500K) for ~5 ns to equilibrate the solvent structure. The simulations can then be reliably quenched to a very low temperature and then slowly heated with an unfrozen protein according to the common community practice of annealing to room temperature for biomolecular studies.

Next, for simulations of enzyme/IL/water (i.e., there are IL and water cosolvents together), the issue of water miscibility is a potential concern. There is a significant amount of experimental data to guide the selection of the system temperature and amount of water (in order to avoid phase separation). However, unless the force field has already been evaluated for its ability to predict experimental phase separation, it is important to remember the precise temperature and composition at which the phase split occurs. If concerns exist regarding miscibility, we suggest trial simulations of the water/IL mixture to evaluate this issue in detail. Finally, in light of the issue of slower timescales arising from high solvent viscosity, we also suggest extending MD simulations to longer timescales and/or using enhanced sampling approaches, which will be discussed further in the conclusions.

4.2.5 Typical Analysis Approaches

This section briefly highlights typical ways that other researchers have characterized MD simulations of enzyme/IL systems. We note that on a practical level, there is no reason any type

of analysis based on studying average or time dependent behavior of the enzyme, solvent, or both cannot be used in the context of MD simulations of nonaqueous biocatalysis. To date the most common way of characterizing enzyme/IL interactions is through the use of structural-based methods like those listed in Table 4.4, most of which are often calculated post priori versus on-the-fly to avoid computational overhead on MD simulations.

Some groups have made a point to focus on the thermodynamic aspects of enzyme/IL interactions as well. For instance, in 2011 Klahn and coworkers quantified the enthalpic changes that occurred upon solvation of an enzyme in IL compared to in water [152], as well as calculated the strengths of the Coulombic and Lennard-Jones interaction energies between the enzyme and IL ions [152-153]. More recent papers by Burney et al. and Jaeger et al. directly quantified the entropic contributions to enzyme/IL interactions [155, 157] by quantifying side chain fluctuations. Jaeger and coworkers also quantified energetic contributions to enzyme/IL interactions by calculating the internal energy of the protein, i.e. summing up all of the molecular interactions within the protein.[155]

A final type of analysis that is commonly performed on data from simulations of enzymes in ILs is the calculation of radial distribution functions [138-139, 154] and spatial distribution functions [151, 158] of the IL ions and/or water molecules surrounding the enzyme surface. These types of analyses have provided insight into the behavior of ions and water, either solution or crystallographic, at the enzyme surface. In the case of Burney et al., this provided insightful information about the change in proportion of IL cations and anions around the enzyme surface with certain enzyme surface charge modifications.[154] In this same paper, the authors also calculated charge distribution functions (i.e. charge density of the IL ions as a function of the radial distance from the enzyme surface) to elucidate trends in solvent structure. Specifically, the goal of the analysis was to test the hypothesis that due to an anion-dominated surface layer created by enzyme charge modifications, the IL would organize in concentric spherical layers of alternating charge.

Since ILs have been shown to change the conformation of substrates, this is another potential area of investigation unique to enzyme/IL simulations and so we briefly discuss here common methods of analysis applied to substrate/IL systems. Similar to the analysis methods for studying enzymes in ILs, many studies of substrates in ILs have focused on determining the solvent's structural and spatial distribution around the substrate to, for example, glean mechanistic aspects of cellulose solvation and dissolution in ILs.[186-191] Thermodynamic analysis of the cellulose solvation process in ILs has revealed an entropic and energetic driving force for cellulose dissolution in ILs [189, 192], and more recently published papers have identified key structural and thermodynamic differences of bundles of cellulose versus individual chains in ILs.[193-196] New and interesting methods of analysis include analyzing the induced changes in the conformational free-energy landscapes of sugar rings by ILs [169], and performing dynamical calculations of mean lifetimes of different binding states of IL anions with hydroxyl groups on cellulose [194].

Table 4.4. Common methods of analysis for enzyme/IL systems.

Method	What is calculated	What it is used for
Root mean square deviation (RMSD)	Mean deviation from a reference structure (usually the xray structure)	Assessing conformational changes and protein unfolding
Root mean square fluctuation (RMSF)	Residue-specific fluctuations averaged over portions or entire simulation	Assessing protein dynamics and backbone flexibility
Secondary structure	α-helix or β-sheet content	Assessing protein stability and loss or gain of structural motifs
Intra/intermolecular hydrogen bonds (h-bonds)	Number of hydrogen bonds between the enzyme and anions, enzyme and water in IL/water solutions, or between residues within the enzyme	Assessing protein stability and solvation in ILs and transient behavior of key non-bonded interactions

Perspective: Challenges and Future Directions

As we look to the future of molecular simulations of biocatalysts in ionic liquids, a hurdle that will need to be overcome is the dearth of experimental data on enzyme/IL systems with which to validate computational models. Part of this problem may be that IL solvents often bring new challenges with standard experimental methods (e.g., NMR, CD, crystallography). The importance of structural information is critical; experimental studies have played an invaluable role in continuing to refine modern force fields for protein simulations (e.g., the identification in 2008 that many force fields incorrectly predicted alpha helicity [197]). The absence of this data has been noted in the literature [153, 155, 157], and many groups have instead taken to comparing their simulation results to experimental data via indirect or qualitative measures. For example, Kim and coworkers found a direct linear correlation between experimentally observed reaction rates of a butyl acetate synthesis reaction and calculated (i.e. simulated) non-bonded interaction energies between different IL solvents and the CALB enzyme that catalyzed the reaction.[145] It has recently been suggested that the following experimental techniques could possibly be used to validate or invalidate computational results of enzymes in ILs: isothermal titration calorimetry to energetically describe substrate binding/unbinding to an enzyme in the presence of ILs, x-ray scattering or small-angle neutron scattering to observe large-scale enzyme structural changes, and Fourier transform infrared spectroscopy to monitor the evolution of enzyme secondary structure.[155]

Another challenge of studying biocatalysis in ILs with computation is related to sampling. Due to strong interactions existing between the enzyme surface and IL ions (specifically anions), as well as the intrinsic properties of the solvent that lead to the previously discussed slow dynamics of many ILs, adequate sampling can sometimes be difficult to achieve.[157] In turn, this can lead to complicated and potentially inaccurate calculations of equilibrium thermodynamic and/or dynamic system properties. The use of enhanced sampling techniques like parallel tempering, umbrella sampling, or metadynamics, has been proposed as a solution to this issue.[157] Using special-purpose machines for MD simulations (e.g. Anton [198]) or hardware accelerators using platforms such as FPGA, Graphics Processing Units (GPU), the Cell Broadband Engine (CBE) and multi-core processors [199] is another potential solution.

The aforementioned areas for future focus and the proposed use of emerging hardware solutions to improve computational studies of biocatalysis in ILs highlight the potential growth of the field and the vast space for exploration and discovery. The field is growing in leaps and bounds every year and we look forward to seeing the exciting new directions researchers take to inspire new experimental work and improve the predictive capabilities of enzyme/IL simulations.

Chapter 5

The General AMBER Force Field (GAFF) Can Accurately Predict Thermodynamic And Transport Properties Of Many Ionic Liquids[5]

Introduction

Ionic liquids (ILs) are a class of solvents consisting of organic salts that are liquid below 100°C. Many ILs possess thermodynamic and electrochemical properties that make them optimal candidates for a large range of applications.[200-201] For example, ILs tend to be thermally stable, noncombustible, and nonvolatile, making them more environmentally favorable compared to common organic solvents and thus increasingly used in green processes such as the separation of organic solvents from ILs by supercritical CO_2. Cellulosic and algal biomass can be solvated and preprocessed by ILs for biofuel production. ILs have also been used to selectively absorb gases and later release them in temperature dependent processes. Because of the favorable electrochemical window, ILs have been used as a replacement for water in certain batteries. Understanding these interactions at the molecular level can help researchers optimize the species of IL used for a given device or process, and it can also guide the discovery/synthesis of new ILs.

Molecular simulation, in particular Monte Carlo (MC) or molecular dynamics (MD) simulations provide a convenient route to predict thermophysical properties of ILs. The accuracy of these predictions relies on the accuracy of the force field used to describe the intra and intermolecular (bonded and non-bonded) interactions of the solvent molecules. Several force fields have been optimized to reproduce experimental data for small organic molecules, both synthetic and biological. Among these force fields are OPLS[202], CHARMM[203], UFF[204], AMBER and its companion the general AMBER force field[182] (GAFF). Since many ionic liquids are similar in structure to some of the molecules used to parameterize these popular force fields, they have

[5]Reproduced in part with permission from K.G. Sprenger, Vance Jaeger, and J. Pfaendtner. The General AMBER Force Field (GAFF) Can Accurately Predict Thermodynamic And Transport Properties Of Many Ionic Liquids. The Journal of Physical Chemistry B, 119: 5882-5895, 2015. Copyright 2015 American Chemical Society.

commonly been used as templates for the creation of IL force fields, both in functional form and actual parameter values. However, most of these force fields have been developed for only a small subset of ILs. Inspired by prior successes, we sought to perform a comprehensive evaluation of a common force field (GAFF) in order to determine how well standard molecular simulation force fields perform for a new class of solvents (i.e. ILs) for which the parameters were not optimized. Though GAFF has been widely used in IL simulations for years[205-208], it is precisely this widespread usage that motivated our investigation, because a systematic study of the accuracy of GAFF across a wide spectrum of ILs has not, to our knowledge, yet been performed. We note that while only MD simulations were performed in this study, the force fields generated using this method are not specific to MD, but can also be used in MC simulations.

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

The transport properties of ILs are not described well by classical force fields when full $+e/-e$ charges are assigned to the cations and anions.[172] The electronic point charges are commonly derived from quantum calculations in vacuum. It has been shown, both computationally and experimentally, that this description of a cation or anion in a vacuum with a full charge does not accurately describe the liquid system.[172-174] In a liquid, neighboring molecules can screen charge, and the electron density of a molecule can be affected by nearby molecules of the opposite charge. Because of this, previous researchers have suggested that classical force fields should assume that the charge of the cation and anion is actually some fraction of $+e/-e$.[209] Maginn and coworkers have suggested and compared several methods to scale the ions' charges, and they have found that scaling is necessary to predict the shear viscosity or self-diffusivity of an IL using MD.[172] To overcome the shortcomings of calculating the electronic structure *in vacuo*, *ab initio* MD can be used with crystal structures of solidified ILs in order to scale the charges and very accurately reproduce the properties of liquid phase ILs. A more naïve approach is to simply scale the charges by multiplying by a factor of 0.7-0.9[206, 209-212], as this is near what is typically found in the *ab initio* methods. While this approach sacrifices some accuracy, it is far more amenable to large-scale deployment of molecular simulation for predicting IL properties.

Ultimately, if IL force fields can be constructed quickly and easily without refinement, screening for the optimal ILs for certain applications could be accomplished with greater efficiency, and complex simulations involving ILs could be performed with greater ease. For example, simulating sugars, proteins, and enzymes in ILs and IL/water mixtures is becoming an increasingly relevant area of research, supported by many recently published papers by us and others.[140-141, 143, 157, 169, 213-214] An even more interesting possibility is that we will be able to predict the thermophysical properties of yet undiscovered ILs or of complex interactions between solutes and ILs. These latter simulations could help to screen a large library of ILs and direct chemists in their efforts to discover useful cation-anion pairs. Other protocols have been proposed for predicting

electrostatic point charges and properties of ILs.[172] These protocols require more rigorous calculations to develop the force fields. The approach we adopt herein, even though it could result in a modest loss of accuracy, can be implemented by a researcher with a basic understanding of standard MD using widely available, and free, programs. Specifically in this paper we propose a simple and fast protocol for accurately predicting the thermophysical properties of ILs using GAFF.

An extensive series of molecular dynamics simulations has been conducted featuring 19 combinations of ionic liquids at room conditions (298.15 K and 1 bar), including alkylammonium [RAM], 1-alkyl-3-methylimidazolium [RMIM], 1-alkylpyridinium [RPyr], and 1-alkyl-3-methylpyridinium [RMPyr] cations, and formate [HCO_2], acetate [Ac], alkylsulfate [RSO_4], bis[(trifluoromethyl)sulfonyl]imide [Tf_2N], trifluoromethanesulfonate [Otf], and chloride [Cl] anions. Calculations of IL density, heat capacity, molar enthalpy of vaporization, self-diffusivity, and shear viscosity have been computed for the ILs and compared to experimental data. Self-diffusivity calculations have also been performed at high temperatures to see the influence on the accuracy of GAFF's predictions. To further evaluate the effectiveness of the GAFF force field for ILs, the resulting densities, molar enthalpies of vaporization, and self-diffusion coefficients are compared to results from other published force fields, including a modified force field based on the OPLS-AA/AMBER framework and the CHARMM 22 force field.

Representations of the IL cations and anions are shown in Figure 5.1. As was noted earlier, many recent studies have focused on combining water and ILs to look at protein or enzyme structure and/or stability, thus additional simulations were performed on pure boxes of water using the TIP3P and SPC/E water models. This was done for two reasons: first, we surmised that the accuracy of TIP3P and SPC/E, given their widespread use, would be a useful standard with which to compare the accuracy of GAFF in predicting IL properties and second, we wished to provide a starting point for future work to assess the ability of GAFF to predict IL/water mixture properties.

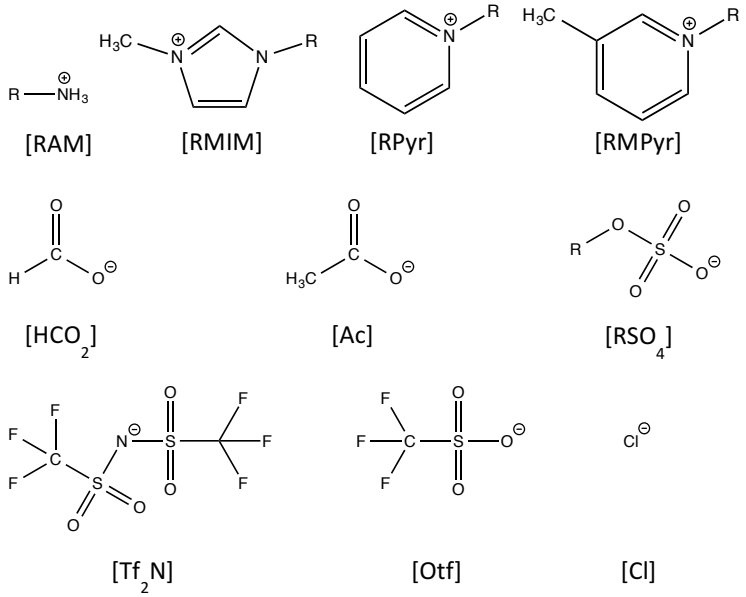

Figure 5.1. Ionic liquid ions. R = ethyl (Et or E), butyl (B), pentyl (P), hexyl (H), and octyl (O).

Methods

5.2.1 GAFF

To represent the ILs, GAFF[182] was the chosen force field, the functional form of which is shown in Eqn. 1. Despite the reputation and widespread use of AMBER as a force field to study proteins and nucleic acids, AMBER lacks the parameters necessary to study many organic molecules. Thus, GAFF was designed to improve upon this aspect of AMBER and permit the detailed study of organic molecules such as those present in ILs in conjunction with studies of biomolecules. In a similar fashion to AMBER, bond stretching and angle bending terms in GAFF are evaluated with a simple harmonic functional form and the dihedral term is evaluated with a cosine function. Coulombic and 12-6 Lennard-Jones terms are used to calculate the electrostatic and van der Waals non-bonded interactions, respectively, considering all pairs of atoms (*i* and *j*) in different molecules or in the same molecule but separated by a minimum of 3 bonds. One-four interactions, or non-bonded interactions separated by 3 bonds, are reduced by a scale factor, in order to achieve identical parameters for intra and intermolecular interactions. Parameters include the force constants k_r, k_θ, and V_n, the equilibrium structural parameters r_{eq} and θ_{eq}, the partial atomic charges q_i and q_j, and the Lennard-Jones well depth, ε, and radii, A_{ij} and B_{ij}; *n* is multiplicity and λ is a phase angle for torsional angle parameters.

$$V_{tot} = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\phi - \lambda\right] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \qquad (5.1)$$

Quantum Mechanics Calculations

Individual quantum mechanics (QM) calculations were performed on all ions, i.e. [RAM] (R = ethyl (E), butyl (B), and pentyl (P)), [RMIM] (R = E, B, hexyl (H), and octyl (O)), [RPyr] (R = P and H), and [RMPyr] (R = E) cations, and [HCO_2], [Ac], [RSO_4] (R = methyl (Me) and ethyl (Et)), [Tf_2N], [Otf], and [Cl] anions. Geometry optimization and energy calculations were performed at the Hartree-Fock (HF) level of theory using the 6-31G(d)//6-31G(d) basis set and the electronic structure program Gaussian.[215] All force constants and equilibrium and Lennard-Jones parameters were taken directly from the GAFF force field. Antechamber[216] was used to assign electrostatic point charges by the RESP method.[217] After the determination of partial charges, tleap[218] and ACPYPE[179] were used to generate input files for each ion. Atomic charges were scaled by multiplying each by a factor of 0.8. These methods are very similar to some of those employed by Maginn and coworkers.[206, 219] This work aims to determine how universal this approach is for ILs.

Ionic Liquid Simulations

Cubic boxes of 125 nm^3 in size and containing 200 to 1000 ion pairs were generated using Packmol[178] at close to experimental liquid densities. IL boxes were simulated with periodic

boundary conditions and used the GROMACS[220-221] 4.6.5 package and GAFF force field. Under 1.4 nm, electrostatics were calculated explicitly; above 1.4 nm, particle-mesh Ewald (PME) summations were used to calculate long-range electrostatic interactions. Van der Waals interactions were shifted to 0 at 1.2 nm, in order to eliminate artifacts. All bonds between hydrogen and heavy atoms were frozen at equilibrium values.

IL ions were placed randomly in the box to carry out the liquid-phase simulations. Following energy minimization, each of the 19 ILs was simulated in the NVT ensemble for 10 ns, at 500 K, to achieve thorough mixing of the ions and to eliminate any artifacts in the initial configurations, thus establishing coordinates independent of the Packmol input. Constant temperature was maintained with a global stochastic thermostat[54].

From these 10 ns NVT simulations, unique starting structures for subsequent simulations were taken from points along the simulation trajectories at 2, 4, 6, 8, and 10 ns. For each IL, the 5 unique starting structures then underwent 5 ns of pressure equilibration in the NPT ensemble, at room conditions (298.15 K and 1 bar), using the Berendsen barostat.[109] The boxes then underwent a further 10 ns of pressure equilibration in the NPT ensemble at the same conditions, using a timestep of 1 fs and the Parrinello-Rahman barostat[222] to accurately capture the correct ensemble fluctuations for calculations of the densities and heat capacities. The reported density and heat capacity values represent the values averaged across the 5 simulations.

One of the 5 simulation boxes for each IL was used to initiate triplicate non-equilibrium MD NPT simulations, 1 ns in length, wherein a cosine acceleration factor, or shear rate, of 0.03 nm/ps^2 was applied to each particle. This periodic perturbation method was used to calculate the shear viscosity of the ILs.[223] NPT simulations of length 10 or 20 ns depending on the IL (see Results and Discussion) were run to calculate the self-diffusivity of the ions of 4 ILs.

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

Results and Discussion

An overall picture of the results for IL density, heat capacity, molar enthalpy of vaporization, and shear viscosity is shown via parity plots in Figure 5.2; results will later be discussed in detail.

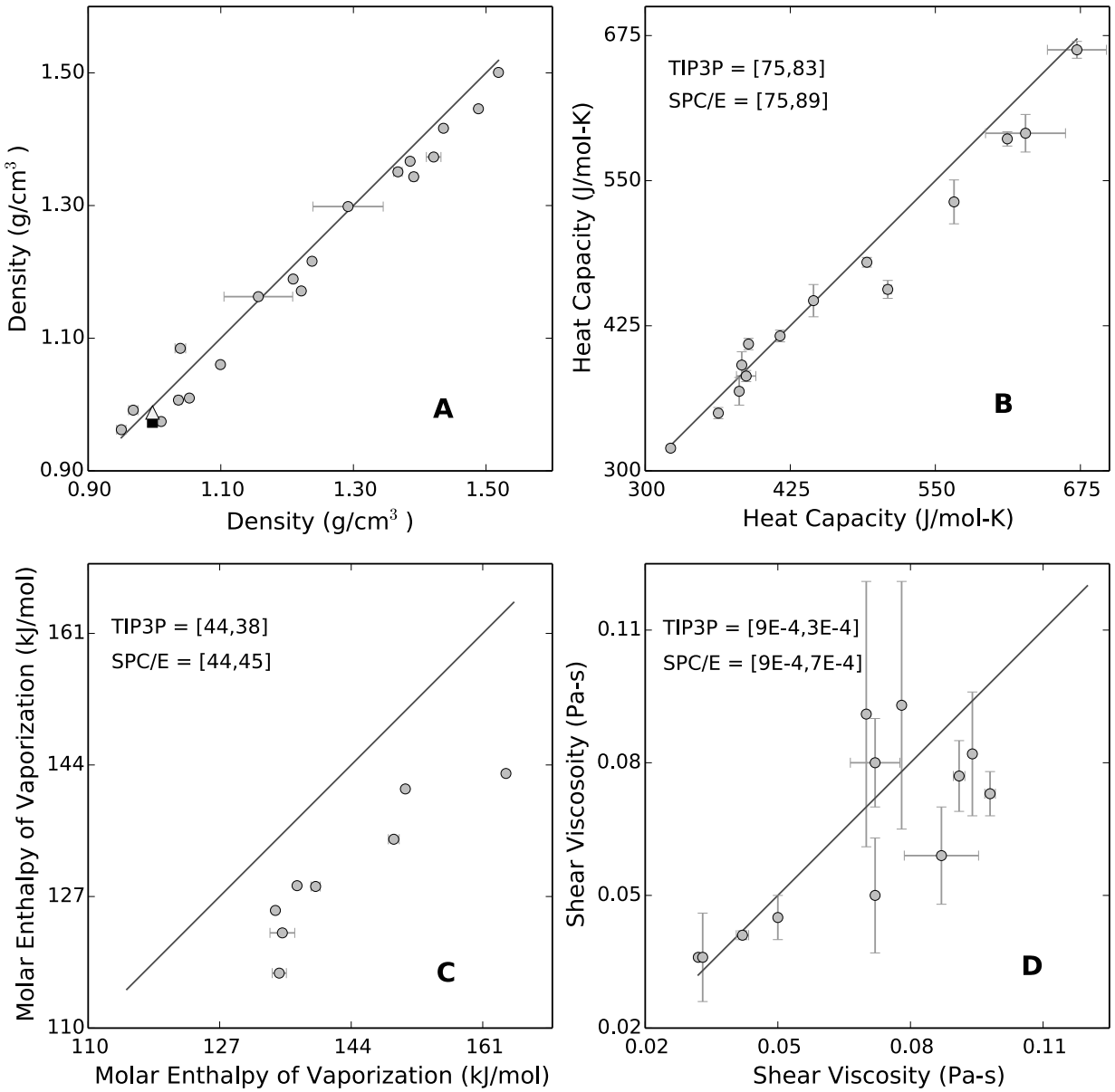

Figure 5.2. Simulated vs. experimental IL properties: A) density (g/cm^3), square = TIP3P water, triangle = SPC/E water; B) heat capacity (J/mol-K); C) molar enthalpy of vaporization (kJ/mol); and D) shear viscosity (Pa•s). The x and y axes show experimental and simulated results, respectively. Brackets show results for TIP3P and SPC/E water with experimental results listed first.

5.3.1 Density

Using the equilibrium volume of the simulation cell, we computed the density of the ILs using the standard definition. Computed and experimental densities of the 19 ILs and TIP3P and

SPC/E water are reported in Table 5.1. When multiple experimental data points were available, an average of the values was taken to get the single reported experimental value in Table 5.1.

Table 5.1. Experimental and simulated IL densities (g/cm^3). Values in parentheses represent simulated densities from Vega and Abascal.[224]

IL	Experiment	Simulation	Error (%)
[BAM][HCO_2]	0.968[225]	0.992	2.4
[BMIM][Ac]	1.053[226-228]	1.010	-4.1
[BMIM][$MeSO_4$]	1.209[229-233]	1.189	-1.6
[BMIM][Tf_2N]	1.436[234-238]	1.416	-1.3
[EAM][HCO_2]	1.039[225]	1.085	4.4
[EMIM][Ac]	1.099[239-241]	1.060	-3.6
[EMIM][$EtSO_4$]	1.238[242-245]	1.216	-1.7
[EMIM][Otf]	1.386[231, 239, 246-247]	1.367	-1.4
[EMIM][Tf_2N]	1.519[236, 244, 248]	1.501	-1.2
[EMPyr][$EtSO_4$]	1.221[249-251]	1.171	-4.1
[EMPyr][Tf_2N]	1.489[252]	1.446	-2.9
[HMIM][Cl]	1.036[253-256]	1.007	-2.8
[HMIM][Tf_2N]	1.367[240, 257-260]	1.351	-1.2
[HPyr][Tf_2N]	1.391[261-262]	1.343	-3.4
[OMIM][Cl]	1.010[230, 253-255, 263-264]	0.974	-3.5
[OMIM][Otf]	1.157[265-266]	1.163	0.5
[OMIM][Tf_2N]	1.292[258, 267-268]	1.298	0.5
[PAM][HCO_2]	0.950[225]	0.962	1.3
[PPyr][Tf_2N]	1.421[269]	1.373	-3.4
Mean Absolute Error			2.4
Max Absolute Error			4.4
Water Model	**Experimental**	**Simulation**	**Error (%)**
TIP3P	0.997[224]	0.973 (0.980[224])	-2.4
SPC/E	0.997[224]	0.988 (0.994[224])	-0.9

A parity plot showing overall trends in the data across the full spectrum of simulated densities is shown in Figure 5.2A. Uncertainty in the experimental data, when possible to assess, is drawn in the plot with error bars representing the standard deviation of all experimental or simulation data. We note that some data points are based upon a single experimental source and thus no error bars are available. Simulated error bars are too small to be seen due to the good reproducibility of the simulations. In general, GAFF appears to slightly underestimate the densities of most of the 19 ILs; however, the same is true of TIP3P water with excellent agreement seen between SPC/E and experiment.

For the five points that lie above the parity line in Figure 5.2A, i.e. overestimate the density, three of the points correspond to the three ILs with [RAM] cations. In contrast to most other experimental densities, we note that only a single source could be found in the literature for the experimental densities of the [RAM]-based ILs. Interestingly, the simulated density error of these

specific ILs appears to decrease with increasing chain length, from 4.4% (C_2) to 2.4% (C_4) to 1.3% (C_5), and the same trend is also mildly seen for the [Tf_2N]-based ILs, from 1.2% (C_2) to 0.5% (C_8). This is not surprising given the extensive use of hydrocarbon-containing compounds in the development of GAFF. The other two points that lie just above the parity line represent two of the three [OMIM]-based ILs. Notably, these two ILs have the largest experimental standard deviations of the 19 simulated ILs (see Figure 5.2A), however they also represent the ILs with the smallest deviation in simulated density from the reported experimental values (0.5% for both). The largest deviation from experimental values occurs for [EAM][HCO_2] (4.4%), followed closely by [EMPyr][$EtSO_4$] and [BMIM][Ac] (4.1% for both ILs).

The trend of underestimating IL density is not unique to the GAFF force field. Simulated densities for 33 [RMIM]-based ILs, computed using a refined OPLS-AA force field[165], are compared against results of this work and to experimental results in the following parity plot.

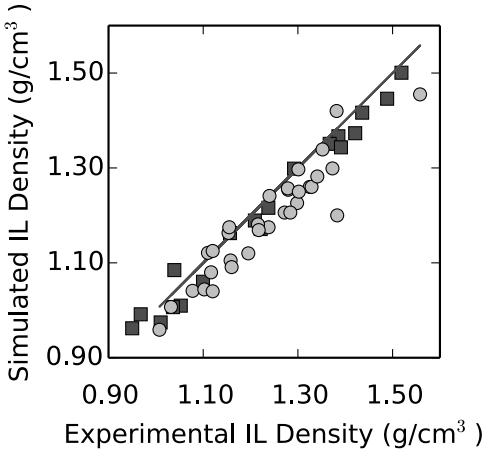

Figure 5.3. Simulated vs. experimental IL densities for this work (black squares) and the work of Sambasivarao and Acevedo[165] using a refined IL-specific OPLS-AA force field (gray circles).

The majority of the points, as with GAFF, lie below the parity line. Contrary to GAFF, however, increasing the chain length of the cation generally leads to larger deviations from experiment; since the OPLS-AA force field was parameterized for types of molecules such as carbohydrates, rather than hydrocarbons, this makes sense.[270] For estimating densities, these results clearly show that GAFF, with no tuned parameters, provides results of comparable accuracy to an IL-specific force field.

5.3.2 Heat Capacity

Constant pressure heat capacities for the ILs were calculated using the classical statistical mechanical definition,

$$C_p = \left(\frac{\left\langle (\partial H)^2 \right\rangle}{nk_B T^2} \right)_P = \left(\frac{\left\langle H^2 \right\rangle - \left\langle H \right\rangle^2}{nk_B T^2} \right)_P \tag{5.2}$$

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

Simulated heat capacities for the 19 ILs and corresponding experimental values (averaged across the specified experimental sources) for 14 of the 19 ILs are reported in Table 5.2; results for TIP3P and SPC/E water are also reported in Table 5.2. These results are also graphically displayed in Figure 5.2B, which shows the simulated and experimental error of each data point (note that some points have no experimental error bar due to a single source, and that for some points, simulation errors are too small to be seen). Error bars represent the standard deviation of all experimental or simulation data.

Table 5.2. Experimental and simulated IL heat capacities (J/mol-K). Values in parentheses represent simulated heat capacities from Vega and Abascal[224].

IL	Experiment	Simulation	Error (%)
[BAM][HCO_2]	N/A	258	N/A
[BMIM][Ac]	383[271]	391	2.1
[BMIM][$MeSO_4$]	416[231]	416	0.0
[BMIM][Tf_2N]	566[238, 272-273]	532	-6.1
[EAM][HCO_2]	N/A	185	N/A
[EMIM][Ac]	322[239]	320	-0.7
[EMIM][$EtSO_4$]	387[231, 274-275]	382	-1.4
[EMIM][Otf]	363[239, 276]	350	-3.6
[EMIM][Tf_2N]	509[248, 277-278]	456	-10.3
[EMPyr][$EtSO_4$]	389[279]	409	5.2
[EMPyr][Tf_2N]	491[280]	480	-2.3
[HMIM][Cl]	381[281]	369	-3.3
[HMIM][Tf_2N]	628[275, 279, 282-283]	591	-5.9
[HPyr][Tf_2N]	612[279]	586	-4.2
[OMIM][Cl]	445[281]	447	0.4
[OMIM][Otf]	N/A	572	N/A
[OMIM][Tf_2N]	672[272, 279]	663	-1.4
[PAM][HCO_2]	N/A	289	N/A
[PPyr][Tf_2N]	N/A	561	N/A
Mean Absolute Error			3.6
Max Absolute Error			10.3
Water Model	Experiment	Simulation	Error (%)
TIP3P	75[224]	83 (78[224])	9.9
SPC/E	75[224]	89 (87[224])	18.7

Similar to density, GAFF slightly underestimates heat capacities of the majority of the ILs. It appears, however, that GAFF more accurately estimates heat capacities that lie within the lower half of the experimental range of heat capacities of the 14 ILs. Incidentally, it can be pointed out that the 4 most underestimated heat capacities, all in the upper range of Figure 5.2B, represent ILs with fluorinated anions, specifically [Tf_2N]. This underestimation appears to lessen quite

drastically with increasing chain length, from 10.3% (C_2) to 1.4% (C_8). As was noted earlier in regards to density, this result is not unusual in light of the fact that GAFF was parameterized with a significant fraction of hydrocarbon-containing compounds and increasing the alkyl chain length has the net result of reducing the impact of inaccuracies introduced from the fluorinated components.

In reality, the quantity that we calculated in this work represents only the excess, or residual portion of the heat capacity, which accounts for the intermolecular interactions in the condensed phase.[284] Conversely, the other portion of the heat capacity is an ideal gas contribution that takes into account intramolecular interactions.[284] These two contributions to the heat capacity are typically represented as shown in Eqn. 3.

$$\langle H \rangle = \langle H^{ig} \rangle + \langle H^{res} \rangle = \langle \Phi^{\text{int}} + K + Nk_BT \rangle + \langle \Phi^{nb} + PV - Nk_BT \rangle \tag{5.3}$$

Typically, ideal gas portions of heat capacities can be found through experiments, but as this type of experimental data is currently not available for ILs, the values are calculated from a frequency analysis of optimized cation and anion structures from ab initio MD simulations.[285] Since GAFF appears to give good agreement with experimental heat capacities across a broad range of ILs using the quick screening method we are proposing here, we conclude that the residual heat capacity is the dominant term and thus deemed it unnecessary to compute the ideal gas portions of the heat capacities.

5.3.3 Molar Enthalpy of Vaporization

Experiments taking place over the last decade or so have served to refute the idea of ILs as nonvolatile liquids, and on the contrary have proved that to some extent ILs can be distilled.[286] Because of this revelation, there has been a push to understand the nature of ILs in the vapor phase, and thus to be able to accurately model and even predict molar enthalpies of vaporization of ILs is of growing importance.

Previous studies have shown that ILs most likely transition into the gas phase via single, neutral ion pairs (i.e. one cation and one anion).[284] Thus, as was described in the Methods section, the IL molar enthalpies of vaporization were calculated by simulating a single cation/anion pair in the gas phase. The equation for calculating the molar enthalpy of vaporization is shown below,

$$\begin{aligned} \Delta h_{vap}(T,P) &= \langle h_{gas}(T) \rangle - \langle h_{liquid}(T,P) \rangle \\ &= \left[\langle u_{gas}(T) \rangle + RT \right] - \left[\langle u_{liquid}(T,P) \rangle + Pv \right] \\ &\approx \left[\langle u_{gas}(T) \rangle + RT \right] - \left[\langle u_{liquid}(T,P) \rangle \right] \\ &= \Delta u_{vap}(T,P) + RT \end{aligned} \tag{5.4}$$

where Chevron brackets represent ensemble averaged molar quantities. The gas is assumed to be ideal, and Pv contributions (i.e. pressure-volume work, where P is the pressure and v is the molar volume) in the liquid can be considered negligible when compared to the internal energies and thus can be neglected. Additionally, the liquid internal energies (u_{liquid}(T,P)) were divided by the

total number of ion pairs in the system, since the desired final quantity was the enthalpy of vaporization (Δh_{vap}) for a single ion pair.

Due to limited experimental data, experimental molar enthalpies of vaporization and results from this work for only 8 of the 19 ILs are reported in Table 5.3, as well as in a parity plot in Figure 5.2C; simulation error bars are too small to be seen, and multiple experimental sources could only be found for the four [Tf_2N]-based ILs, thus some points have no experimental error bars. Error bars represent the standard deviation of all experimental or simulation data. For the [Tf_2N]-based ILs in Table 5.3, the reported experimental value is an average of those from the specified sources. Also reported in Table 5.3 is data for TIP3P and SPC/E water, as well as molar enthalpies of vaporization of the [RMIM][Tf_2N] ILs from 2 other simulation groups[287-288] for comparison to the results of this work.

Table 5.3. Experimental and simulated IL molar enthalpies of vaporization (kJ/mol). Values in parentheses represent simulated molar enthalpies of vaporization from Santos et al.[287], Kelkar and Maginn[288], and Vega and Abascal[224].

IL	Experiment	Simulation	Error (%)
[EMIM][Tf_2N]	135[289-290]	117 (159[287], 146[288])	-13.0
[BMIM][Tf_2N]	135[289-290]	122 (174[287], 151[288])	-9.5
[HMIM][Tf_2N]	139[289-290]	128 (184[287], 157[288])	-8.0
[OMIM][Tf_2N]	150[289-290]	134 (201[287], 162[288])	-10.1
[PPyr][Tf_2N]	134[291]	125	-6.7
[HPyr][Tf_2N]	137[291]	128	-6.3
[EMIM][$EtSO_4$]	164[289]	143	-12.9
[OMIM][Otf]	151[289]	141	-6.7
Mean Absolute Error			9.1
Max Absolute Error			13.0
Water Model	Experiment	Simulation	Error (%)
TIP3P	44[224]	38 (42[224])	-12.9
SPC/E	44[224]	45 (49[224])	2.4

Our results show that GAFF is able to calculate molar enthalpies of vaporization of the 8 ILs with fair agreement to experiment, with the largest percent error occurring for [EMIM][Tf_2N] (-13.0%; note this does not exceed the error of the TIP3P water model). Additionally, at least in the case of the 4 [RMIM][Tf_2N] ILs, GAFF predicts the molar enthalpies of vaporization with similar or better accuracy than the computed values of other published studies that used parameters from a force field based on the AMBER/OPLS-AA framework extended to [RMIM][Tf_2N] ILs[287], and the CHARMM 22 force field[288], illustrated in Figure 5.4.

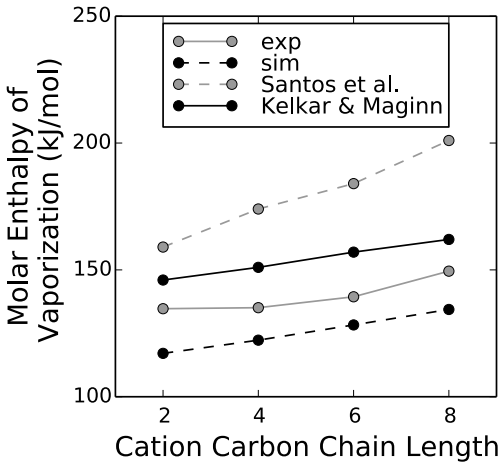

Figure 5.4. Simulated vs. experimental IL molar enthalpies of vaporization for [RMIM][Tf_2N] ILs (R=E, B, H, and O). Simulation results are also shown from the work of Santos et al.[287], using a refined OPLS-AA force field, and Kelkar and Maginn[288], using the CHARMM 22 force field.

For the 6 [Tf_2N]-based ILs in Table 5.3, GAFF captures the well-documented trend, both experimentally and computationally, of increasing vaporization molar enthalpy with increasing cation carbon chain length. For the four [Tf_2N]-based ILs, our results, in agreement with those of Kelkar and Maginn[288], show around a 5-6 kJ/mol increase in the molar enthalpy of vaporization for each 2-carbon increase in cation chain length. This trend also generally agrees with the experimental results, though virtually no increase in the molar enthalpy of vaporization is seen with an increase in carbon chain length from C_2 to C_4, experimentally, and a larger increase (10 kJ/mol) accompanies a jump from a length of C_6 to C_8. Both past simulations and experiments have proposed that these increases are due to additional van der Waals interactions brought about by the increasing chain lengths.[287-288] Given the significantly increased difficulty of accurately predicting this property over others such as density or heat capacity, in conjunction with the use of an unrefined force field, the errors yielded by GAFF in predicting the molar enthalpies of vaporization of these particular ILs are not unreasonable; however, it is recommended that for the property of molar enthalpy of vaporization, as well as for self-diffusivity as it turns out (see next subsection), one should use this method to merely draw qualitative trends rather than as a means to obtain accurate predictions.

5.3.4 Self-Diffusivity

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

The Einstein relation shown in Eqn. 5 was used to compute the self-diffusivity of the cations and anions through calculation of the mean square displacement, or MSD, of the ion centers-of-mass,

$$D = \frac{1}{6}\lim_{t\to\infty}\frac{d}{dt}\left\langle \sum_{i=1}^{N}\left[\vec{r}(t)-\vec{r}(0)\right]^2 \right\rangle \tag{5.5}$$

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

As with the molar enthalpy of vaporization, limited experimental data led to our decision to calculate the self-diffusivity of only the 4 [RMIM][Tf_2N] ILs. Computed and experimental self-diffusivities for the cation and anion of each of the ILs are reported in Table 5.4, along with self-diffusivities from the work of Köddermann and Paschek[292] for later comparison and TIP3P and SPC/E water data. All production simulations were run in triplicate for 10 ns except where later noted and fitting for calculation of the self-diffusion coefficients was done over the range of 1-9 ns. For ease of visualizing the trends, the data is also graphically displayed in Figures 5.5 and 5.6; error bars represent the standard deviation of all experimental or simulation data, and simulation error bars are generally too small to be seen, whereas points lack experimental error bars due to the availability of only a single experimental source.

Table 5.4. Experimental and simulated IL self-diffusivities (10^{-11} m^2/s). Simulated self-diffusivities from Köddermann and Paschek[292] and Vega and Abascal[224] are included for comparison.

Il	Cation Exp[293]	Cation Sim	Error (%)	Anion Exp[293]	Anion Sim	Error (%)
[EMIM][Tf_2N]	4.98	2.29	-53.9	3.10	1.11	-64.1
	5.88[a]	6.00[a,292]	+2.0	3.70[a]	2.30[a,292]	-37.9
[BMIM][Tf_2N]	2.76	1.13	-51.8	2.20	0.83	-62.2
	3.37[a]	3.00[a,292]	-11.0	2.69[a]	2.20[a,292]	-18.4
[HMIM][Tf_2N]	1.69	0.87	-48.7	1.54	0.61	-60.6
	2.11[a]	1.30[a,292]	-38.5	1.93[a]	1.20[a,292]	-38.0
[OMIM][Tf_2N]	1.17	0.41	-65.1	1.17	0.31	-73.4
	1.48[a]	0.50[a,292]	-66.3	1.48[a]	0.48[a,292]	-67.5
Mean Abs Error[b]			54.9			65.1
			29.5			40.4
Max Abs Error[b]			65.1			73.4
			66.3			67.5
Water Model	Exp	Sim	Error (%)			
TIP3P	230[224]	586 (549[224])	154.8			
SPC/E	230[224]	264 (254[224])	14.7			

[a]Evaluated at 303 K and 1 bar
[b]Evaluated first for the data from this work, then for the data from Köddermann and Paschek[292]

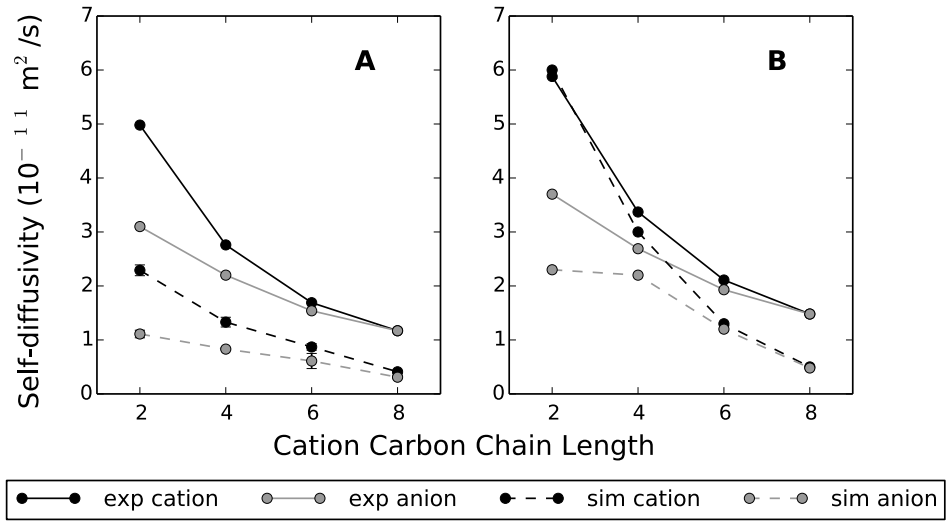

Figure 5.5. Simulated vs. experimental IL cation and anion self-diffusivities for [RMIM][Tf_2N] ILs (R=E, B, H, and O): A) simulation results of this work compared to experimental values[293] at 298.15 K and 1 bar; B) simulation results from Köddermann and Paschek[292] compared to experimental values[293] at 303 K and 1 bar.

The results show that GAFF is able to accurately capture the trend of decreasing cation/anion self-diffusivity with increasing cation chain length, and even to some extent, the relatively large nonlinear trend in self-diffusivity of the cation in moving from a cation chain length of C_2 to C_4. The trend of decreasing cation/anion self-diffusivity with increasing cation chain length has been observed by others and can be explained by an increased steric hindrance brought on by longer cation chain lengths that serve to slow down the diffusion of the cations through the IL.[285] GAFF is also able to capture the interesting trend of the anion self-diffusivity tracking that of the cation, a trend that has been observed by Cadena et al.[285] In other words, our results show that the self-diffusivity of the [Tf_2N] anion is highest when paired with the fastest cation, [EMIM], and that it is lowest when paired with the slowest cation, [OMIM]; as others have noted, this suggests that cations and anions diffuse through the liquid in pairs or clusters, rather than on their own.[285] Our results also show that for the 4 IL cation chain lengths tested, the self-diffusivity of the cation at 298.15 K is always higher than that of the anion, though studies show this is not the case for all ILs.[285] From a quantitative perspective, deviations in cation/anion self-diffusivity from experiment range from 49-73%. There is consistently a greater deviation from experiment of the simulated anion self-diffusivities than of the simulated cation self-diffusivities, with the largest percent error occurring for [Tf_2N] in [OMIM][Tf_2N].

Using a force field based on the OPLS-AA/AMBER framework and specifically modified to reproduce densities and self-diffusion coefficients of ions of [RMIM][Tf_2N] ILs, molecular dynamics simulations were carried out by Köddermann and Paschek[292] to calculate self-diffusion coefficients at 303 K and 1 bar. A similar method as was used as in this work was employed to perform the actual diffusion calculations. As Table 5.4 shows, the IL-refined force field results are much more accurate than the GAFF results at the lower cation chain lengths of 2 and 4 carbons. As the cation chain length increases to 8, however, the error gap between the refined OPLS and GAFF force fields becomes nearly negligible, easily visible in Figure 5.5. Overall, there is about

a 25% increase in the mean absolute error for the GAFF cation and anion data sets over the IL-refined OPLS data sets. Another interesting and informative way to compare the simulation data against each other and to experiment is by plotting the ratio of the cation to anion self-diffusivity as a function of cation carbon chain length. This is shown in Figure 5.6 below, where gray markers represent the experimental data sets at 298.15 K and 303 K (the ratios are the same at these two temperatures), and black markers represent the two simulation data sets at 298.15 K and 303 K. As Figure 5.6 shows, the ratios of cation to anion self-diffusivity generated with the IL-refined OPLS force field are, in every case except for [EMIM][Tf_2N], closer to experimental ratios than results generated with GAFF, though GAFF captures the trend of decreasing cation to anion self-diffusivity ratios with increasing cation chain lengths very well.

Considering the OPLS force field is specifically refined for this family of ILs and for the property of self-diffusion, and that other groups using this same force field without the property refinement saw diffusion results off by at least an order of magnitude from experimental values[294-295], it is surprising that simply using an unrefined force field such as GAFF with scaled atomic charges yields such relatively accurate results. However, as was suggested earlier in regards to the molar enthalpy of vaporization, it is perhaps inadvisable to use this method for the purpose of obtaining accurate predictions of self-diffusion coefficients. Rather, we suggest using this method as a way to obtain meaningful qualitative trends of self-diffusivity, or for the purposes mentioned in the Conclusion such as a starting point for force field refinement.

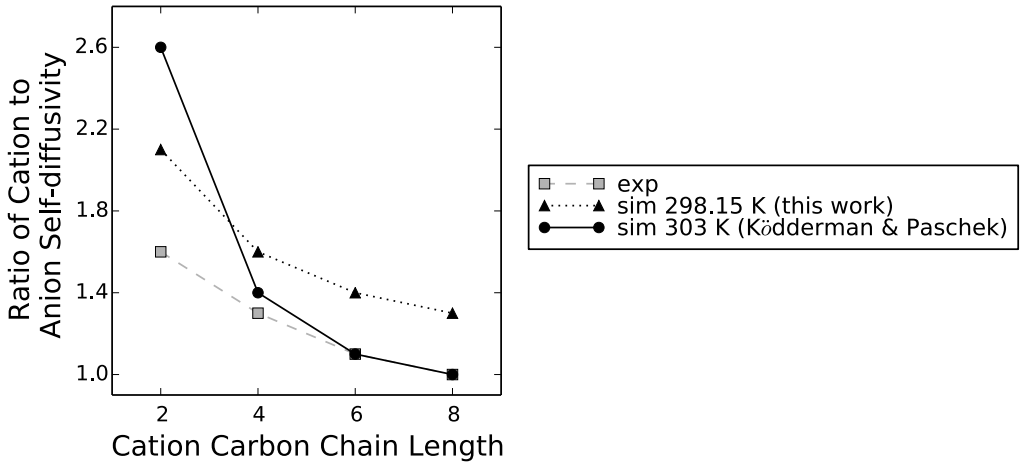

Figure 5.6. Simulated vs. experimental[293] cation to anion self-diffusivity ratios for [RMIM][Tf_2N] ILs (R=E, B, H, and O). Simulation results from Köddermann and Paschek[292] are included for comparison.

When considering the quantitative and qualitative accuracy of the results and trends of IL self-diffusivity, respectively, it is important to consider the ability of each IL to reach diffusive behavior over the timescale of the simulations. In a review article on computing IL properties with simulation, Maginn suggests computing the time-dependent quantity beta, or $\beta(t)$, as a way to measure whether a system is in the diffusive regime.[284] The equation to calculate $\beta(t)$ is shown below,

$$\beta(t) = \frac{d\log(\Delta r^2)}{d\log(t)} \tag{5.6}$$

where t is the simulation time and Δr^2 is the MSD of the ion centers-of-mass, similar to Eqn. 5.

The system can be said to have reached diffusive behavior when $\beta(t)$ has reached a value of 1; values below 1 thus indicate sub-diffusive behavior.[284] The triplicate TIP3P water simulations all reached $\beta(t)$=1 after only 1 ns, while the triplicate simulations of [RMIM][Tf_2N] (R = E, B, and H) had all reached diffusive behavior after 5 ns. Neither simulation of [OMIM][Tf_2N], however, reached diffusive behavior over the 20 ns; after leveling off at around 17 ns, the maximum $\beta(t)$ value reached for either simulation was 0.6. Thus, for ILs with long-chained ions, such as [OMIM][Tf_2N], it would be advisable to extend the simulation time, or increase the simulation temperature, to have more confidence in the calculated ion self-diffusivities. Indeed, of the 4 [RMIM] cations, the greatest percent error relative to experiment is seen for [OMIM] (65%), and as was noted earlier, of the 4 [Tf_2N] anions, [Tf_2N] in [OMIM][Tf_2N] has the greatest deviation (73%).

5.3.5 Shear Viscosity

Past simulations of IL shear viscosity have roughly fallen into two categories: 1) equilibrium MD simulations that focus on the intrinsic pressure and momentum fluctuations of the system, and 2) non-equilibrium MD simulations that monitor the linear response of the system after it is driven away from equilibrium. Nearly all previous attempts using equilibrium methods have led to overestimations of IL shear viscosities by at least an order of magnitude, while calculations using non-equilibrium methods have tended to agree better with experiment and generally require significantly shorter simulation times for calculations.[284] For example, in a study performed by Hess, the shear viscosity of a Lennard-Jones fluid using the SPC and SPC/E water models was calculated using two equilibrium and two non-equilibrium methods; of the four methods, the non-equilibrium, periodic perturbation method (PPM) provided the best results.[296] Accordingly, we adopted the PPM method, first developed by Gosling et al.[223], to assess GAFF's ability to predict IL shear viscosities.

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

$$\rho\frac{\partial u_x(z,t)}{\partial t} = \rho a_x(z,t) + \eta\frac{\partial^2 u_x(z,t)}{\partial z^2} \tag{5.7}$$

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

$$a_x(z,t) = a_0\cos(kz) \qquad k = \frac{2\pi}{l_z} \tag{5.8}$$

where k is the wavenumber and l_z is the box height in the z-dimension. The acceleration profile is assumed to take this shape because it satisfies the conditions that the acceleration and velocity profiles be both periodic (since the simulations feature periodic systems) and smooth (so local shear rates are small).[296] This assumption, together with the assumption that the initial velocity $u_x(z,t=0)$ is zero, leads to the following solution to the velocity profile,

$$u_x(z,t) = a_0\tau(1 - e^{-\frac{t}{\tau}})\cos(kz) = u_0\cos(kz) \tag{5.9}$$

where τ, the system relaxation time, is given by Eqn. 10.

$$\tau = \frac{\rho}{\eta k^2} \tag{5.10}$$

At sufficiently long times, $e^{-t/\tau}$ goes to zero, and the amplitude of the steady-state velocity profile, u_0, is equal to $a_0\tau$. By fitting this steady-state velocity profile to a $\cos(kz)$ form, one can then measure u_0 and use Eqn. 11 to calculate the IL shear viscosity.

$$\eta = \frac{a_0}{u_0}\frac{\rho}{k^2} \tag{5.11}$$

As was described earlier in the Methods section, 1 ns NPT simulations were run in triplicate to calculate the shear viscosity. Simulations were run at shear rates of 0.01, 0.03, 0.05, and 1.0 nm/ps^2, and though the shear viscosity calculated using the smallest perturbation should theoretically be nearest to the true shear viscosity at the hydrodynamic limit (i.e. k=0)[297], the best overall agreement with experiments was achieved with a shear rate of 0.03 nm/ps^2. Thus, Table 5.5 shows the computed and experimental shear viscosities for 17 of the 19 ILs (experimental data at 298.15 K was unavailable for [EMPyr][Tf_2N] and [OMIM][Otf]), as well as for TIP3P and SPC/E water, at a shear rate of 0.03 nm/ps^2. Reported experimental values represent averages of the values from the specified experimental sources. The data is also displayed in a parity plot in Figure 5.2D, though the range is restricted to highlight good fits (note that some points have no experimental error bars due to a single source, and that for some points, simulation error bars are too small to be seen). Error bars shown in Figure 5.2D represent the standard deviation of all experimental or simulation data.

Table 5.5. Experimental and simulated IL shear viscosities (Pa•s) using a shear rate of 0.03 nm/ps^2. Values in parentheses represent simulated shear viscosities from Vega and Abascal[224]. Bolded ILs are not included in Figure 5.2D, which is restricted to a range that highlights good fits.

IL	Experimental	Simulation	Error (%)
[BAM][HCO_2]	0.070[225]	0.091	29.7
[BMIM][Ac]	0.349[226-227, 260, 279, ...]	0.055	-84.2
[BMIM][$MeSO_4$]	0.094[299]	0.082	-12.8
[BMIM][Tf_2N]	0.050[237, 268, 300]	0.045	-10.6
[EAM][HCO_2]	0.032[225]	0.036	11.3
[EMIM][Ac]	0.135[227, 239, 301]	0.027	-80.4
[EMIM][$EtSO_4$]	0.098[245, 260, 302-304]	0.073	-26.1
[EMIM][Otf]	0.042[239, 305]	0.041	-2.4
[EMIM][Tf_2N]	0.033[268, 306-307]	0.036	9.3
[EMPyr][$EtSO_4$]	0.156[249, 279]	0.053	-66.2
[EMPyr][Tf_2N]	N/A	N/A	N/A
[HMIM][Cl]	18.10[253]	0.380	-98.0
[HMIM][Tf_2N]	0.072[257, 260, 268, 308-...]	0.080	10.2
[HPyr][Tf_2N]	0.087[261-262, 279]	0.059	-31.7
[OMIM][Cl]	20.93[253, 303, 310]	0.419	-98.0
[OMIM][Otf]	N/A	N/A	N/A
[OMIM][Tf_2N]	0.091[268, 311-312]	0.077	-15.3
[PAM][HCO_2]	0.078[225]	0.093	19.1
[PPyr][Tf_2N]	0.072[262]	0.050	-30.2
Mean Absolute Error			17.4[a], 37.4[b]
Max Absolute Error			31.7[a], 98.0[b]
Water Model	**Experimental**	**Simulation**	**Error (%)**
TIP3P	9.0E-4[224]	3.1E-4 (3.2E-...)	-65.1
SPC/E	9.0E-4[224]	7.0E-4 (7.3E-...[224])	-22.2

[a]Excludes bolded ILs.

[b]Includes bolded ILs.

As Table 5.5 shows for the sets of [Tf_2N] and [HCO_2]-based ILs, GAFF is able to reproduce the experimental trend of increasing shear viscosity with increasing cation chain length, even despite the significantly large errors that were seen among the triplicate simulations of [BAM][HCO_2] and [PAM][HCO_2] (see Figure 5.2D). Additionally, as Figure 5.2D shows, below an experimental shear viscosity of approximately 0.1 Pa•s, GAFF is able to predict IL shear viscosity with good quantitative accuracy; over this range of experimental values, which accounts for just over 70% of the total data set, errors range from as little as 2% to a maximum of 34%. Above an experimental shear viscosity of 0.1 Pa•s (bolded ILs shown in Table 5.5), GAFF predicts shear viscosities with deviations of 66-98%; it is important to note that these deviations, however, are comparable to the 65% deviation in the simulated shear viscosity of TIP3P water compared to experiment. Nonetheless, a few similarities can be found among these 5 most deviant IL values. Recurring anions include [Cl] and [Ac]. In the case of the [Cl]-based ILs, we do not recommend

using GAFF to calculate transport properties. ILs are generally considered to be highly viscous fluids, and [Cl]-based ILs have significantly higher shear viscosities than most ILs, as Table 5.5 shows. While GAFF does correctly predict that the 2 [Cl]-based ILs have the highest viscosities of the 17 ILs at the temperature used in this study (298.15 K), the simulation dynamics of the [RMIM][Cl] ILs were visibly so slow as to imply that a glass-like state may have been reached. This may reflect that these particular systems did not come close to reaching the hydrodynamic limit, where the shear viscosity is a number rather than a function, due to too small of a box length (and thus too large of a wavenumber). Indeed, a study by Hu and Margulis on the shear viscosity of [HMIM][Cl], using OPLS-AA force field parameters, showed that with extremely large box sizes and higher temperatures, much better agreement with experiment could be achieved.[297]

To test whether GAFF conforms to this same trend that higher temperatures lead to more accurate IL shear viscosity estimates, additional simulations were performed at temperatures above 298.15 K for which experimental data was available, for 3 of the 5 ILs with the most poorly predicted shear viscosities (see bolded ILs in Table 5.5), namely [EMIM][Ac], [EMPyr][$EtSO_4$], and [HMIM][Cl]. Results are shown in Table 5.6 below, where reported simulated values are averages of 3 simulations run in triplicate, and reported experimental values are averages of the values of the cited literature sources.

Table 5.6. Experimental and simulated IL shear viscosities (Pa•s) using a shear rate of 0.03 nm/ps^2 and temperatures above room condition.

IL	Temperature (K)	Experiment	Simulation	Error (%)
[EMIM][Ac]	298.15	0.135[227, 239, 301]	0.027	-80.4
	353.15	0.016[239, 298]	0.006	-59.5
	393.15	0.005[298]	0.003	-36.9
[EMPyr][$EtSO_4$]	298.15	0.156[249, 279]	0.053	-66.2
	323.15	0.048[249, 279]	0.022	-54.5
	343.15	0.024[249, 279]	0.012	-51.8
[HMIM][Cl]	298.15	18.10[253]	0.380	-98.0
	353.15	0.238[255, 298]	0.063	-73.4
	393.15	0.043[298]	0.019	-55.4

Results show that, in line with the findings of Hu and Margulis[297], as the simulation temperature is increased above room temperature the error in the calculated shear viscosities decreases for each of the 3 ILs, by over 40% in the cases of [EMIM][Ac] and [HMIM][Cl] from 298.15 K to 393.15 K.

5.3.6 Opportunities for Improvement

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

Conclusions

Densities, heat capacities, molar enthalpies of vaporization, self-diffusivities, and shear viscosities of 19 ionic liquids have been calculated using the general AMBER force field and compared to experiment. Results show that GAFF is able to reproduce these properties with good accuracy compared to experiment and with similar accuracy compared to other published force fields. No refinement to GAFF was required for such accuracy. Specifically, deviations in simulated IL densities from experimental values were ca. 1-4%, heat capacity deviations ranged from 0-10%, and deviations in molar enthalpies of vaporization from experiment were ca. 6-13%. In general, across the range of ILs tested, GAFF tended to underestimate the densities, heat capacities, and molar enthalpies of vaporization. Significantly larger errors occurred for calculations of IL transport properties; cation and anion self-diffusivity values were all underestimated by GAFF and deviated from experiment in the range of 50-70%. GAFF was, however, able to capture important trends in the data, such as decreasing self-diffusivity with increasing cation chain length, as well as anion self-diffusivities tracking those of the cations. Deviations in IL shear viscosity from experimental values varied widely, from as little as 2% to as much as 98%. The largest errors occurred for ILs with [Cl], [RSO_4], and [Ac] anions, and may indicate that those simulations were unable to reach the hydrodynamic limit and thus that larger box sizes and higher temperatures could be used to achieve better agreement with experiment. This theory was tested for 3 ILs and results showed errors in the simulated shear viscosities indeed decreased with increased temperatures. Excluding the most deviant ILs from the data set at 298.15 K, the range of viscosity error was found to be much more reasonable, with a max absolute error of ~30%. Considering the difficulty of calculating dynamic properties of ionic liquids with simulation[284], GAFF does surprisingly well at predicting these properties and capturing important trends.

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

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

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

Chapter 6

The Lytic Polysaccharide Monooxygenases *Sc*LPMO10B and *Sc*LPMO10C Are Stable in Ionic Liquids as Determined by Molecular Simulations[6]

Introduction

From both an economic and environmental standpoint, a major limitation to the process of converting lignocellulosic biomass to fuels and chemicals is in the pretreatment methods needed to sufficiently breakdown the cellulosic biomass for further enzymatic hydrolysis. Expensive processing conditions (e.g. high temperatures and pressures) and harsh environmental impacts from caustic chemicals have broadened the field of biocatalysis towards a search for new solvents that are more economically viable and environmentally friendly.[155] Ionic liquids (ILs), which are organic salts that are liquid at room temperature, have been gaining ground as possible solvents for biocatalysis in recent years due to their properties of low flammability and vapor pressure, high recoverability, and highly tunable solvent properties. The countless combinations of cations and anions that make up an IL present a vast exploration space for improving enzyme activity and stability over traditional organic solvents.

Molecular simulation has proven to be an invaluable tool at providing new understanding of the atomic level interactions of enzymes with ILs.[136, 138-140, 145, 151-155, 157-158] Computational studies have assessed the effects of ILs on both the structure and dynamics of enzymes to measure stability and retained activity. For example, a possible link has been found to exist between transient salt bridge/hydrogen bond behavior and enzyme stability[155], and IL-induced conformational change of an enzyme's active site (i.e. substrate binding pocket) has been directly linked to enzymatic activity[145]. Additionally, interaction energies calculated between the enzyme and IL have provided information about the kinetics of enzyme-catalyzed reactions in ILs[145], and

[6]Reproduced in part with permission from K.G. Sprenger, Alaksh Choudhury, Joel Kaar, and J. Pfaendtner. The Lytic Polysaccharide Monooxygenases *Sc*LPMO10B and *Sc*LPMO10C Are Stable in Ionic Liquids as Determined by Molecular Simulations. The Journal of Physical Chemistry B, 120: 3863-3872, 2016.

principal component analysis has been performed to gain insight into slow, correlated motions of enzymes in ILs[140]. Molecular simulation has also been used as a predictive tool in determining the role of ILs as competitive inhibitors to enzyme substrates.[140] Due to the extensive experimental literature available, a large number of molecular simulations of enzymes in ILs to date feature lipases such as *Candida antarctica* lipase B (CALB) and *Candida rugosa* lipase (CRL).[145, 152-154, 157-158] Less common but growing in examples are simulations of biomass-degrading enzymes in ILs, although such examples have involved only cellulases and xylanases.[140, 155]

The nascent discovery of the enzyme family lytic polysaccharide monooxygenases (LPMOs) has changed the paradigm of how the cellulose degradation system is understood. Unlike the hydrolytic mechanism of endo- and exoglucanases, LPMOs oxidize glycosidic bonds in cellulose, creating new chain termini in the interior of the fibril rather than at the ends of the fibril that can then be acted upon by advancing cellulases[313]. It has been proposed that LPMOs increase the accessibility of other cellulolytic enzymes to the cellulose surface, thereby increasing their activity[314]. Specifically, the mechanism of activity of these enzymes entails oxidation of the C_1 or C_4 carbon in the polysaccharide by reducing dioxygen using a bound copper ion, and internal electrons transferred by an unknown mechanism[315-318]. The copper in the active site is coordinated by three nitrogen ligands (two from the N terminal histidine and one from a highly conserved histidine) in a T-shaped coordination plane termed a histidine brace[316], as well as by a fourth group such as a water molecule or acetate ion (Figure 6.1). Notably, activity enhancement by LPMOs has been reported in both free enzyme cocktails[319] and complex cellulosomes[320] on phosphoric acid-swollen microcrystalline cellulose (PASC). One of the most widely used commercial cellulolytic cocktails, CTec2, by Novozymes, also now includes certain fungal LPMOs to enhance the activity of the mixture.[321]

Given the immense potential of LPMOs to accelerate biomass conversion for the production of biofuels and other chemicals, it is of considerable practical interest to investigate the stability of LPMOs in relevant conditions for biomass processing. ILs have generated widespread interest as novel solvents for the dissolution of cellulosic material[322-324]. With both computational and experimental techniques, we have extensively studied the activity and stability of enzymes in these solvents, while furthermore developing rational methods to stabilize enzymes against IL-induced inactivation.[141-142, 154] In the context of biomass processing, much of this effort has focused on elucidating and improving the stability of cellulolytic enzymes in these solvents[325]. Ultimately, evaluating the stability of LPMOs in ILs is a necessary step towards engineering LPMOs for improved utility and synergistic activity with cellulases for industrial biomass conversion.

Herein, we have used molecular dynamics (MD) simulations as a computational screening tool to predict the suitability of the bacterial LPMOs *Sc*LPMO10B and *Sc*LPMO10C (CelS2), for use in ILs. These enzymes are recently discovered cellulose-active LPMOs from the bacteria *Streptomyces coelicolor*.[326] Structures for the two enzymes were obtained from PDB entries 4OY6 and 4OY7, respectively.[326] While *Sc*LPMO10C oxidizes only the C_1 carbon, *Sc*LPMO10B reportedly can oxidize both the C_1 and C_4 carbons on polysaccharide chains.[326] Structurally, these LPMOs are similar to previously characterized LPMOs containing the highly conserved histidine brace and a flat cellulose-binding surface. We investigated the enzymes' structure and stability in three ILs commonly used in biomass processing applications: 1-butyl-3-methylimidazolium chloride [BMIM][Cl], 1-ethyl-3-methylimidazolium acetate [EMIM][Ace], and 1-allyl-3-methylimidazolium chloride [AMIM][Cl]. Each enzyme was evaluated in 10 and 20% (w/w) IL as well as in pure water as a control. Each system was prepared in duplicate with different random

initial solvent configurations and simulated for 250 ns in the NPT ensemble, for an aggregate of eight microseconds of MD simulations. Following this, we analyzed the trajectories to study the structure, dynamics and active site behavior of the systems. To the best of our knowledge, there have been only a couple of molecular simulation studies to date that have reported on the behavior of LPMOs[313, 327], and both in aqueous conditions. Thus, this study represents the first to examine the stability of LPMOs in ILs and therefore assess the potential utility of these enzymes in IL media.

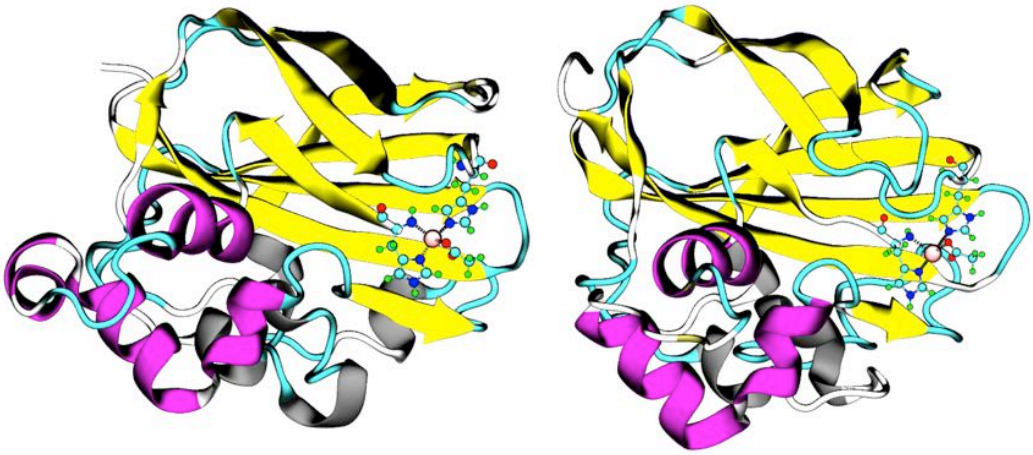

Figure 6.1. Side-by-side comparison of the structures of LPMO enzymes *Sc*LPMO10B, left, and *Sc*LPMO10C, right. Magenta and gray coloring indicate regions with α- and 3_{10}-helical character, respectively, and yellow and cyan coloring indicate regions with β-barrels and turns, respectively. The location of the copper active site within each enzyme is also highlighted. Each copper ion is coordinated to a bidentate histidine residue, a different monodentate histidine residue, and a monodentate acetate ion ligand (substituted for a coordinating water molecule in *Sc*LPMO10C for consistency with *Sc*LPMO10B). Dotted black lines signifying bonds between atoms coordinated to the central copper ion in the active site are included to guide the eye.

Methods

6.2.1 Enzyme/IL Simulations

The simulations utilized AMBER compatible force fields implemented in GROMACS 4.6.5[50] and the TIP3P water model[181]. The AMBER99SB-ILDN force field[51] was used to describe the enzymes, and the ILs were parameterized using the general AMBER force field (GAFF)[182] based on our previous work[140, 155, 166]. Parameterization of the Cu-histidine brace was performed as described in the Supporting Information. As suggested in the literature[172], we improved prediction of dynamical properties by scaling the IL point charges by 0.8. Cubic boxes of the solvent and protein were generated using Packmol.[178] Table 6.1 lists the size of each simulation box and the number of all solvent molecules for each system.

Table 6.1. Size of each simulation box and the number of all solvent molecules for each system.

		Box Size (nm^3) / Number of Solvent Molecules			
		10 Wt%		20 Wt%	
Enzyme	IL	Trial 1	Trial2	Trial 1	Trial 2
*Sc*LPMO10B	[AMIM][Cl]	419.0 / 35,259	417.0 / 35,259	418.2 / 31,806	417.7 / 31,806
	[BMIM][Cl]	417.4 / 35,013	419.3 / 35,013	414.5 / 31,368	417.0 / 31,368
	[EMIM][Ace]	419.5 / 35,073	418.7 / 35,073	417.7 / 31,455	419.3 / 31,455
		Trial 1		Trial 2	
	Water	420.0 / 38,652		419.6 / 38,652	
*Sc*LPMO10C	[AMIM][Cl]	402.6 / 33,666	400.8 / 33,666	399.4 / 30,375	398.8 / 30,375
	[BMIM][Cl]	400.4 / 33,435	398.5 / 33,435	400.0 / 29,955	398.3 / 29,955
	[EMIM][Ace]	401.5 / 33,507	403.4 / 33,507	401.0 / 30,039	399.5 / 30,039
		Trial 1		Trial 2	
	Water	403.2 / 36,900		402.8 / 36,900	

After a steepest descent energy minimization of the initial coordinates, the system was slowly heated to 298 K over 1 ns and then the system was briefly mechanically equilibrated at 1 bar using the Berendsen barostat[109]. For all MD simulations, a time step of 2 fs was attained by constraining the bonds between hydrogen and heavy atoms using the LINCS algorithm.[110] Periodic boundary conditions were applied to all dimensions. Lennard-Jones interactions were calculated over 1.2 nm and shifted to ensure no artifacts were created from discontinuities in energy. Particle mesh Ewald summation was used to account for long-range electrostatic interactions. The production simulations in the NPT ensemble maintained temperature at 298 K using the Bussi-Danadio-Parrinello thermostat[54] and pressure at 1 bar using the Parrinello-Rahman barostat[222].

6.2.2 Parameterization of the Cu-Histidine Brace

The active site of each enzyme consists of a copper ion coordinated to a bidentate histidine residue, a different monodentate histidine residue, and a monodentate acetate ion ligand (see Figure A2.1). For consistency with enzyme *Sc*LPMO10B, an acetate ion was substituted for a water molecule observed experimentally to be coordinated to the copper ion in enzyme *Sc*LPMO10C. Assigned an overall charge of +1, a quantum mechanics calculation was performed on the full coordinated copper ion active site to determine the atomic point charge of each individual atom. Geometry optimization and energy calculations were performed at the Hartree-Fock (HF) level of theory using the 6-31G* basis set and the electronic structure program Gaussian, from which ESP partial charges were obtained (see Table A2.1). Force constants and equilibrium

geometry parameters were based on the work of Wise and Coskuner[328]. After determination of partial charges, tleap[218] and ACPYPE[179] were used to generate input files.

Results and Discussion

6.3.1 IL Effects on Enzyme Structure

To understand if the ILs induced changes in the structure of *Sc*LPMO10B and *Sc*LPMO10C, we calculated the root mean squared deviation (RMSD) of the Cα atoms of each enzyme from the PDB crystal structure. The entire time series is shown in Figure 6.2 and the mean RMSD values (calculated over the 2nd half of the trajectories) are tabulated in Table 6.2. Data points were collected from the trajectories every ns.

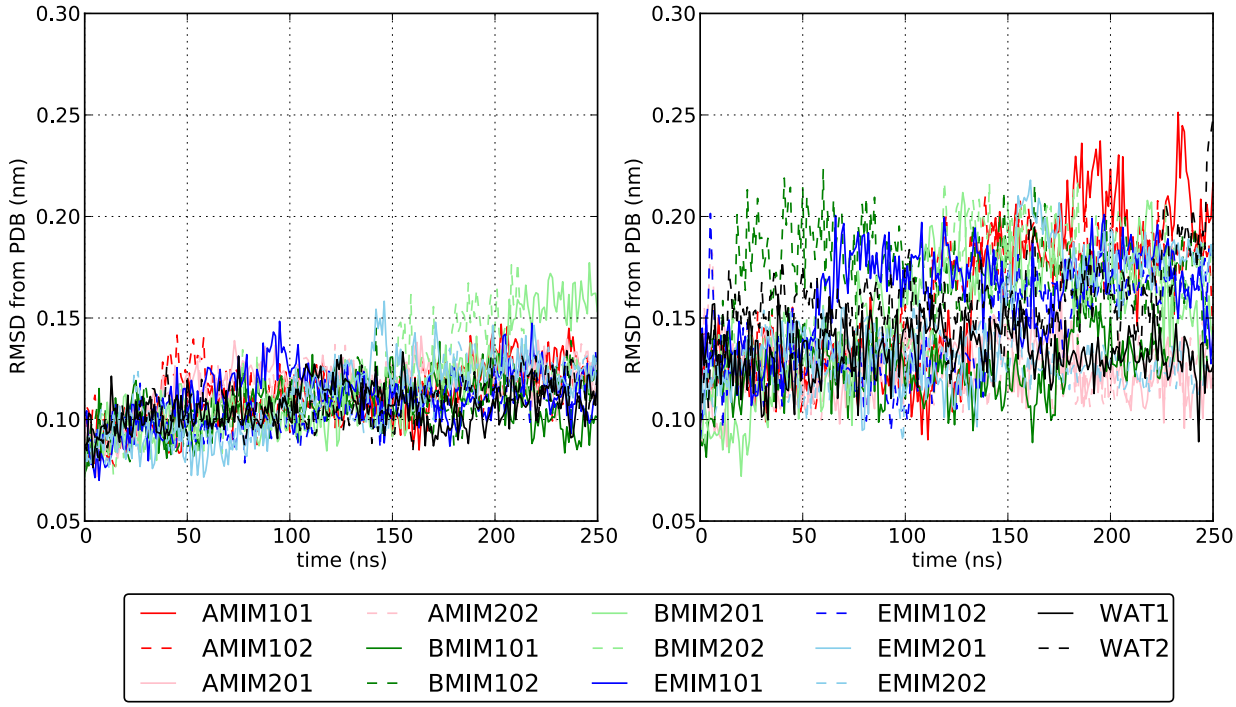

Figure 6.2. Side-by-side comparison of the time series of the Cα RMSD of *Sc*LPMO10B (left) and *Sc*LPMO10C (right) from the crystal structure for each enzyme/solvent combination. Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer).

Table 6.2. Average RMSD and standard deviation of Cα atoms from the crystal structure for each enzyme/solvent system, calculated over the second half of the 250 ns trajectories. The error bars are one standard deviation of the RMSD over the same period.

		Average Cα RMSD from PDB (nm)			
		10 wt%		20 wt%	
Enzyme	IL	Trial 1	Trial2	Trial 1	Trial 2
*Sc*LPMO10B	[AMIM][Cl]	0.12±0.013	0.12±0.0093	0.12±0.0072	0.12±0.0078
	[BMIM][Cl]	0.11±0.012	0.11±0.0096	0.13±0.023	0.13±0.017
	[EMIM][Ace]	0.11±0.010	0.11±0.0080	0.12±0.0090	0.12±0.013
		Trial 1		Trial 2	
	Water	0.11±0.0099		0.11±0.0099	
*Sc*LPMO10C	[AMIM][Cl]	0.19±0.025	0.18±0.014	0.13±0.014	0.12±0.012
	[BMIM][Cl]	0.13±0.017	0.17±0.016	0.17±0.016	0.18±0.016
	[EMIM][Ace]	0.17±0.017	0.16±0.016	0.17±0.027	0.13±0.0098
		Trial 1		Trial 2	
	Water	0.13±0.0099		0.16±0.021	

The data in the table show that the differences between the enzymes in any of the solvents were quite small, which is not necessarily surprising as both enzymes show similar structural motifs with a β-barrel, α-helical and 3_{10}-helical regions (Figure 6.1). The data does show, however, a general trend of slightly higher RMSDs for *Sc*LPMO10C than for *Sc*LPMO10B, which is the case for both IL solvents and pure water. Inspection of the time series (Figure 6.2) shows basically the same behavior, and that the largest peaks in RMSD of the IL systems are generally on par with those of the water simulations. The time series does show that two of the systems, *Sc*LPMO10B/20%[BMIM][Cl]-trial 1 and *Sc*LPMO10C/10%[AMIM][Cl]-trial 1, display deviations that appear to exceed those of the water simulations. To ensure this was only a statistical fluctuation, we extended these two systems for an additional 500 ns each. For the system with [BMIM][Cl] we saw no further changes in the RMSD (Figure A2.2), and for the system with [AMIM][Cl] we saw the RMSD increase by about 1 Å over the extended time of 750 ns. The magnitude of the RMSD values, overall, is very small and indicative of a high degree of stability. Moreover, there are no apparent differences in the structural changes over time in the IL-containing systems (regardless of wt% IL) versus water-containing systems, a feature previously identified as a potential predictor of enzyme activity in ILs[155].

A prior study by our group investigated a family of three cellulase enzymes and found, in at least one case, widespread denaturation of α-helical regimes by aqueous solutions of [EMIM][Ace].[155] The three-helix bundles in *Sc*LPMO10B and *Sc*LPMO10C are close to the active site (c.f., Figure 6.1) and therefore could interfere with catalytic activity should they unfold upon exposure to the ILs. Furthermore, the hydrophobic packing of each bundle is different, making this an interesting comparison. Figure 6.3 highlights the differences in the distribution of hydrophobic, polar, and charged residues within the 3-helix bundles of the two enzymes. The results show

similar charge motifs exist for the 3-helix bundles: the overall charge of the bundle is 0 in enzyme *Sc*LPMO10B and -1 in *Sc*LPMO10C. These charge motifs are very similar to the surface charges of two of the cellulases studied in our previous work (surface charges of 0 and -4 were reported for *T. viride* and *P. horikoshii*, respectively), and notably different from the surface charge of +13 that was reported for *T. maritama*.[155] *T. maritama* was found to undergo a significantly higher loss of secondary structure in IL than the other two enzymes, supporting the hypothesis that a positive surface charge is more susceptible to an attack by anions, which can ultimately lead to unfolding.[154] Thus, the surface charge motifs of the 3-helix bundles indicate a capacity to prevent penetration by anions and retain a measure of stability in ILs, which is supported by the results of the following analysis.

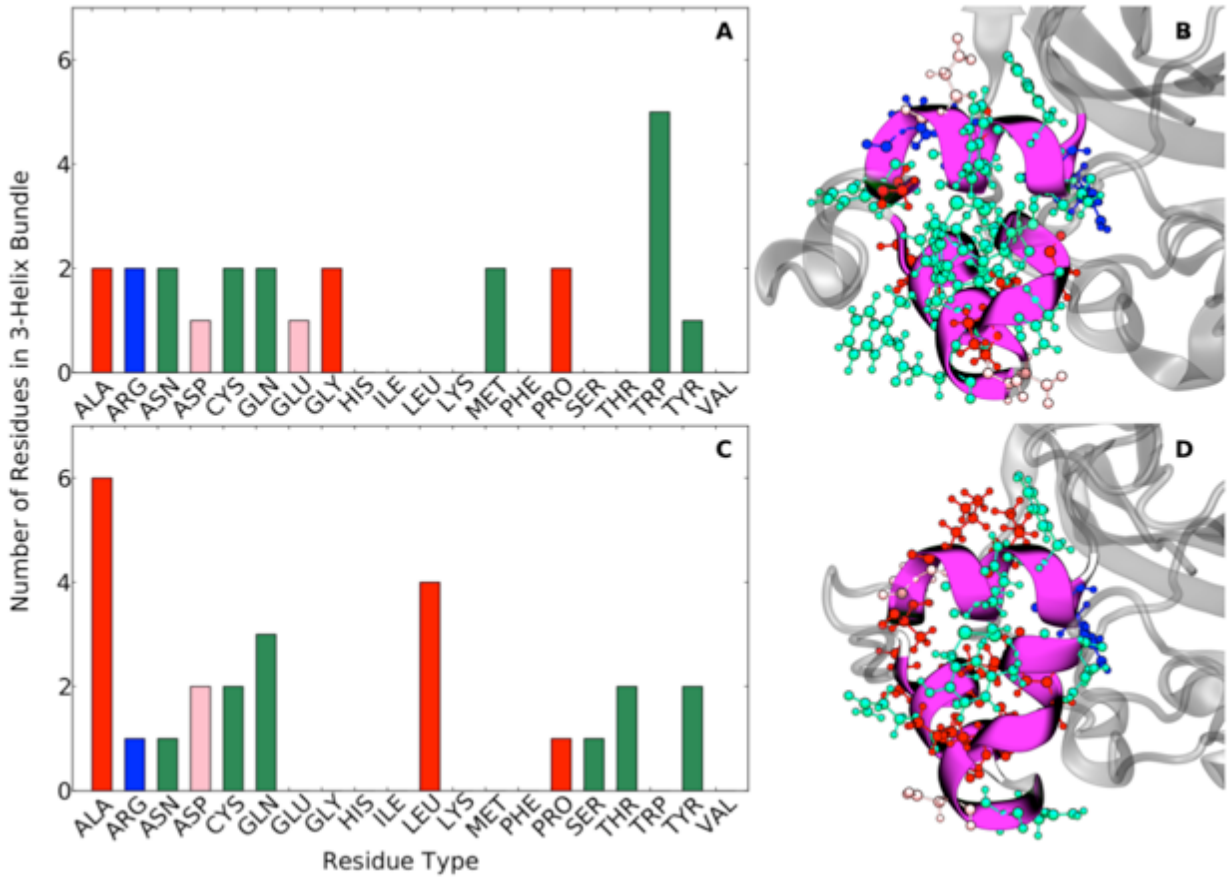

Figure 6.3. Composition of the residues within the 3-helix bundles of enzymes *Sc*LPMO10B (A) and *Sc*LPMO10C (C), with positively charged amino acids highlighted in blue, negatively charged amino acids in pink, hydrophobic amino acids in red, and polar amino acids in green. Zoomed in representations of the enzymes' 3-helix bundles (B, D) are also shown, with residues colored to correspond to the plots at the left.

Despite similar charge motifs of the helical regions in the enzymes, Figure 6.3 shows *Sc*LPMO10C has nearly double the amount of hydrophobic residues in this region (11 vs. 6), and *Sc*LPMO10B contains a greater number of polar residues (14 vs. 11). Since stronger hydrophobic packing may be considered a very course indicator of stability, the results imply *Sc*LPMO10C could be more tolerant to ILs than *Sc*LPMO10B; however, subsequent analysis shows the bundles

are very similar in their ability to withstand the effects of ILs. Figure 6.4 shows the RMSD of the enzymes' Cα atoms for the 3-helix bundles from the PDB crystal structure. As the times series show, the bundles are remarkably stable in both *Sc*LPMO10B and *Sc*LPMO10C.

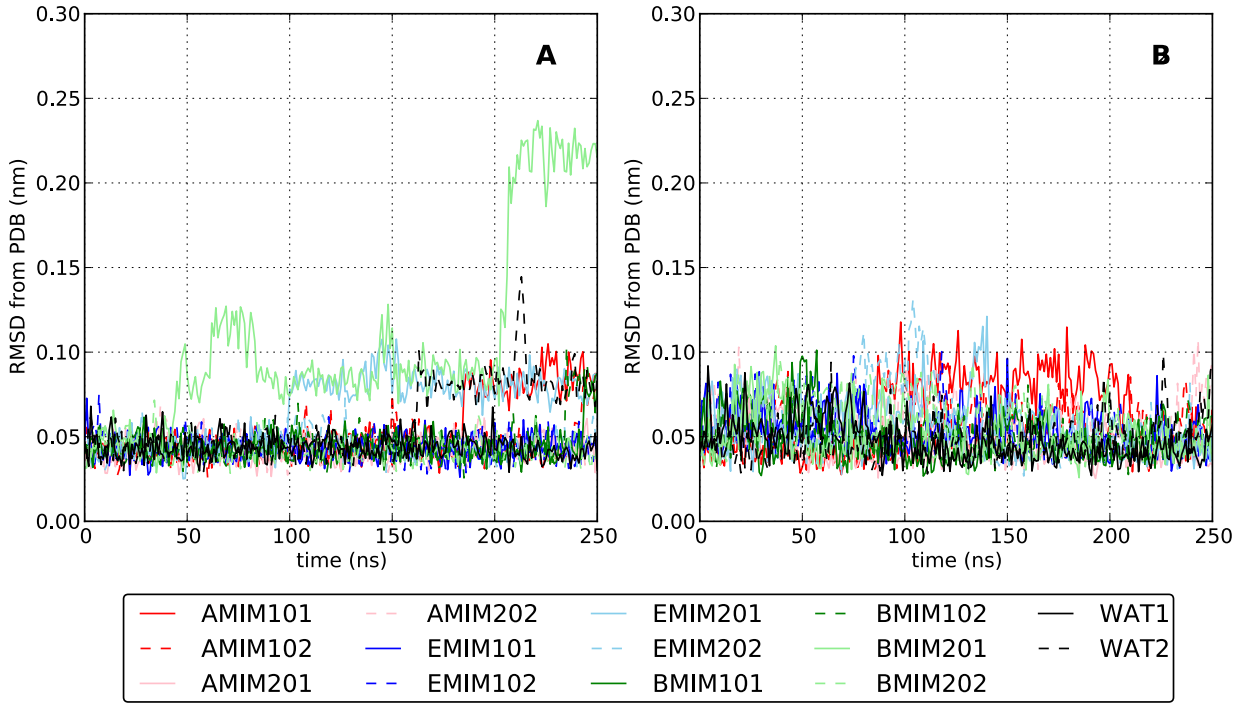

Figure 6.4. Side-by-side comparison of the stability of the 3-helix bundle of each enzyme: A) RMSD of the helical regions of *Sc*LPMO10B in water and in each IL; and B) RMSD of the helical regions of *Sc*LPMO10C in water and in each IL. Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer).

The two exceptions in the time series shown in Figure 6.4 where the enzymes do not appear to be stable are *Sc*LPMO10B/20%[BMIM][Cl]-trial 1 (BMIM201), as before (Figure 6.2), as well as one of the water simulations (*Sc*LPMO10B/water-trial 2, WAT2). The full 750 ns time series for *Sc*LPMO10B/20%[BMIM][Cl]-trial 1 is shown in Figure 6.5 and we saw no further significant changes in the RMSD. A snapshot from the simulation of the structure of *Sc*LPMO10B compared to the crystal structure at the height of the structural fluctuations shows a mild displacement of one of the helices in the bundle. Figure 6.6 shows this change has little effect on the overall secondary and tertiary structure of the enzyme, implying it is likely the cause for the increase in RMSD of the whole enzyme discussed earlier. Figure 6.5 also shows the 250 ns time series for *Sc*LPMO10B/Water-trial 2, along with a snapshot from the simulation at the peak RMSD. The snapshot shows the increase in RMSD is caused by the same displacement of a particular helix in the three-helix bundle as in the [BMIM][Cl] simulation. While the similarity of the deviations could indicate a possible first step in a mechanism of solvent-induced denaturation, the fact that both an IL solvent and water led to these deviations negates the idea that ILs alone are responsible for causing denaturation of the enzyme. It is also important to note that in the case of either simulation, only a single trial displayed the deviant behavior, and thus more trials would need to be run to determine if this behavior is of any real consequence in discussing the stability of *Sc*LPMO10B in these particular solvents.

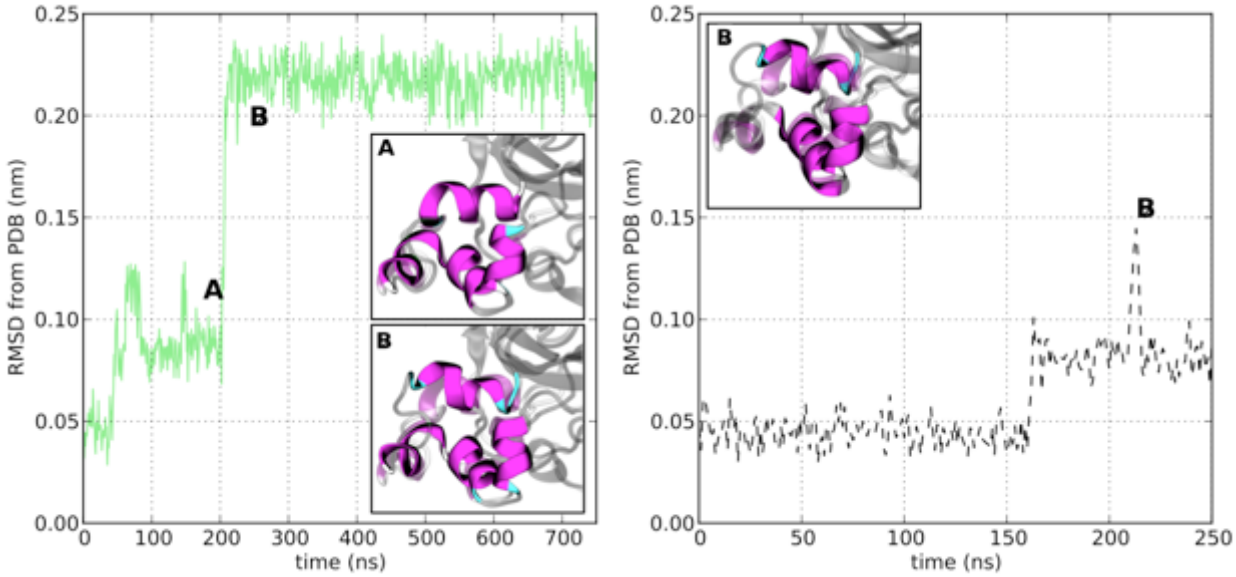

Figure 6.5. Side-by-side comparison of the Cα RMSD of the helical regions compared to the crystal structure for extended systems *Sc*LPMO10B/20%[BMIM][Cl]-trial 1 (left) and *Sc*LPMO10B/Water-trial 1 (right). Plot insets show snapshots from the simulations of the enzyme's structure compared to the crystal structure (zoomed in), with letters marking the time at which the snapshots were extracted from the trajectories. The enzyme's crystal structure is colored entirely transparent silver, whereas the NPT structures are colored transparent dark gray, with α-helical regions colored purple to highlight structural differences in this region.

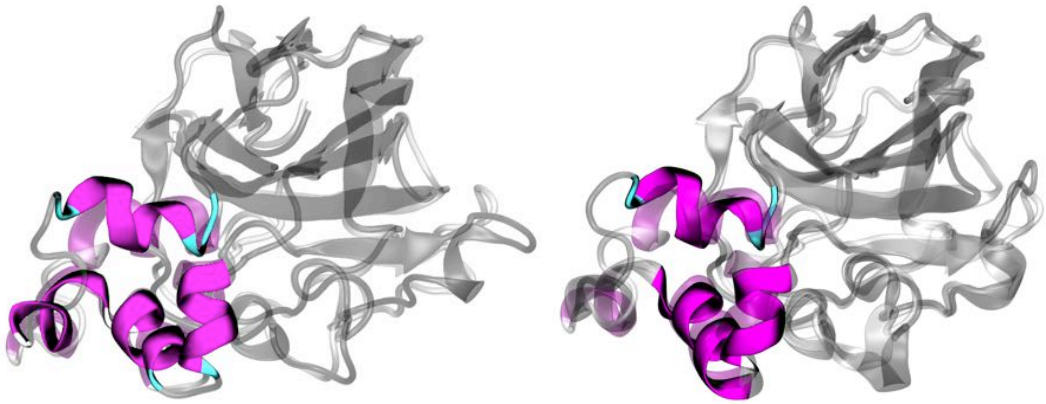

Figure 6.6. Zoomed out versions of (left) the bottom inset of the left plot of Figure 6.5, corresponding to the *Sc*LPMO10B/20%[BMIM][Cl]-trial 1 simulation, and (right) the inset of the right plot of Figure 6.5, corresponding to the *Sc*LPMO10B/Water-trial 1 simulation. The snapshots were extracted from points in the simulations at which maximum structural fluctuations occurred (c.f. Figure 6.5).

To further examine enzyme/IL interactions in relation to the hydrophobic packing differences of the two enzymes, we calculated pairwise or radial distribution functions (RDFs) of the cations (positively charged nitrogen atoms) and anions (negatively charged oxygen atoms used for [Ace]) from the center-of-mass (COM) of the Cα atoms of the 3-helix bundles of the enzymes (Figure 6.7). Looking first at the RDFs between the bundles and surrounding cations in Figure 6.7, the results are very similar for the two enzymes across the range of ILs tested. Peaks occur for both enzymes at distances of ~0.5 and 0.8 nm from the 3-helix bundles and a third peak occurs at ~1.5 nm for *Sc*LPMO10B. These results indicate some short-range organization of the cations in concentric layers about the enzymes. The following order of IL density near the bundles is true for both enzymes: 10% [BMIM][Cl] > 10% [EMIM][Ace] > 10% [AMIM][Cl] > 20% [BMIM][Cl] > 20% [EMIM][Ace] > 20% [AMIM][Cl]. The one notable exception to this ordering is one trial of 20% [BMIM][Cl], which shows an increased density over other ILs near the α-helical region of *Sc*LPMO10B. We hypothesize that an increased number of self-interactions in the 20% IL solutions compared to the 10% IL solutions is contributing to the lower RDF values of the 20% IL solutions.

Interestingly, the density of anions around the 3-helix bundles does not follow the same trends as the density of cations. Figure 6.7 shows notably higher RDF values for the eight solutions with the [Ace] ion (i.e. four solutions with each enzyme) than solutions with the [Cl] ion. The impact of this result is compounded when considering the relative size of the anions: despite the larger size of the [Ace] ion, which in other cases might prove a hindrance, significantly more of them are reported near the surface than are [Cl] ions. More interesting is perhaps the difference in anion organization around the bundles of the two enzymes: sharp peaks exist at ~0.3 nm for every IL solution with enzyme *Sc*LPMO10C, while none of these peaks are present in solutions with enzyme *Sc*LPMO10B. These differences in anion organization are more notable when comparing the water/bundle RDFs for the enzymes (Figure 6.8), which show essentially no differences in the organization of water around the helical regions of the two enzymes. Despite the trends in cation/anion organization indicated by Figure 6.7, in light of the previous RMSD analysis, we conclude that any structural organization of the ILs around the 3-helix bundles does not lead to destabilization of those regions in either enzyme.

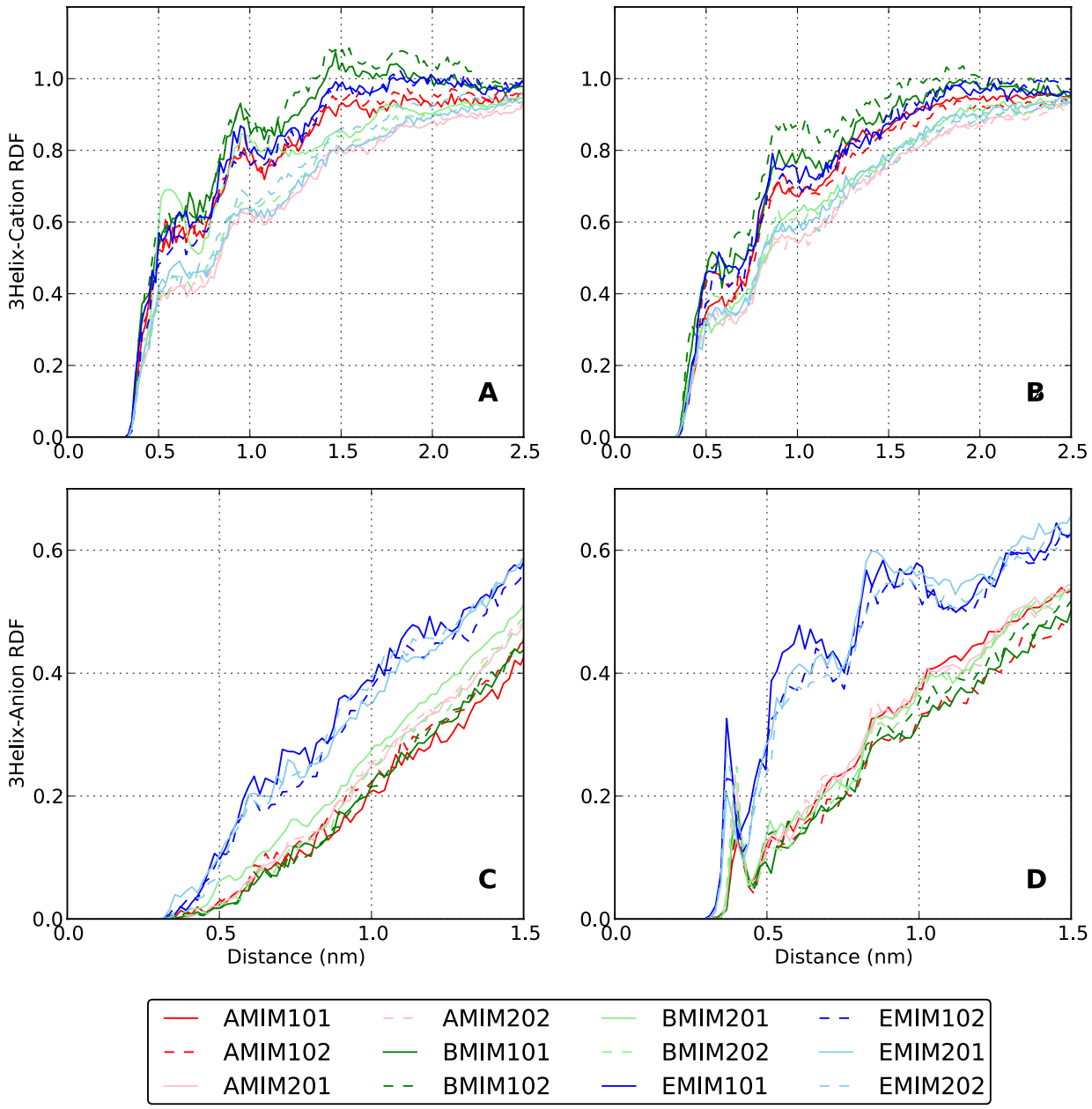

Figure 6.7. Smoothed pairwise distribution functions between the COM of the Cα atoms of the 3-helix bundle in: A) *Sc*LPMO10B, and surrounding cations; B) *Sc*LPMO10C, and surrounding cations; C) *Sc*LPMO10B, and surrounding anions; and D) *Sc*LPMO10C, and surrounding anions. Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer).

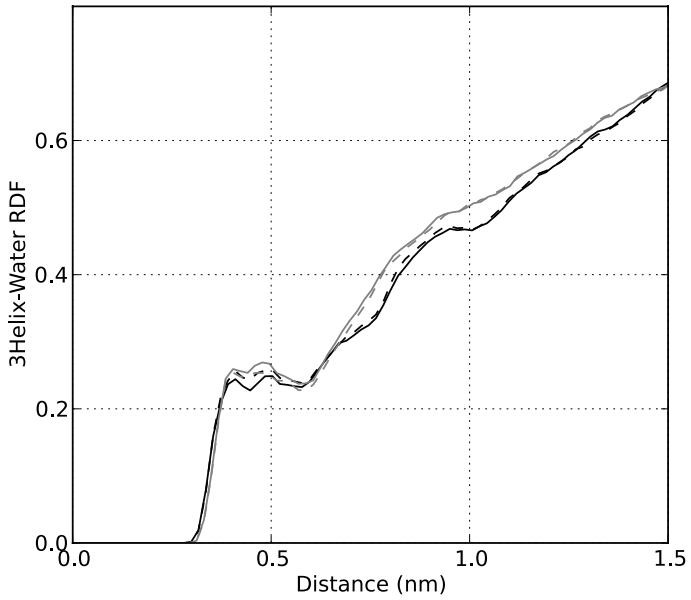

Figure 6.8. Smoothed pairwise distribution functions between the oxygen atoms in water molecules and the centers-of-mass of the 3-helix bundles in *Sc*LPMO10B (black) and *Sc*LPMO10C (gray). The two trials with each enzyme are shown with solid and dotted lines.

6.3.2 IL Effects on Dynamical Fluctuations

Following the structural analysis, we used root mean squared fluctuation (RMSF) of the enzymes' Cα atoms to assess whether specific regions in each enzyme experience larger than average fluctuations upon exposure to IL. The full 250 ns RMSF profiles for each simulation are given in Figure A2.3 and Figure 6.9 shows the ΔRMSF for each simulation, where the change shown is the difference between each simulation and the average RMSF of the two simulations in water.

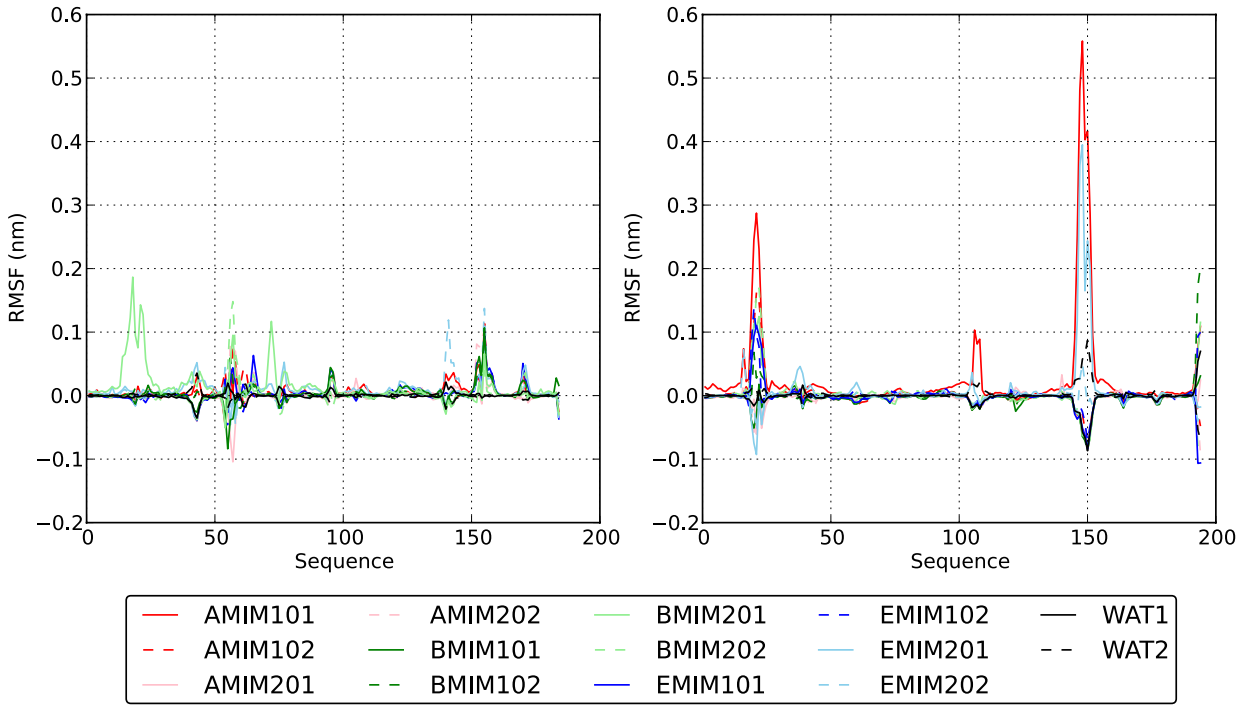

Figure 6.9. Side-by-side comparison of the difference in RMSF between each simulation and the average RMSF of the two water simulations, for *Sc*LPMO10B (left) and *Sc*LPMO10C (right). Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer).

As with the RMSD analysis, the RMSF results generally show that the enzymes do not appreciably change with the exception of a few of the simulations. As with the prior RMSD results, *Sc*LPMO10B showed significant changes from water in one 20% [BMIM][Cl] simulation (BMIM201). This deviation occurs around residues 15-25, which, as shown in Figure 6.10, corresponds to the exact region of the enzyme where partial displacement of one of the three-helix bundles was seen to occur during the earlier analysis of the α-helix RMSD. In the same system, around residues 52-60, corresponding to a flexible loop region of the enzyme (Figure 6.10), several of the solvents (including one of the simulations in water) showed some deviation from the average water values. These deviations, however, are much smaller than the largest features in the *Sc*LPMO10C system. As with *Sc*LPMO10B, the same region in the three-helix bundles causes the observed deviations around residues 15-25 in *Sc*LPMO10C (Figure 6.10). Around residues 144-154, there are significant deviations from the mean water RMSF values (ΔRMSF 0.4-0.6 nm) for one 10% [AMIM][Cl] simulation (AMIM101), and one 20% [EMIM][Ace] simulation (EMIM201). Similar to *Sc*LPMO10B once again, these residues in *Sc*LPMO10C correspond to a loose loop region of the enzyme (Figure 6.10). As these deviations occur for just one instance out of two simulations, the results indicate that, overall, the fluctuations we observe in IL and pure water are approximately the same. However, it is important to note that the deviant loop in both enzymes around residues 15-25 falls within a region of the enzymes termed the L2 loop[327], which represents at least 50% of the putative substrate-binding surfaces.[326] Thus, it is possible that the observed deviations in this region alter binding of the enzymes to the cellulose surface, which could alter their activity.

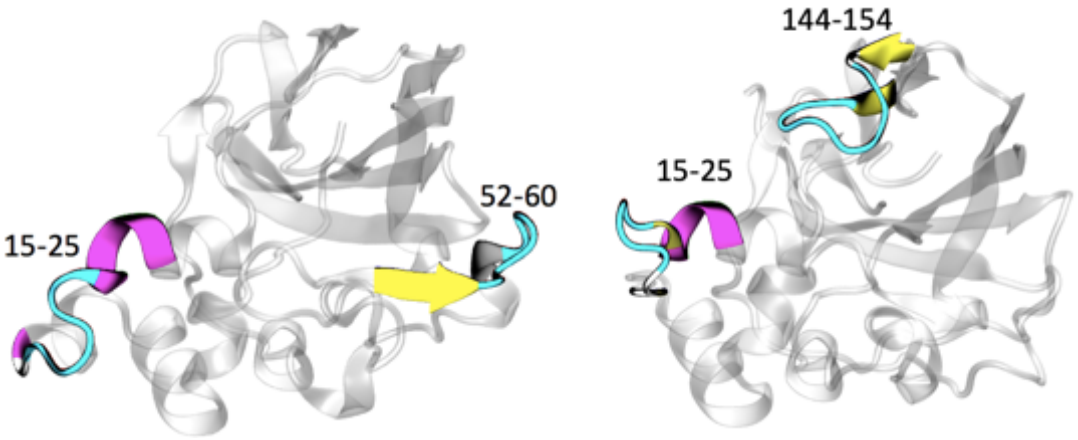

Figure 6.10. Crystal structures of *Sc*LPMO10B (left) and *Sc*LPMO10C (right), highlighting regions of high RMSF values in accordance with Figure 6.9. Number ranges next to highlighted areas indicate the corresponding sequence range.

6.3.3 IL Effects on Enzyme Active Sites

Finally, we investigated the active site structure of the enzymes, including, in particular, whether the ILs alter the access of water to the histidine brace at the active sites of each enzyme and potentially interfere with the catalytic reaction. Although for bacterial LPMOs like the two examined here there is a conserved alanine to preclude water from the active site[326], we investigated if the exclusion of water was disrupted by the addition of the ILs. To answer this question we performed a two-fold analysis. First, we examined the pairwise distribution functions (Figure 6.11) for Cu-water distances. The data in Figure 6.11 were calculated in the same fashion as a traditional RDF, but only the distances between a single Cu ion and all of the waters (O atom) are tabulated. All simulations showed essentially the same behavior with a few small exceptions. In the *Sc*LPMO10B/20%[AMIM][Cl]-trial 1 simulation (AMIM201), there was a notable increase in water 4 Å from the active site above that of the pure water simulations. At a similar distance from the active site in enzyme *Sc*LPMO10C, compared to the pure water simulations, the IL solvents showed a greater distribution of water concentration around the active site than with *Sc*LPMO10B. The greatest increase in water around the active site occurred with 20% [AMIM][Cl] (AMIM201 and AMIM202, lines overlap), whereas a clear decrease in water content occurred with 10% [EMIM][Ace] (EMIM201). In contrast to the other aspects of the simulations we investigated where only one of the replicates showed any significant response compared to water, these results indicate a more consistent response by [AMIM][Cl] (20wt%) in the presence of either enzyme. In general, however, our results show there is little reason to suspect the ILs investigated here would spuriously change the structure/dynamics of the LPMO active site and increase water transport.

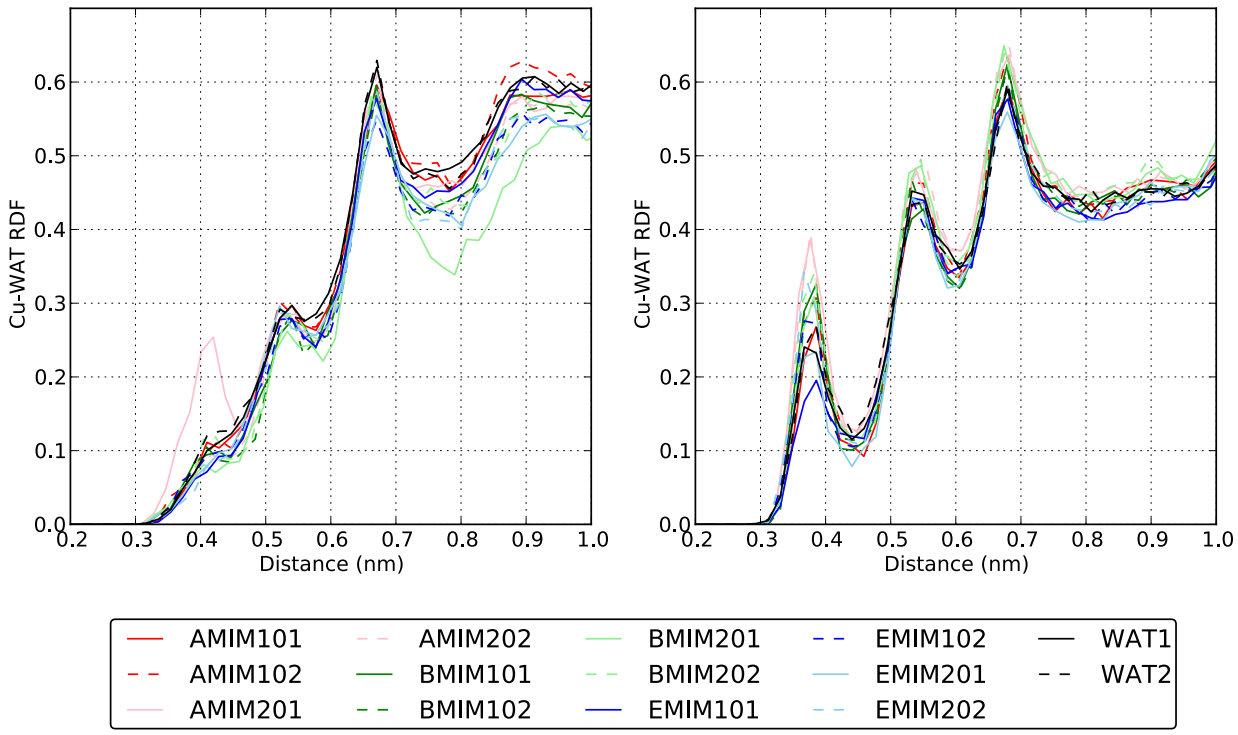

Figure 6.11. Side-by-side comparison of smoothed pairwise distribution functions between the copper ion in the active site and surrounding water for *Sc*LPMO10B (left) and *Sc*LPMO10C (right). Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer).

To further examine the ability of the conserved alanine residues to preclude water from the active site, we inspected the distance between the Cu ion and Cβ atom of the alanine residue near the active site. Figure 6.12 shows the 250 ns time series for each simulation. Consistent with the results from Figure 6.11 for enzyme *Sc*LPMO10B, only the 20% [AMIM][Cl] (trial 1) simulation showed a prominent deviation from the water results, implying the extended widening of the Cu-alanine distance is the reason for the higher concentration of water near the active site seen in Figure 6.11. This does not seem to be the reason for the increase in water content of the *Sc*LPMO10C/20%[AMIM][Cl] simulations, which show little fluctuation of the Cu-alanine distance in Figure 6.12. For *Sc*LPMO10C, only 10% [AMIM][Cl] (trial 1) and 20% [EMIM][Ace] (trial 1) showed any significant fluctuations in distance, though the largest fluctuations in either of these simulations are on par with the fluctuations of one of the pure water simulations.

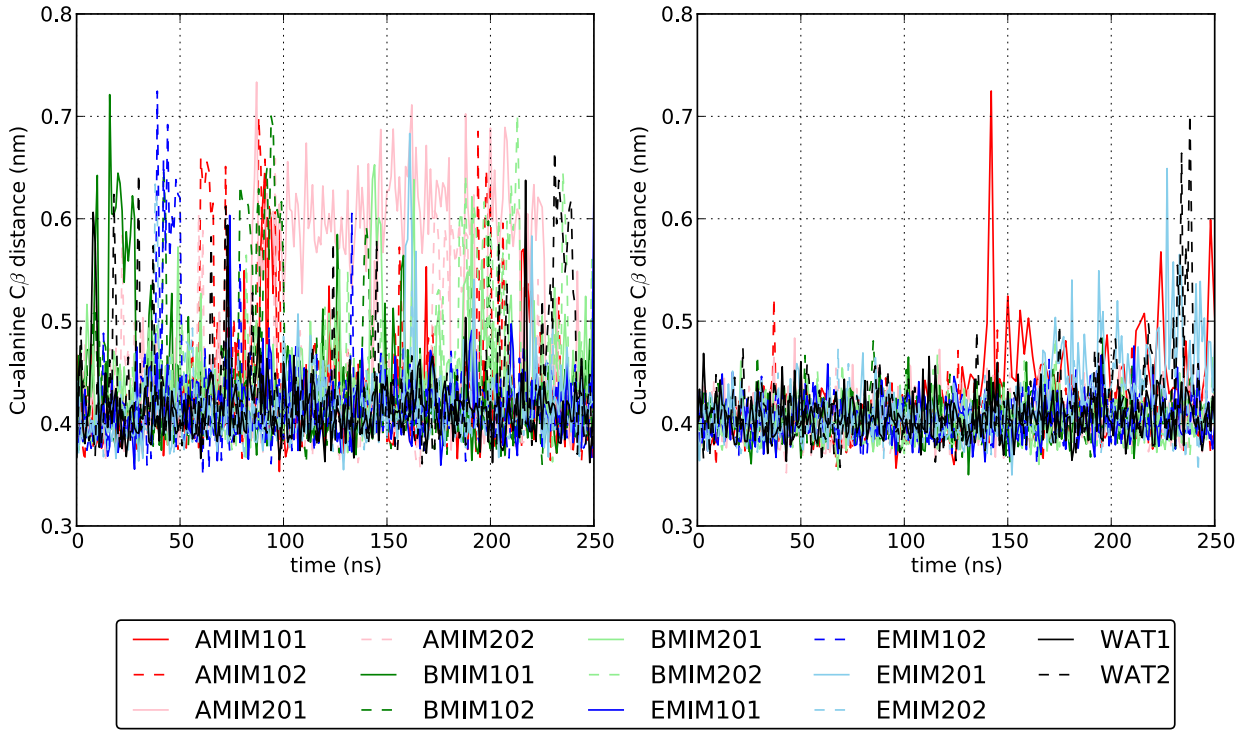

Figure 6.12. Side-by-side comparison of the distance between the copper ion and alanine residue (Cβ atom) near the active site for *Sc*LPMO10B (left) and *Sc*LPMO10C (right). Legend entries indicate the IL, wt% in water (10 or 20), and trial number (final integer).

Conclusions

The results presented here show that on a 250-750 ns timescale, the bacterial LPMOs *Sc*LPMO10B and *Sc*LPMO10C retain substantial similarities in structure and dynamics in ILs as in water. Even on inspection of microscopic aspects such as key structural elements (active site neighboring helical bundles) and solvation properties (water access to the active site), there is every indication from the MD simulations that the LPMOs, like cellulases and xylanases, are excellent candidates for biomass degradation in ILs. Specifically, the results of our structural and dynamical investigations showed: 1) the global structure of the enzymes remains intact in IL; 2) local helical structure that is susceptible to IL-denaturation (and also close to the active site) also remains intact and shows no significant changes from what occurs in pure water (this appears to be due both to highly stable hydrophobic packing in both enzymes as well as a surface charge motif that prevents penetration by anions, which can cause denaturation); 3) the fluctuations of the enzymes in IL and in water are roughly the same; and 4) the structure/dynamics of the enzymes' active sites are generally unaffected by the presence of IL. These findings are also consistent with a comprehensive simulation study of cellulases[140, 155], which indicated that IL-induced enzyme deactivation can provide signatures in MD simulations in the form of significant deviations when comparing aqueous systems to IL systems. Future experimental characterization of LPMO function in ILs, in the form of activity assays, will provide needed final validation of the conclusions suggested by these simulations. A natural next step for simulations entails the rational

design of LPMO family enzymes to have increased stability in ILs. Specific to the LPMOs studied here, increasing the rigidity of loops that were found to deviate from the crystal structure as well as engineering residues around the conserved alanine near the active site to further increase stability are of importance. Another future step for simulations includes engineering enzyme cocktails to function in ILs.

Chapter 7

Elucidating Sequence and Solvent Specific Design Targets to Protect and Stabilize Enzymes for Biocatalysis in Ionic Liquids[7]

Introduction

Ionic liquids (ILs) are a special class of solvents for biocatalysis where enzyme/host interactions are known to play a large role in determining enzyme structure and function. There has therefore been intense interest in recent years to determine the effect of ILs on enzymes for biocatalysis, studied through both experiments[135-150] and computational methods[136, 138-140, 145, 151-155, 157-158, 329]. Recent experimental studies have shown enzyme surface charge modifications have a large impact on enzymatic activity in the presence of ILs.[142-143] Complementing such experiments, simulations have been able to provide a molecular picture of how ILs interact with enzymes, elucidating, for example, how controlled surface charge modifications affect the long-range structure of the IL solvent.[154] Molecular dynamics (MD) simulations in particular hold great promise for augmenting experimental structural measurements from NMR and x-ray crystallography, whose application is often limited in the highly concentrated IL solutions required by many industrial applications.[330]

Despite the rapidly growing number of computational studies of enzymes in ILs, however, there has been a lack of experimental evidence with which to validate simulation force fields and methodology and quantify predictions from MD. New insight from NMR[141] and crystallography[8] has provided a unique set of data around which to benchmark molecular simulations of enzymes in ILs, which we use herein to address the question of MD's predictive power in enzyme/IL simulations. This data was acquired for several forms of lipase, which is both a model enzyme of known structure and catalytic mechanism as well as an industrially important enzyme, in complex

[7]Reproduced in part with permission from K.G. Sprenger, J.G. Plaks, Joel Kaar, and J. Pfaendtner. Elucidating Sequence and Solvent Specific Design Targets to Protect and Stabilize Enzymes for Biocatalysis in Ionic Liquids. Physical Chemistry Chemical Physics, 19: 17426-17433, 2017. Copyright 2017 Royal Society of Chemistry.

with imidazolium ILs in aqueous solution. The findings from these studies provided, for the first time, detailed experimental information about the preferred locations and binding modes of IL cations and anions on enzyme surfaces. These are at atomic resolution and can be directly compared with predictions from MD.

This data set, with quantitative information on the stability and structure of a series of lipase mutants, also affords a unique opportunity for MD simulations to provide new insight into the mechanism of IL-induced enzyme destabilization. While it is well-known that some ILs can lead to large-scale denaturation of proteins, the mechanism by which an IL solvent may change the relative folding/unfolding free-energies has not been fully explored. Recent simulation work by our group quantified the unfolding rate of a miniprotein as a function of the IL hydrophobicity[331], but did not determine the net effect of the solvent on ΔG. Herein, we add new understanding to this effect by quantifying IL/enzyme interactions in the wild-type and mutant and folded and unfolded states, providing new understanding of the IL destabilization mechanism.

The goals of the current study are two-fold. First, we examined the predictive capabilities of current interatomic potentials by validating the use of MD simulations to study enzyme/IL interactions through comparing directly with atomistically resolved structural experiments.[8, 141] We have evaluated the surface binding characteristics of the IL 1-butyl-3-methylimidazolium chloride ([BMIM][Cl]; 0, 5, 10, and 20% (v/v) in water) to wild-type Bacillus subtilis lipase A (WT-lipA), as well as a quadruple mutant of lipase A (QM-lipA). Our results show MD simulations can recapitulate the experimentally measured molecular scale interactions between enzymes and ILs, suggesting they may be used in a predictive capacity in the future. We expand on the current understanding of these interactions by elucidating the specific molecular scale mechanism of IL binding in a systematic manner. Second, we extended our simulations to study the binding characteristics of ILs to the unfolded state of WT-lipA and QM-lipA. Our results provide unprecedented insight into the mechanism of enzyme stabilization in ILs by surface point mutations.

Computational Methods

7.2.1 Binding Site Determination

Structures of WT-lipA crystallized in 5, 10, and 20% (w/w) aqueous IL solutions were analyzed with PYMOL[332] to determine the surface protein residues that were involved in binding IL cations during the experiments. Following the experimental definitions of a contact, all protein residues within 4 Å of a bound cation were initially included in the definition of the respective binding site (n.b., a "binding site" is a site in the folded protein, as opposed to a linear sequence of residues). In some cases, cations were discovered to be bound by residues from two different protein monomers (e.g., there are two monomers in the unit cell of WT-lipA), and between protein residues in different but adjacent unit cells. As WT-lipA is active in the monomeric state, rather than in the dimeric state, it is likely that some of these cation-binding residues are not part of true binding sites (i.e., those with the ability to strongly attract and bind cations in the monomeric protein form). However, without any prior knowledge to help eliminate these residues beforehand, it was necessary to include all cation-binding residues in the initial binding site definitions.

Residues of different protein monomers that bound the same IL cation were assigned to different binding sites to reflect their unique locations on monomeric WT-lipA. This led to the identification of 12 major locations for IL cation binding on the surface of WT-lipA, as well as a rough estimate of the residues actively involved in binding (i.e., crystal structures only provide a "static" snapshot of binding). Analysis of the density of IL cations around each residue in Visual Molecular Dynamics (VMD)[47] throughout our simulation trajectories was done to further refine the binding site definitions, summarized in Table A3.1.

7.2.2. Simulation Protocol

To parameterize the IL, the General Amber Force Field (GAFF)[182] was used with the RESP[217] partial charge calculation method (appropriately scaled[172]), following our established protocol for creating de novo IL force fields[333]. GAFF has been shown to accurately reproduce many thermodynamic and transport properties of a wide range of ILs.[166] For the enzymes (wild-type lipase (WT-lipA) and a quadruple-mutant lipase (QM-lipA)), the Amber99SB-ILDN force field[51] was used, and the TIP3P water model was used for water[181]. Cubic boxes of the enzyme/IL solutions were generated with Packmol[178] (~36-43K atoms), and solutions were constructed with 0, 5, 10, and 20% [BMIM][Cl] (v/v) in water to match experiments. To prohibit bias in our occupancy calculations, simulations were setup without the [BMIM][Cl] initially bound to the enzymes in the crystal structures, though all crystallographic waters around the enzymes were included. Topology for the enzymes was initially built with the hydrogen coordinate data in the published crystal structures left in place, which resulted from the best-fitting score in the structure determination from the X-ray diffraction data. Additional simulations to support our findings were performed with different side chain protonation states of key surface residues as described in the Results and Discussion section, bringing the total sampling time to just over 5 μs.

A steepest descent energy minimization of initial coordinates for each system was run for 10,000 steps. NPT equilibration was run for 5 ns at 298 K and 1 bar with the Bussi-Donadio-Parrinello thermostat[54] and Berendsen barostat[109]. Production simulations where then carried out in the NPT ensemble at the same temperature and pressure using the same thermostat and Parrinello-Rahman barostat[222]. All MD simulations used a time step of 2 fs and periodic boundary conditions in all dimensions. Bonds between hydrogen and heavy atoms were constrained with the LINCS algorithm.[110] Long-range electrostatic interactions were calculated with Particle mesh Ewald summations (PME) and a cutoff of 1.4 nm, and Lennard-Jones interactions were calculated over 1.2 nm and shifted to eliminate artifacts in energy discontinuities. Neighbor lists were updated every 10 MD steps with a cutoff of 1.4 nm. While a single trial was run for each enzyme/IL system, production simulations were run for 0.5 μs each, allowing for partitioning of the trajectories into 50 ns chunks to obtain accurate error estimates of the data (convergence was established at 50-100 ns for most cation-binding sites; Figure A3.1).

Both WT-lipA and QM-lipA were subjected to high temperatures (500 K) in water to nearly completely eliminate their native secondary structures (radius of gyration of ~2.0 nm). Following this, steered MD at 300 K was used to guide the enzymes back to a radius of gyration near that of the original crystal structures, in vacuum. The enzymes were then re-solvated in 20% aqueous [BMIM][Cl]. Results showed 500 ns to be more than sufficient to converge cation binding on the surface of lipase (Figure A3.1), thus a single trial was run in each case to elucidate trends

in binding specificity. The final structures were subjected to energy minimization, NPT equilibration, and 500 ns production NPT simulations.

Results and Discussion

7.3.1. Comparing the Locations of Cation Binding to the Folded State from MD with Experiments

Twelve major locations for IL cation binding on the surface of WT-lipA were identified from the crystal structures and refined through MD, as described in the Computational Methods section. Figure 7.1 provides a visual representation of the type and location of each binding site on the enzyme's surface, where the unique coloring of each site is correlated with figures presented throughout the manuscript. Table A3.1 of the Supporting Information (SI) includes a full list of the residues identified in each of the binding sites on WT-lipA.

Half of the sites include aromatic residues that interact with the imidazolium ring of IL cations through pi-stacking interactions. Other binding site residues mainly consist of amino acids with hydrophobic side chains or long-chained amino acids that, despite a positively-charged head group (i.e., LYS in site 5 and ARG in sites 1, 7), also interact through hydrophobic interactions with the butyl tail of IL cations. Notably, only one site consists of a negatively charged amino acid (i.e., ASP in site 7), indicating hydrophobic interactions between surface residues and BMIM cations generally dominate over electrostatic interactions.

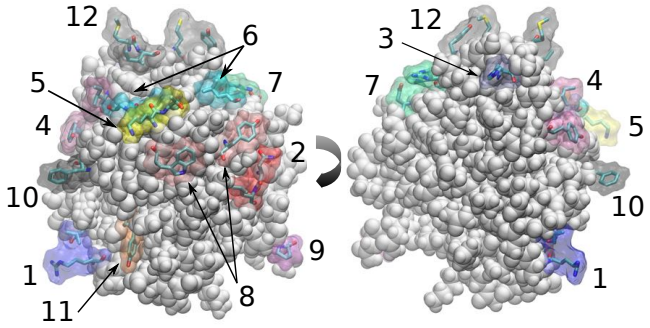

Figure 7.1. Visual representation of the 12 IL cation binding sites on the surface of WT-lipA determined from crystal structures derived from experiments. Binding site residues are shown in both licorice and space-filling representation, and hydrogen atoms are excluded for clarity.

To determine whether atomistic simulations are suitable for predicting *a priori* the interfacial behavior of ILs at enzyme surfaces, the percent occupancy for every residue on WT-lipA by IL cations was calculated at all IL concentrations using MD simulations initiated from a randomized solvent box (i.e., the IL ions were not initially bound at the experimental locations). Percent occupancy refers to the fraction of frames over the MD trajectory for which an IL cation

was within 4 Å of a given protein residue, following the experimental definition of a contact. Figure 7.2 shows these results with the same coloring scheme used in Table A3.1 and Figure 7.1. The results show the expected increase in cation binding to most residues with increasing IL concentration, in line with crystallographic and NMR findings (n.b. buried residues maintain approximately zero percent occupancy at all concentrations). This can be explained by increased random collisions of cations with the surface at higher IL concentrations, as well as by faster replacement of one cation for another at binding sites. At all concentrations, simulations and experiments agree upon the most strongly binding residues; there are no residues on the enzyme's surface that simulations predict can bind cations more strongly than those determined by experiments.

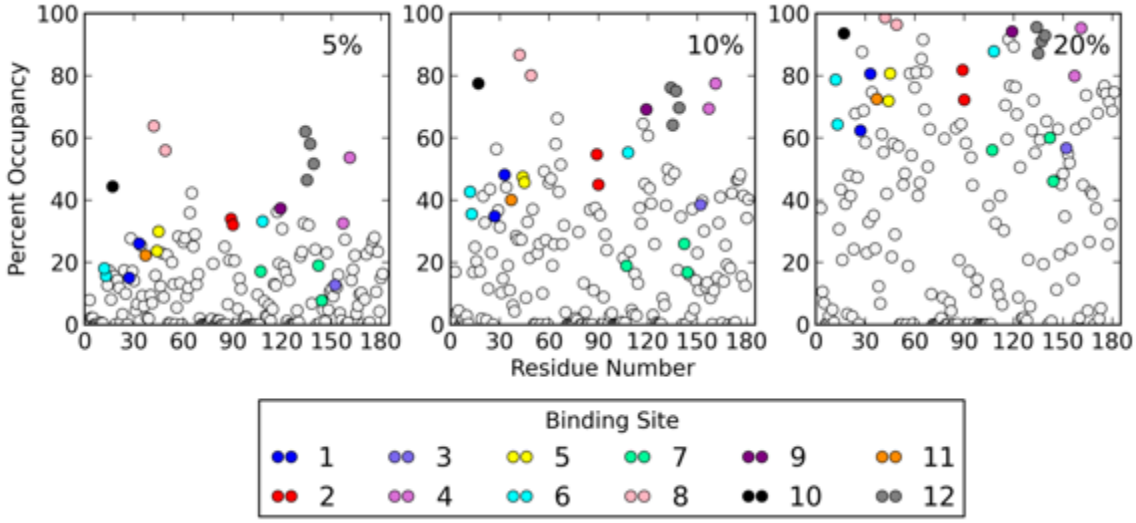

Figure 7.2. Percent occupancy of residues in WT-lipA by IL cations as a function of IL concentration in water. Colored circles represent experimentally identified binding sites, and uncolored circles represent all other protein residues.

These results demonstrate the utility of using molecular simulations to predict the best target residues for future rational design efforts to stabilize enzymes against IL-induced denaturation. To our knowledge, this study represents the first quantitative comparison of measured structure and MD simulated structure of enzymes in ILs. The IL and enzyme are both simulated with AMBER family force fields, although many others have been used to simulate enzyme/IL systems.[333] In the future, this data set and our analysis approach could be used for similar validations of MD performed with different IL/enzyme potentials.

Looking beyond just the top binding residues, Figure 7.2 shows that at all IL concentrations, our simulations predict approximately seven of the 12 binding sites to be in the top 10% of all residues in WT-lipA for binding IL cations. Standard deviations in cation occupancy calculated over the entire 5% IL solution simulation trajectory (10 data points spaced 50 ns apart) for all 12 binding sites are less than 5% in every case (n.b., with the exclusion of a single outlier at 50 ns for binding site 11); results for the top five binding sites are shown in Figure A3.1. It is reasonable to assume that the more concentrated IL solutions would converge even faster; therefore, we conclude our results to be converged at all concentrations and note that similar accuracy in prediction for the top sites could be achieved with much less sampling time (Figure A3.1). In 4 of the remaining sites, cations were stabilized by crystallographic artifacts, including monomer-

monomer contacts within the unit cell and between unit cells (Figure A3.2), which create non-native binding sites.

The origin of the low percent occupancy of binding site 3 predicted by the simulations at all IL concentrations was traced to repulsion of the imidazolium cation by the positively charged histidine residue in binding site 3. It is not surprising that a histidine residue with a hydrogen atom left in place as in the best-fit structure (see Computational Methods) repels IL cations, leading to low occupancy values. At the pH under consideration for typical enzyme/IL studies, histidine is usually found with a positively charged side chain (+1). However, the pKa of histidine in the IL environment is unknown, so we also considered the neutral side chain form. Therefore, two additional 500 ns simulations were performed with both isomers of the neutral side chain histidine. In both cases, this resulted in large increases in cation binding, qualifying site 3 to be in the top 10% of residues in WT-lipA for binding IL cations, in line with the experiments.

As Figure 7.2 shows, it becomes increasingly difficult at the higher IL concentrations to separate out the important binding residues from those that are frequented by random collisions due to a high solvent exposed surface area. Additional constraints can be imposed on the data to establish a baseline across the different IL concentrations and reduce the random noise. An example is the residence time of cation binding: while percent occupancy measures the frequency of binding to a given protein residue, residence time measures the strength of binding to that residue or, in effect, how long on average a cation stays bound to that site. These results could then be further filtered by calculating the average time for the cation to diffuse 4 Å from the bulk IL solutions to the enzyme surface; residues with residence times above these cutoff diffusion values would indicate important binding sites for further consideration. Calculations of cation residence time for all IL concentrations are described in the SI and are shown in Figure A3.3 for WT-lipA and Figure A3.4 for QM-lipA. The results are similar to the results of Figure 7.2 in that nearly half of the binding sites are predicted to be in the top 10% of all residues in WT-lipA for cation interaction strength. Notably, residue V62 was identified from Figure A3.3 as a candidate for mutagenesis due to its ability to bind cations for relatively long times (Figure A3.5). In support of this conclusion, while no cations were found to be bound to V62 in the crystal structures, NMR experiments[8, 141] indicated V62 was notably perturbed by the presence of IL. Overall, our results indicate that current molecular force fields can recapitulate experimental observations of the preferred enzyme surface binding locations of IL cations, albeit with some differences that can be easily explained.

It was also investigated whether IL perturbed the structure of the enzymes during our simulations, which we hypothesized could impact our binding calculations. The root-mean-squared deviation (RMSD) of WT-lipA's Cα atoms from the energy-minimized crystal structures shows global structure remained stable in all IL solutions over the entire 500 ns (Figure A3.6). Figure A3.6 also shows the RMSD decreased with increasing IL concentration, indicating that for this enzyme/IL system, the enzyme was globally stabilized in the IL solutions compared to in water. Plotting the root-mean-squared fluctuation (RMSF) of the enzyme's Cα atoms in the presence of the different IL solutions highlights several flexible loop regions where fluctuations were suppressed in the IL solutions (Figure A3.7). Figure A3.8, which plots the RMSF as a function of cation percent occupancy for all IL concentrations, shows that these changes in enzyme conformational dynamics did not lead to prominent differences in ion binding.

7.3.2. Comparing the Locations of Anion Binding to the Folded State from MD with Experiments

Arguably as important, if not more important, than accurately predicting binding as a means of stabilizing enzymes in IL solutions, is predicting lack of binding. Simulations were performed for 500 ns with QM-lipA in the same IL solutions to determine the ability of our simulations to predict reduced IL binding to mutated surface residues. Based on prior experiments, the mutations K44E and R57E in QM-lipA reduced anion binding while the mutations G158E and Y49E in QM-lipA reduced cation binding.[8, 141] Combined, while reducing binding, the mutations contributed to the dramatic stabilization of QM-lipA relative to wild-type lipA in the presence of [BMIM][Cl].[141]

Tables 7.1 and 7.2 show percent occupancies calculated from our simulations for the four residues before and after mutation, respectively (Figure A3.9 shows data for all residues in QM-lipA). The simulations predict all four mutations would lead to near complete elimination of anion binding to those residues at all IL concentrations (Figure 7.3). These findings indicate molecular simulations can directly elucidate the effect of mutations on the binding of ILs to enzymes in the absence of experimental validation, which can be extremely challenging given that the use of high-field NMR or crystallography is not always permissible.

Table 7.1. Percent occupancy of IL cations and anions to residues in WT-lipA before mutation for all IL solutions.

WT-LipA	5% IL		10% IL		20% IL	
Residue	Cation	Anion	Cation	Anion	Cation	Anion
K44	24	15	47	28	72	54
Y49	56	12	80	18	96	39
R57	28	20	49	40	81	72
G158	27	13	8.6‡	8.6‡	63	33

Table 7.2. Percent occupancy of IL cations and anions to mutated residues in QM-lipA for all IL solutions.

QM-LipA	5% IL		10% IL		20% IL	
Residue	Cation	Anion	Cation	Anion	Cation	Anion
E44	42	1.1	71	2.3	86	5.5
E49	34	2.1	47	3.8	80	8.2
E57	35	0.85	56	1.7	78	4.7
E158	47	6.6	60	5.5	82	14

Also in line with experiments, the Y49E mutation is predicted to result in large reductions in cation binding in all solutions, again indicating that hydrophobic/aromatic interactions are often more important than electrostatic interactions between BMIM cations and enzyme surface residues. Our simulations indicated the other three mutations, however, would result in minor increases in cation binding (relative to the decreases in anion binding).

As all four mutations were ultimately stabilizing to the enzyme in the experiments, anion binding likely plays a larger role in IL-induced enzyme deactivation for the systems we studied. This is in agreement with prior experimental studies that showed the activity of enzymes in anhydrous ILs was anion-dependent[334], although the mechanism of inactivation of enzymes by the anion in neat ILs may differ from that in IL solutions.

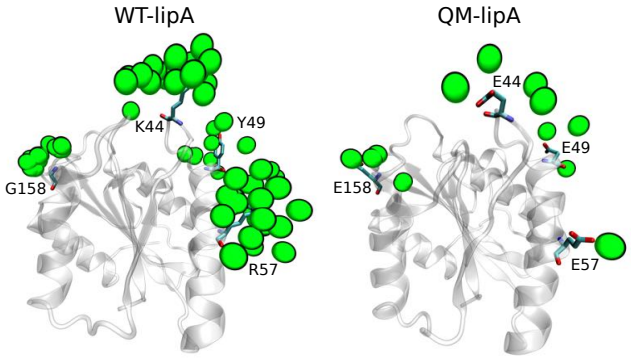

Figure 7.3. Superimposed IL anion binding poses from simulations of 20% IL around residues in (left) wild-type lipase and (right) mutant lipase. Anions are shown as green spheres. C, O, and N atoms are shown in in cyan, red, and blue, respectively, and protein secondary structure is shown in gray. Hydrogen atoms are not pictured for clarity.

7.3.3. Analysis of IL Binding to the Unfolded State

After validating the accuracy of MD force fields for studying the interactions of enzymes with both IL cations and anions, we performed additional simulations to investigate characteristics of IL binding to the unfolded state of lipase. WT-lipA and QM-lipA were denatured at high temperature in vacuum and resolvated in 20% aqueous [BMIM][Cl] (see Computational Methods for more details). We analyzed the changes due to the denaturation and mutation by comparing, on a site-by-site basis, how each residue differentially binds IL cations and anions throughout the simulation. Figure 7.4 shows this difference, comparing the tendency for a residue to bind ILs strongly or weakly in the wild-type with the change in binding on mutation.

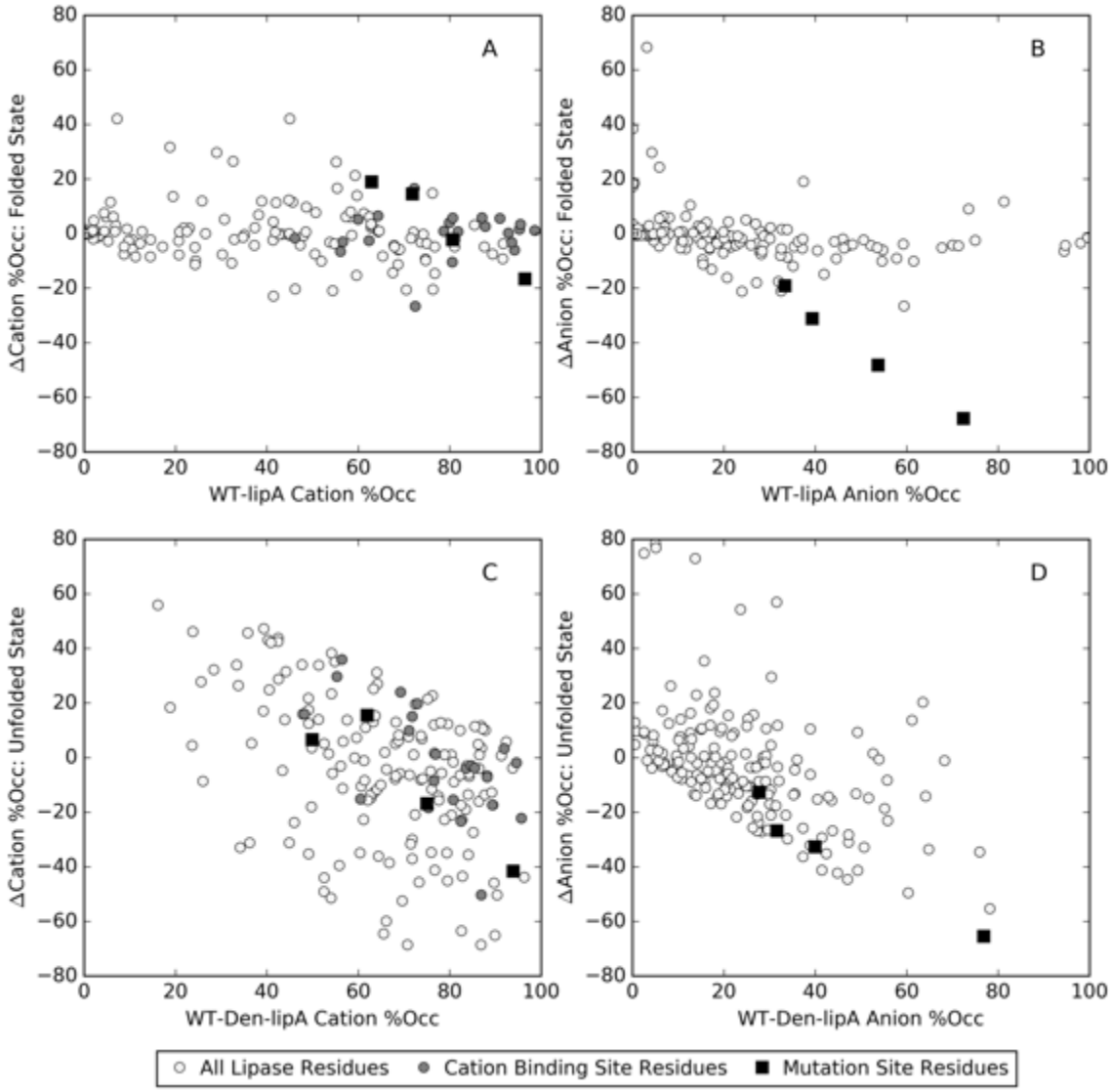

Figure 7.4. Determining the role of mutations in changing residue specific binding and protein folding. The top panels (A/B) show the percent occupancy by IL cations (A) and anions (B) in the wild-type (WT-lipA) (x-axis) vs the change in percent occupancy (mutant minus wild-type) from a 20% aqueous [BMIM][Cl] solution at 300 K. The bottom panels (C/D) show the effect of ion binding on the denatured lipase (WT-Den-lipA). Shaded circles represent experimentally identified binding sites, open circles represent all other protein residues, and squares represent mutated residues. All y-axes (change in ion binding occupancy) are presented as mutant minus wild-type.

Looking first at the change in cation binding to the folded state of the enzymes (Figure 7.4A), the results show that at the four sites of mutation (black squares), cation binding increased to two of the mutated residues, remained approximately the same to one residue, and decreased to the fourth residue, consistent with the data in Tables 7.1 and 7.2. For the majority of residues in lipA, including the 12 identified surface binding sites (gray circles), the mutations lead to little change in cation binding between the mutant and wild-type enzyme in the folded state (i.e., the

average change in percent occupancy is ~0 across all residues). Also, consistent with the results of Tables 7.1 and 7.2, Figure 7.4B shows all four mutations lead to large decreases in anion binding to those particular residues in the folded state. However, similar to the cation-binding results, Figure 7.4B shows the mutations result in little change in anion binding to most other residues. Therefore, we can infer that the mutations generally led to local rather than global changes in binding to the folded state of lipA.

Remarkably different behavior emerges when considering changes in IL binding to the unfolded state of wild-type and mutant lipA. Figures 7.4C and 7.4D show large reductions in cation and anion binding, respectively, to residues in the unfolded state of the mutant enzyme that had high percent occupancies in the unfolded state of the wild-type enzyme (WT-Den-lipA), implying that mutations lead to global rather than local changes in binding to the unfolded state of lipase. These results are even more pronounced when considering the large structural changes the enzymes underwent in the simulations (Figure A3.10). Figures 7.4C and 7.4D also show some increases in both cation and anion binding to residues in the unfolded mutant that had relatively low percent occupancies in the unfolded wild-type. This is likely due to the inhibition of IL binding to the higher occupancy sites by the mutations, leading IL ions to seek out other, less favorable protein contacts. Overall, however, subtracting the average percent occupancy of the unfolded mutant from that of the unfolded wild-type leads to a 4.2% decrease in both cation and anion binding to the unfolded state, compared to a negligible 0.46% increase in cation binding and a 1.6% decrease in anion binding to the folded state of lipase (calculated only for the 20% IL solution for consistency with the unfolded state simulations).

These findings suggest a new method for rationally designing enzymes to have increased stability in ILs by reducing binding to the unfolded state of the enzyme, without dramatically altering binding to the folded state. This could be accomplished by first using the approach described herein as a computational screen t0 identify the best candidates for point mutations to then guide experimental efforts towards this result. Specifically, by reducing binding to the unfolded state (in the absence of altering binding to the folded state), the stability of the unfolded state relative to the folded state would presumably decrease. This, in turn, would lead to a stabilizing effect on the folded state given that the overall change in free energy of unfolding of the enzyme (i.e., $\Delta\Delta G_{Den\text{-}WT}$) would increase. This hypothesis could potentially be directly tested through equilibrium dialysis experiments with the folded and unfolded state of an enzyme in solution with IL. Still, the question remains to be answered as to why there is reduced IL binding to the unfolded state relative to the folded state, which we address in the following section.

7.3.4. The Counter-Ion Effect and its Influence on Enzyme Stabilization in Aqueous IL Solutions

Pair correlation functions for the folded and unfolded state of the enzymes are displayed in the left- and right-hand plots of Figure 7.5, respectively. Consistent with earlier findings, the results show anion binding is decreased (left plot; cyan vs. green line) and cation binding is marginally increased (left plot; red vs. blue line) to the folded state of QM-lipA due to the presence of the mutations. The results also show both cation and anion binding increase to the unfolded state of the wild-type compared to the folded state of the wild-type (blue and green lines in right-

vs. left-hand plots, respectively). This is expected given the many hydrophobic core and charged residues exposed to IL in the unfolded state that are hidden in the folded state.

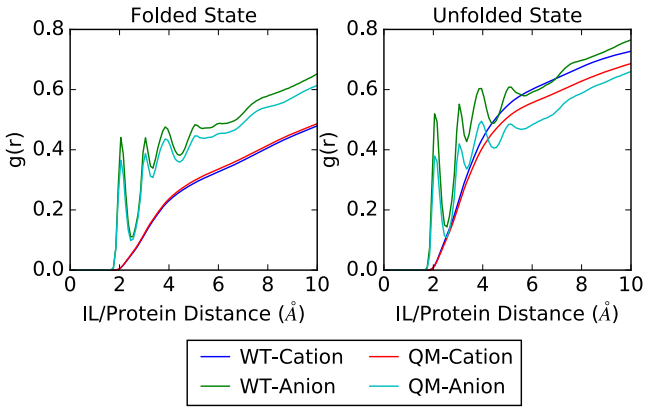

Figure 7.5. Pair correlation functions for cations and anions within 4 Å of (left) the folded state and (right) the unfolded state of wild-type and mutant lipase.

Contrary to the results for the folded state, the results show both anion binding (right plot; cyan vs. green line) and cation binding (right plot; red vs. blue line) to the unfolded state are decreased due to the presence of the mutations, which points to a mechanistic difference in IL binding to the folded and unfolded state that we call here the "counter-ion effect". In the folded state, it was previously determined that the mutations affect binding on a local scale, and that the main driving force behind the increases in cation binding to the mutation sites was electrostatic in nature. This would result in balanced charges at the mutation sites between the positively charged cations and negatively charged mutations, leading to our hypothesis that IL anions are not needed to act as counter-ions at negatively-charged mutation sites in the folded state.

Conversely, in the unfolded state, it was determined that mutations likely have global effects on IL binding. Supporting this notion, Figure 7.5 shows the decreases in anion binding between the mutant and wild-type are much larger in the unfolded state compared to the folded state. This may be due to the ability of the mutated residues to cast a wider net of anion repulsion in the flexible, ever-changing structure of the unfolded state. We further hypothesize that global exclusion of the anion from the surface of the unfolded state hinders binding of [BMIM] cations to nearby hydrophobic residues, because if cations were to bind to those residues, this would result in charge imbalances at the protein/solvent interface (i.e., positively charged cations without neutralizing counter-ions). This would explain the decreases in both cation and anion binding to the unfolded state observed in Figure 7.5, as well as earlier findings of the same average decrease in percent occupancy of 4.2% for cations and anions to the unfolded state. Thus, we believe the counter-ion effect plays a governing role in determining enzyme stabilization in IL solvents (i.e., in determining differences in IL binding to the folded and unfolded states of an enzyme), presenting new opportunities to leverage this information for future rational design efforts.

Conclusions

Over the past 10-15 years, great strides have been made in the quest to understand the interactions and role of small molecule factors on enzyme structure and function. The use of enzymes in IL solvents, specifically the massive design space offered by the solvent choice and amino acid diversity, has presented a unique challenge that calls for expanded use of molecular scale simulations to guide and interpret experimental efforts. The work presented herein has addressed this in three unique ways. First, we demonstrated the capability of MD to predict IL/enzyme binding sites that were determined by molecular resolution structural experiments. Second, we demonstrated the sensitivity of MD to predict subtle binding differences that emerged in experiments only from a few point mutations. Our results further demonstrated how simulations can be used quantitatively to reconcile differences between MD and molecular scale structural measurements such as dimer stabilization and increasing the likelihood of correct pKa assignment of side chains. We expect the finding that current molecular force fields are accurate for studying enzyme/IL interactions to have large impacts on the fields of biology and biophysics, where molecular simulations are becoming an increasingly important tool of research. Overall, these results provide strong evidence for the efficacy of employing molecular simulations to guide future rational design efforts to improve enzyme function and thus the utility of enzymes in non-conventional solvents like ILs for biocatalysis.

Finally, building on both the experimental measurements of how ILs bind in the native state and the validated MD simulation approach, we used MD to compare the binding between globally denatured and folded enzymes. We gained new fundamental insight into the mechanism of enzyme stabilization through a "counter-ion effect" in IL solvents. This approach of quantitatively comparing the binding behavior in the folded/unfolded contexts could be a powerful computational screening tool in the future and could assist in rapid identification of the best candidates for point mutations to stabilize enzymes against IL-induced denaturation.

While the results presented here denote an important step towards our understanding of IL-induced denaturation of enzymes from non-specific surface interactions, molecular insight into IL-induced deactivation of enzymes due to specific, active-site interactions is still largely missing. Additionally, the general findings in this work need to be replicated with a multitude of other enzyme/IL systems, as well as with other IL/enzyme potentials. Current work to accomplish this involves the use of automation tools and machine learning techniques to compile a database of features/descriptors that relate the binding propensity of IL substituents to enzyme surfaces, such as binding frequency, duration, and mobility of IL species once bound. Assessment of both the strong and weak IL binding sites is important; commonalities across the panel of enzymes and ILs may reveal universal features to guide design of weak-binding IL sites. Lastly, surface-binding observations could be paired with a description of the protein surface that includes the local sequence and secondary structure. Ultimately, the combination of molecular simulations with experiments will create a powerful tool to guide rational protein engineering for the case of solvent or small-molecule induced inhibition.

Chapter 8

Effect of an Ionic Liquid/Air Interface on the Structure and Dynamics of Amphiphilic Peptides[8]

Introduction

Ionic liquids (ILs) are well known for displaying a number of chemical and physical properties that have marked them as highly desirable solvents for a wide range of applications.[166, 335] In addition to their ability to dissolve a variety of substances, ILs also have a negligible vapor pressure, low chemical reactivity, and are liquid at ambient conditions, rendering them more environmentally-friendly solvents than other commonly used organic chemicals.[335] Specifically, these properties make ILs not only attractive as solvents for traditional chemical reactions, but also those catalyzed by enzymes. In recent years, both experiments and computational studies have been performed on enzyme/IL systems and have shown that many enzymes are able to retain their catalytic properties even in such a solvent as IL that is so different from their native environment[155, 336-337]. The potential for ILs and enzymes to work together to perform biocatalytic processes, which typically require harsher chemicals and high temperatures and pressures, is very promising. However, despite this potential, there is still much to understand about enzyme/IL interactions before the two can be used together for a wide variety of applications. For many enzymes, even slight changes in environmental pH, ionic strength, temperature, and especially solvent composition are known to lead to structural changes or even catalytic deactivation and/or complete protein denaturation.

Because preserving the structure of proteins and enzymes is necessary for maintaining their function and activity, it is vital to have a molecular scale understanding of the effects of these variables on protein structure if one is to predict *a priori* whether a given protein or enzyme will function well in a particular solvent environment. Most computational and experimental studies to date have focused on the effects of ILs on larger, biological proteins and enzymes.[338] This approach, while very applicable to the polypeptide being studied, can lead to difficult interpretation of broader results due to the many degrees of freedom such large system sizes introduce. A small

[8]Reproduced in part with permission from K.G. Sprenger, Kovas Palunas, Tobias Weidner, and J. Pfaendtner. Effect of an Ionic Liquid/Air Interface on the Structure and Dynamics of Amphiphilic Peptides. Journal of Molecular Liquids, 236: 404-413, 2017. Copyright 2017 Elsevier.

handful of experimental studies have been performed with model protein compounds such as amino acids, peptides, or mini-proteins with well-defined secondary structures to better understand the multifaceted effects of ILs on more complex biomolecules.[338-340] ILs are also increasingly being used as reaction media for peptide synthesis and separations,[341-342] leading to increased focus on peptide/IL interactions that could improve our understanding of protein/IL interactions.

Atomistic simulations have proven to be an invaluable tool at providing insight into the complex interactions of ILs and proteins through the study of model systems. Recently, the enhanced sampling tool metadynamics (MetaD) has been used to study the kinetics of IL-induced unfolding of HP35, the 35-residue subdomain of the villin headpiece,[331] and the unfolding mechanisms of a small β-hairpin peptide in IL solution.[343] Additionally, replica-exchange MD simulations have been employed to study the folding pathways of trp-cage mini-protein and a small β-hairpin region of protein G in IL solutions.[344] Understanding on a molecular level how the structure and dynamics of different secondary structure motifs change when exposed to IL will enable efficient rational design of proteins and enzymes to have improved tolerance in IL solutions and thus will improve the utility of protein/IL mixtures for different chemical processes.

Synthetic peptides composed of leucine-lysine (LK) repeats[11] have been studied extensively in a wide variety of different conditions[4, 21, 345-346]. These studies have shown that LK peptides are not only excellent models for secondary structure motifs, such as α-helices and β-sheets, but can also self-assemble into complex, ordered structures made up of multiple peptides.[347] Moreover, in the right environments, LK peptides can template the ordered growth of mesoscale morphology of inorganic materials like silica.[4, 347-348] A recent study by Lutz et al. showed that at the air/water interface in particular, different silica morphologies could be templated from different LK peptides and that these variations were driven by the specific secondary structures of the peptides.[347] This unique finding, pointing to an interfacial effect with LK peptides, is one of the major motivations for the current study of LK peptides at the interface of air and aqueous IL solutions. The diversity of behavior described above makes LK peptides ideal candidates for examining general protein behavior in non-conventional environments, such as in IL solvents. Additionally, because of the relatively small size of LK peptides, it is possible to exhaustively study their behavior with unbiased molecular dynamics (MD) simulations.

Two synthetic LK peptides were chosen as subjects for this study: LKα14 with an amino acid sequence of LKKLLKLLKKLLKL and LKβ15 with an amino acid sequence of LKLKLKLKLKLKLKL. As their names suggest, these peptides were designed to form α-helices and β-sheets, respectively, at air/water interfaces.[11, 345-346, 348] LKα14 in particular has been shown to maintain a helical structure and self-aggregate to form tetrameric structures both at the air/water interface and in bulk water.[345-347] The ubiquity of the α-helix and β-sheet in natural proteins informed our decision to choose LKα14 and LKβ15 as our model peptides. Formation of the secondary structures of the two peptides arises due to the different periodicities at which the hydrophobic leucine residues appear in the peptides' amino acid sequences. Hydrophobic interactions largely drive the secondary structure formation, allowing the correct backbone hydrogen bond pairings required for different secondary structures to form.[11] A change in the strength of these interactions, perhaps caused by the presence of a particular solvent such as IL, could therefore have a dramatic effect on the structure of the peptides.

This study aims to: 1) provide fundamental insight into the interactions of model LK peptides and IL solvents as a means for predicting protein behavior in non-conventional solvent environments, and 2) evaluate interfacial effects on the structure and transport properties of the

peptides as a step towards understanding mechanisms of LK peptide-induced templating at interfaces. Classical molecular dynamics (MD) simulations were performed on systems of both LKα14 and LKβ15 in three different solvents, namely pure water, a 50% (w/w) mixture of aqueous 1-butyl-3-methylimidazolium chloride ([BMIM][Cl]), and a 50% (w/w) mixture of aqueous 1-butyl-3-methylimidazolium acetate ([BMIM][Ace]). Each peptide/solvent system was also simulated with the addition of a vacuum/solvent interface, approximating the air/solvent interface. Additionally, biased MetaD simulations were run to calculate free energies of adsorption of LKα14 to the vacuum/solvent interface in IL versus in pure water. Unbiased simulations show the presence of IL induces the formation of many unique peptide secondary structures both in solution and at the interface, including several interesting bulk aggregate structures. Biased simulations show that the presence of IL drastically reduces the free energy of peptide adsorption to the vacuum/solvent interface.

Methods

The GROMACS 4.6.5[349] or 5.1.2[104] MD engine was used with the TIP3P water model[181] and the AMBER99SB-ILDN force field[51] to describe the peptides. The ILs were parameterized with the general Amber force field (GAFF)[182], which has been shown to accurately reproduce both thermodynamic and transport properties of a wide range of ILs[166]. Ion charge distributions were generated with Gaussian[175] at the Hartree-Fock level of theory using the 6-31 G* basis set. The resulting IL partial charges were then scaled by multiplying each by a factor of 0.8 to better approximate the dynamical properties of the ILs, as suggested by Maginn et al[172].

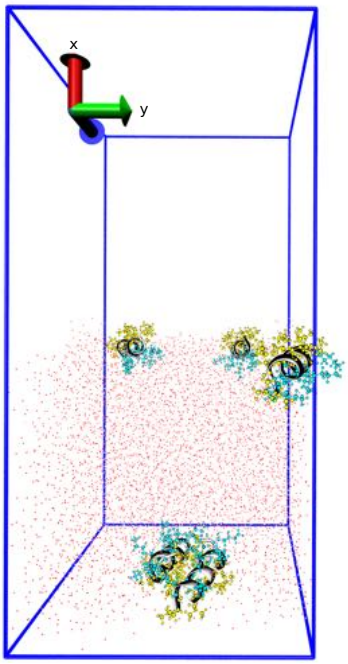

Figure 8.1. Example snapshot of an interfacial simulation of LKα14 peptides in water. Peptide secondary structures are shown in gray, atoms in leucine residues in yellow, and atoms in lysine residues in cyan (CPK format). Solvent oxygen atoms are represented as red beads.

All simulation systems were built by randomly placing five peptides in solvent in a roughly 6.0 x 5.7 x 5.7 nm box (Figure 8.1). The two IL solutions consisted of 50% (w/w) IL in water. The psfgen plugin for Visual Molecular Dynamics (VMD[47]) was used to generate the peptides, which were added to the simulation boxes in randomly unfolded configurations. To mimic experimental conditions, the LKα14 and LKβ15 peptides were capped with a deprotonated carboxylate group, respectively imparting them overall charges of +5 and +6. Cl- counterions were added to the systems to achieve overall charge neutrality. Each Cl- counterion had a charge of -1, leading to two different "types" of Cl- ions in some of our simulations (i.e., [BMIM][Cl] also contains Cl- ions, each with a charge of -0.8). To simulate a vacuum/solvent interface, six nm of vacuum was added above the solvent in the x-dimension for all interfacial systems. Full periodic boundary conditions were employed in all dimensions for all systems, effectively creating two vacuum/solvent interfaces at either end of the solvent box for the unbiased, interfacial simulations. On average, the systems contained 18,500-24,000 atoms and were generated randomly using Packmol[178]. An additional interfacial system containing a single LKβ15 peptide was generated in a similar fashion. Three trials of each unbiased simulation were carried out with reinitialized starting configurations. Given the triplicate trials of the three different solvents, two peptides, and two environments (bulk and interfacial), the aggregate sampling time totals to 7.2 μs, not including biased or supplementary (e.g. lone peptide) simulations. A summary of all the simulations carried out in this work is given in Table 8.1.

Table 8.1. Summary of simulation parameters. Note that all unbiased simulations were run in triplicate. The average number of particles in unbiased simulations is ~18,500 and is ~24,000 in biased simulations.

Simulation	Peptide	Solvent	Number of Peptides	Interfacial (I) vs. Bulk (B)	Length of Simulation (ns)	Type of Simulation
1	LKα14	Water	5	I	200	Unbiased MD
2	LKα14	50%[BMIM][Ace]	5	I	200	Unbiased MD
3	LKα14	50%[BMIM][Cl]	5	I	200	Unbiased MD
4	LKα14	Water	5	B	200	Unbiased MD
5	LKα14	50%[BMIM][Ace]	5	B	200	Unbiased MD
6	LKα14	50%[BMIM][Cl]	5	B	200	Unbiased MD
7	LKβ15	Water	5	I	200	Unbiased MD
8	LKβ15	50%[BMIM][Ace]	5	I	200	Unbiased MD
9	LKβ15	50%[BMIM][Cl]	5	I	200	Unbiased MD
10	LKβ15	Water	5	B	200	Unbiased MD
11	LKβ15	50%[BMIM][Ace]	5	B	200	Unbiased MD
12	LKβ15	50%[BMIM][Cl]	5	B	200	Unbiased MD
13	LKβ15	50%[BMIM][Ace]	1	I	200	Unbiased MD
14	LKα14	Water	1	I	300	Metadynamics
15	LKα14	50%[BMIM][Ace]	1	I	300	Metadynamics

Energy minimization was performed on each system for 1,000 steps using a steepest descent method with a 1.2 nm cutoff for all interactions. Following this step, a two ns NPT simulation at ambient conditions was run on the interfacial systems to equilibrate the box dimensions before adding the vacuum layer. These short runs used an isotropic Berendsen

barostat[109] to maintain the systems at atmospheric pressure. Production simulations were run for 200 ns in either the NVT ensemble (for interfacial simulations) or in the NPT ensemble using the Parrinello-Rahman barostat[222] (for bulk simulations). The simulations were run with a time step of 2 fs using a leap-frog integrator and a velocity rescaling algorithm[54] to maintain the system temperature at 310 K. To calculate van der Waals forces, a Lennard-Jones potential was used with an interaction cutoff of 1.2 nm. Coulombic forces were calculated with the particle mesh Ewald (PME) method and an interaction cutoff of 1.3 nm. All other force calculations were restricted to be between neighboring particles within 1.3 nm of each other to speed up calculations, with neighbor lists updated every 10 steps. The LINCS algorithm was used to constrain all covalent bonds.[110]

VMD was used to analyze the simulation trajectories, along with the PLUMED 2.0 plugin tool[52] and in-house scripts. We note that all analysis performed on unbiased simulation trajectories excluded the first 100 ns of simulation time to avoid the effects of any artifacts from the starting configurations generated by Packmol. Additional secondary structure analysis was carried out with DSSP (the Dictionary of Protein Secondary Structure)[350] and GROMACS tools[349]. Following the protocol of Donovan et al.,[351] interfacial regions were defined to be where the water density was less than the bulk density; this information was obtained from water density data as a function of distance for each unbiased peptide/IL system (cf. Figure 8.5, discussed in the Results and Discussion). Using this method, interfacial regions were defined to be between 0-1 nm and 5-6 nm from the bottom of the simulation boxes. The interfaces themselves were defined as average planes within these regions, at 0.5 and 5.5 nm, respectively, which generally aligns with the locations of the peptides' Cα atoms at the interface (cf. Figure 8.5). Figure A4.1 in the Supporting Information demonstrates the location of the interface in each system remains stable over the course of the simulation, with fluctuations (~0.2-0.4 nm) that are much smaller than the widths of the peptide interfacial density peaks in Figure 8.5 (~1.0 nm). For analyses that required isolation of interfacially-bound peptides, each simulation trajectory was trimmed of all peptides that did not adsorb to an interface (i.e., did not have a center-of-mass that fell within the defined interfacial regions), and was then split up by peptide into multiple single-peptide trajectories. In this way, every peptide's interfacial behavior could be analyzed independently of its bulk behavior.

MetaD simulations were constructed in the same manner as unbiased interfacial simulations with an additional two nm of solution added in the x-direction to allow for accurate binding free energy calculations. A single LKα14 peptide was simulated in both a 50% aqueous [BMIM][Ace] solution and a pure water solution. Energy minimization was performed in each case for 40,000 steps using a steepest descent algorithm. Prior to the addition of the interface, a short two ns NPT simulation was run to equilibrate the systems at 310 K and 1 bar. Steered Molecular Dynamics (SMD) was then employed to initiate four replicas in different parts of collective variable (CV) phase space for the ensuing production simulations. Time-dependent harmonic restraints with a force constant of 10^3 kJ/nm^2 were applied to SMD atoms defined in two CVs: the x-distance between the center-of-mass of all the atoms in the peptide and a frozen particle in vacuum with no charge or Lennard Jones parameters (CV1), and the number of backbone α-helical hydrogen bond contacts in the peptide (CV2). A switching function was used to define contacts for CV2 between oxygen and hydrogen atoms four residues apart (i, i+4), with a reference bond length of 2.5 Å. These same CVs were used in subsequent multiple walker MetaD simulations, biased at 310 K with Gaussian hill widths of 0.05 and 0.010 nm, respectively, an initial hills deposition rate of 3.0 kJ/mol/ps, and a biasfactor of 40. Molecules in the bottom nm of

solution were frozen in the simulations, resulting in a single interface in each system. Upper and lower walls were placed on CV1 in the form of harmonic restraining potentials with force constants of 10^6 kJ/nm^2.

Results and Discussion

8.3.1 IL Effects on LK Peptide Affinity for the Vacuum/Solvent Interface

Ionic liquid solvents had a substantial effect on peptide adsorption to the vacuum/solvent interface in our simulations. The Einstein relation[352], shown in Equation 1, was used in conjunction with the peptide mean squared displacement to determine the effects of ILs on peptide diffusion through solution:

$$D = \frac{1}{6}\lim_{t\to\infty}\frac{d}{dt}\left\langle \sum_{i=1}^{N}\left[\vec{r}(t)-\vec{r}(0)\right]^2\right\rangle \tag{8.1}$$

where D is the diffusivity, t is time, N is the number of particles being sampled for diffusivity (in our case the centers-of-mass of each of the five peptides). Diffusivity calculations were carried out only for bulk systems to avoid artifacts in the results caused by the presence of the interface. Table 2 shows the diffusivity results for both LKα14 and LKβ15 in the two IL solvents and in water; the reported values are averaged across the five peptides in each system and across all three trials. Additionally, data from the first half of each trajectory was excluded to avoid the effects of any artifacts from the initial setup conditions. The results show both IL solvents reduce the diffusivity of the peptides compared to water by approximately an order of magnitude in most cases. This result is expected, considering the high viscosity of ILs relative to water, and demonstrates that our molecular models are at a minimum qualitatively capturing important physics of the systems. Table 8.2 further shows that in all three solvents, the diffusivity of LKα14 is marginally greater than that of LKβ15, and that the diffusivity of both peptides is slightly reduced in [BMIM][Cl] compared to in [BMIM][Ace]. We note that the reported diffusivities may deviate somewhat from microcanonical dynamics due to the use of a thermostat; however, we expect any errors to be proportional across the similar systems studied herein and thus to have little impact on our conclusions. Additionally, our previous work calculating IL self-diffusivities from simulations in the NPT ensemble showed reasonable agreement with experimental data.[2]

Figure 8.2 shows the time-dependent diffusive behavior of each of the five LKα14 and LKβ15 peptides to both the vacuum/water interface and the vacuum/solvent (50% [BMIM][Ace] in water) interface. The results demonstrate that almost immediately, all five peptides in the pure water system migrated to one of the two vacuum/solvent interfaces that exist at x-positions of ~0.5 and 5.5 nm. Conversely, the two IL solutions greatly hindered the ability of the peptides to reach the interfaces. This is demonstrated by the approximately 50 ns average timespan needed to reach an interface compared to about five ns in water, excluding two LKα14 outliers. The results also show that the affinity of the peptides for the interfaces is reduced in IL, supported by the fact that several peptides adsorb to interfaces before detaching themselves and returning to bulk solution in the 200 ns of sampling. In water, however, all five peptides stayed at the interface for the remainder of the simulation after reaching it.

Table 8.2. Diffusivities for all unbiased, bulk systems in 10^{-5} cm^2/s, averaged across the five peptides in all three trials of each system and fit over the first 40 ns of the 2nd half of each trajectory. Error bars represent the standard error of the mean, where the sample size (n=15) is the total number of individual peptide trajectories for the three replicate simulations.

	LKα14	LKβ15
Water	$2.67 \times 10^{-1} \pm 0.02 \times 10^{0}$	$1.56 \times 10^{-1} \pm 0.01 \times 10^{0}$
[BMIM][Ace]	$3.18 \times 10^{-2} \pm 0.01 \times 10^{-1}$	$2.18 \times 10^{-2} \pm 0.01 \times 10^{-1}$
[BMIM][CI]	$2.60 \times 10^{-2} \pm 0.02 \times 10^{-1}$	$1.39 \times 10^{-2} \pm 0.01 \times 10^{-1}$

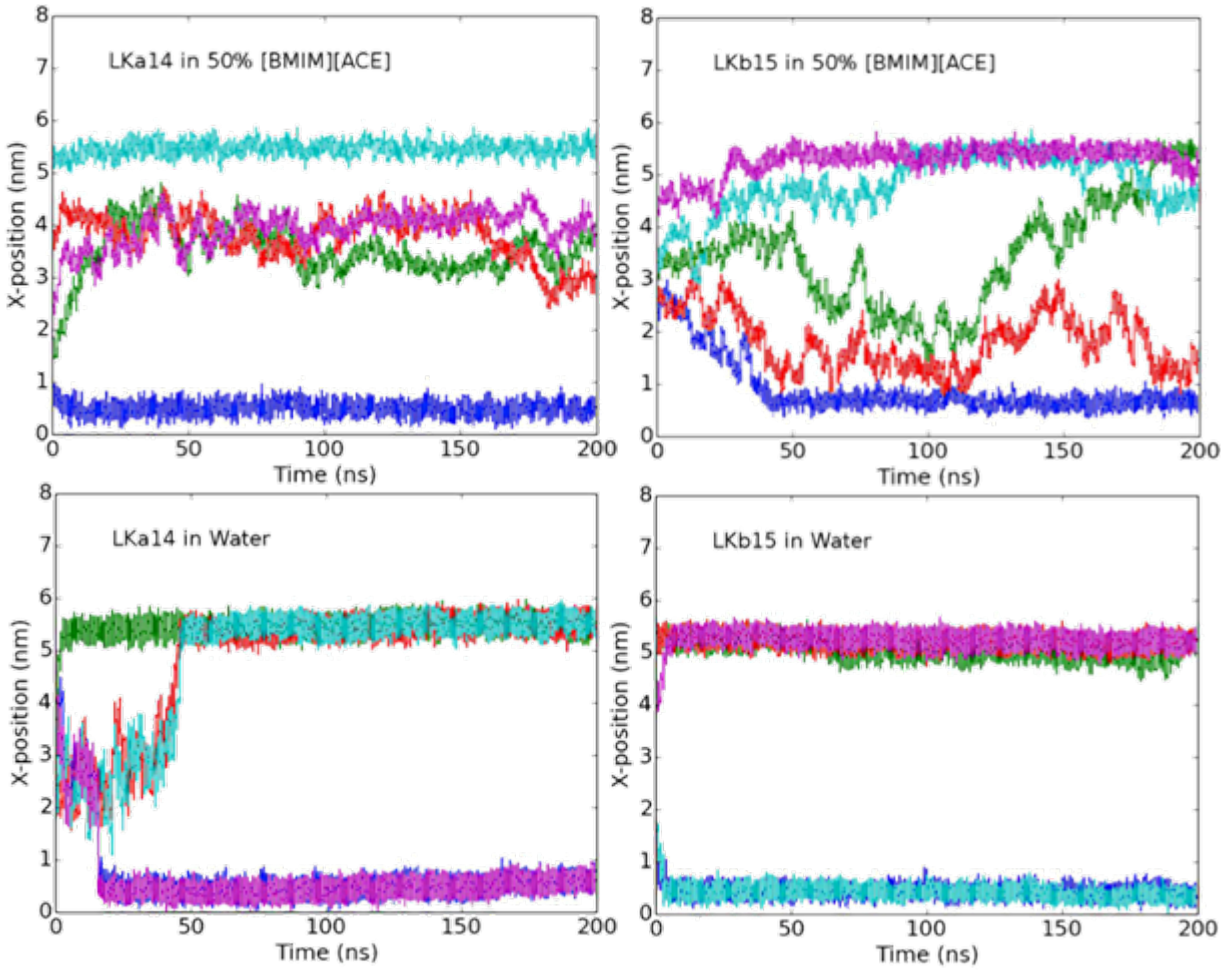

Figure 8.2. Comparison of the center-of-mass position over time of LKα14 and LKβ15 systems in water and 50% aqueous [BMIM][Ace]. Data for a single trial is shown of the three total trials for each simulation. Note that solvent is present between 0 and 6 nm, while the area above 6 nm represents vacuum. The different colors represent the five different peptides in each system.

To further elucidate the effects of ILs on the affinity of the peptides for the vacuum/solvent interface, we performed MetaD simulations (described in the Methods section) to calculate the free energy change on adsorption to the interface. These simulations were performed with LKα14 in both water and in 50% aqueous [BMIM][ACE] at 310 K. The convergence of the biased simulations is shown in Figure 8.3 and was monitored by plotting the free energy difference between the adsorbed and solvated states for the two solvents. The results show it took just over 200 ns (per replica) for the simulations to converge. However, the simulations were extended to 300 ns for further verification of convergence, leading to a total sampling time for each four-replica system of 1.2 μs.

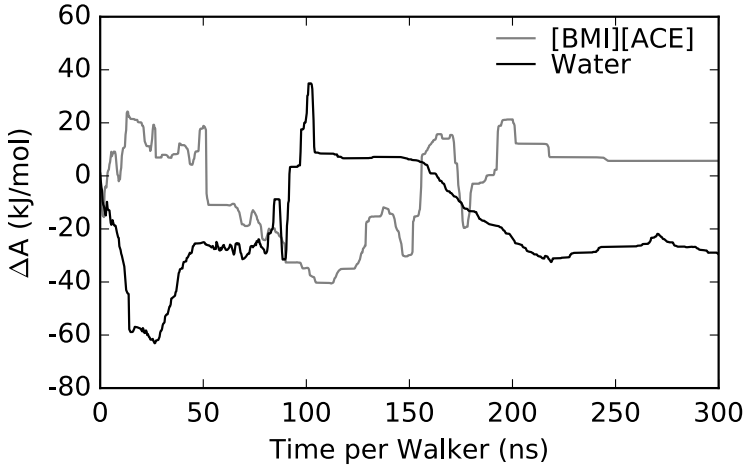

Figure 8.3. Convergence of the distance between LKα14 and the vacuum interface with water and [BMIM][Ace] for metadynamics simulations. Note that the system is still diffusing in *x* during the last 50-100 ns of each run.

Figure 8.4 shows the 1D free energy results for the separation distance between LKα14's center-of-mass and the interface, which is located at ~0.5 nm. As expected, Figure 8.4 shows the peptide has a stronger affinity for the interface in a pure water solution compared to in the bulk solution. In contrast, the IL solution greatly reduces LKα14's affinity for the interface to be nearly the same as that for the bulk solution. Overall, Figure 8.4 shows an estimated 40 kJ/mol difference in interfacial adsorption free energy for LKα14 in water versus in IL solution, obtained by calculating the Boltzmann-averaged free energy difference of the regions from 0-1.5 nm (adsorbed state) and from 2-3 nm (solvated state), and taking the difference for the two solvents. Alternatively, this large free energy change, which is consistent with the findings from our unbiased simulations, can be inferred directly from Figure 8.3 (i.e., $\Delta(\Delta A)_{\text{IL-Water}}$ at 300 ns).

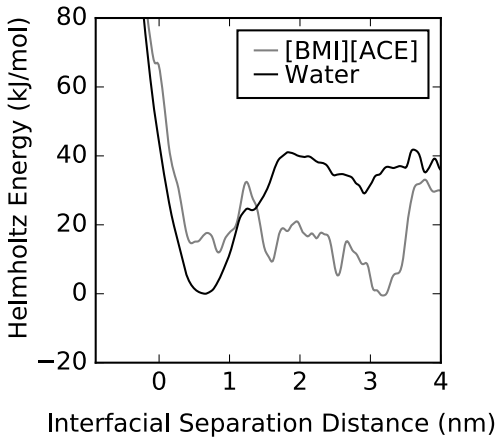

Figure 8.4. Free energy profiles with respect to distance from the interface for LKα14 in water and in aqueous [BMIM][Ace]. The interface is located at approximately 0.5 nm and the free energy is truncated at 4 nm due to the presence of a harmonic wall potential.

The differences in adsorption strength between the IL solvents and water shown in Figure 8.4 can be attributed to many different factors. Since LK peptides are composed of roughly equal numbers of hydrophilic and hydrophobic residues, it is energetically favorable for them to reside at the interface in a pure water solution, as this allows them to remove their hydrophobic sidechains from solution entirely while keeping their hydrophilic sidechains submerged in solution. Some molecules that make up ILs are similarly partially hydrophobic, such as the [BMIM] cation studied herein, and also show a free energy minimum at the interface when they can free their hydrophobic tails from aqueous solution[353]. Thus, we hypothesized that there is competition for the interface between the tails of the [BMIM] cations and the LK peptides in our systems, reducing the free energy of LK peptide adsorption to the interface in the IL solvents. Results shown in Figures 8.5 and 8.6 support this hypothesis by providing insight into the molecular populations at the vacuum/solvent interface in the various systems we tested. We note that in Figure 8.5, data for both backbone carbons (Cα) and hydrophobic sidechain carbons (Cδ) are included in the calculations to demonstrate interfacial peptide orientation, e.g., backbone alpha carbons are always further from the interface than hydrophobic leucine delta carbons. We also note that asymmetries in the peaks for LKα14 and LKβ15 arise due to the uneven number of peptides (5) in each simulation box, resulting in unequal distributions of the number of peptides at each interface.

Looking at the first two rows of plots in Figure 8.5 for the peptides in the IL solutions, we see small peaks in [BMIM] cation density occur around 0.5 and 5.5 nm in each system, near the vacuum/solvent interfaces. These peaks precede the small water peaks that occur at approximately 0.5 nm from either interface, indicating strong ordering of IL cations at the interface. Interestingly, the small water peaks visible in the IL systems are nonexistent in the pure water systems, thus we can infer that the interfacial behavior of the IL induces a small degree of ordering in the water itself near the interfaces. Additionally, we see that the density of the peptides at the interfaces is drastically reduced in the IL solutions compared to in water (bottom plots of Figure 8.5), in line with previous findings. These results, combined with the fact that the peaks in [BMIM] density occur at approximately the same distance from the interfaces as the peptide density peaks, lends support to our theory of competitive inhibition by IL cations for the interface. Specifically, we find evidence of the hydrophobic tails of [BMIM] cations competing for space at the interface; Figure 8.6 shows strong ordering of [BMIM] tail carbons at the interface over ring carbons.

The results also show interesting trends in terms of IL anion affinity for the interface. Figure 8.5 shows peaks in [Ace] density line up with closely with [BMIM] peaks at the interface, outcompeting nearby water molecules. In contrast, the peaks in [Cl] density in Figure 8.5 occur at approximately 0.5 nm from either interface, lining up instead with the water density peaks. These results are in line with previous conclusions that larger, more flexible ions like [Ace] tend to have a greater affinity for the interface, while smaller, harder ions like [Cl] tend to avoid the interface and remain in the bulk.[354] Furthermore, Figure 8.6 shows the [Cl] density peaks also align with the peaks of [BMIM] ring carbons, implying a strong electrostatic layer near the interface where the positively charged rings of [BMIM] cations are densely packed next to the negatively charged [Cl] ions. We note that while there are two "types" of [Cl] ions in our simulations (see Methods), Figures A4.2 and A4.3 in the Supporting Information show that for either LKα14 or LKβ15 in the [BMIM][Cl] solutions, respectively, the fully-charged [Cl] counterions do not behave in a disruptive manner when compared to their lesser charged brethren.

To compensate for the reduced peptide density at the interface in the IL solutions, we see additional peaks arise in the density profiles of both LKα14 and LKβ15 in bulk solution in Figure 8.5. Large density peaks of LKα14 exist near the center of the bulk IL solutions, whereas a smaller secondary peak of LKβ15 exists near one of the interfaces in the [BMIM][Ace] solution and the peaks are noticeably broader (particularly in the tail regions) in the [BMIM][Cl] solution compared to in water. In general, there appears to be a greater number of LKβ15 peptides adsorbed to the interface in the IL solution than LKα14 peptides, which is also apparent in Figure 8.2. The primary explanation for this behavior is likely related to the different aggregation tendencies of the peptides. Previous work suggests that LKα14 can form dimers and tetramers in bulk solution.[21, 347] We saw formation of LKα14 dimers and tetramers in our IL solutions as well, which correlates with the slower diffusion to the interface of some of the LKa14 peptides in the IL solutions as seen in Figure 8.2; once formed, these bulk aggregates remained for the rest of the 200 ns simulations and never migrated to an interface (n.b., we saw similar bulk aggregation behavior in all three trials in both IL solutions). Peptide-peptide interactions were also identified to be the cause for the slower diffusion to the interface of some of the LKα14 peptides in water as well (see Figure 8.2), though in this case the bulk aggregates dispersed relatively quickly and the peptides all migrated to an interface within approximately 50 ns (n.b., all three trials showed similar aggregation behavior and dispersion within ~50 ns). We therefore conclude that ILs can promote aggregation of LKα14 peptides due to the increased solution viscosity that reduces peptide diffusion compared to in water.

Aggregation of LKα14 peptides in the bulk IL solutions led to less LKα14 peptides at the interface than LKβ15 peptides, which are not known to form such aggregate complexes and did not exhibit any aggregation behavior in our simulations. In general, however, and as previously discussed the IL solutions slow the interfacial migration of both LKα14 and LKβ15 compared to water and reduce the affinity (i.e. interfacial adsorption energy) of both peptides for the interface. This is likely due to the specific interactions at play in the peptide/IL systems that are not present in the peptide/water systems, namely increased hydrophobic and electrostatic interactions between LK peptides and IL ions in bulk solution. It is probable that many of the low free energy solution states in Figure 8.4 occur when [BMIM] cation tails can interact favorably with the peptides' hydrophobic leucine sidechains. These bulk interactions would decrease the driving force of the peptides to free their hydrophobic sidechains from solution, thereby reducing the free energy change upon adsorption from the bulk as seen in Figure 8.4.

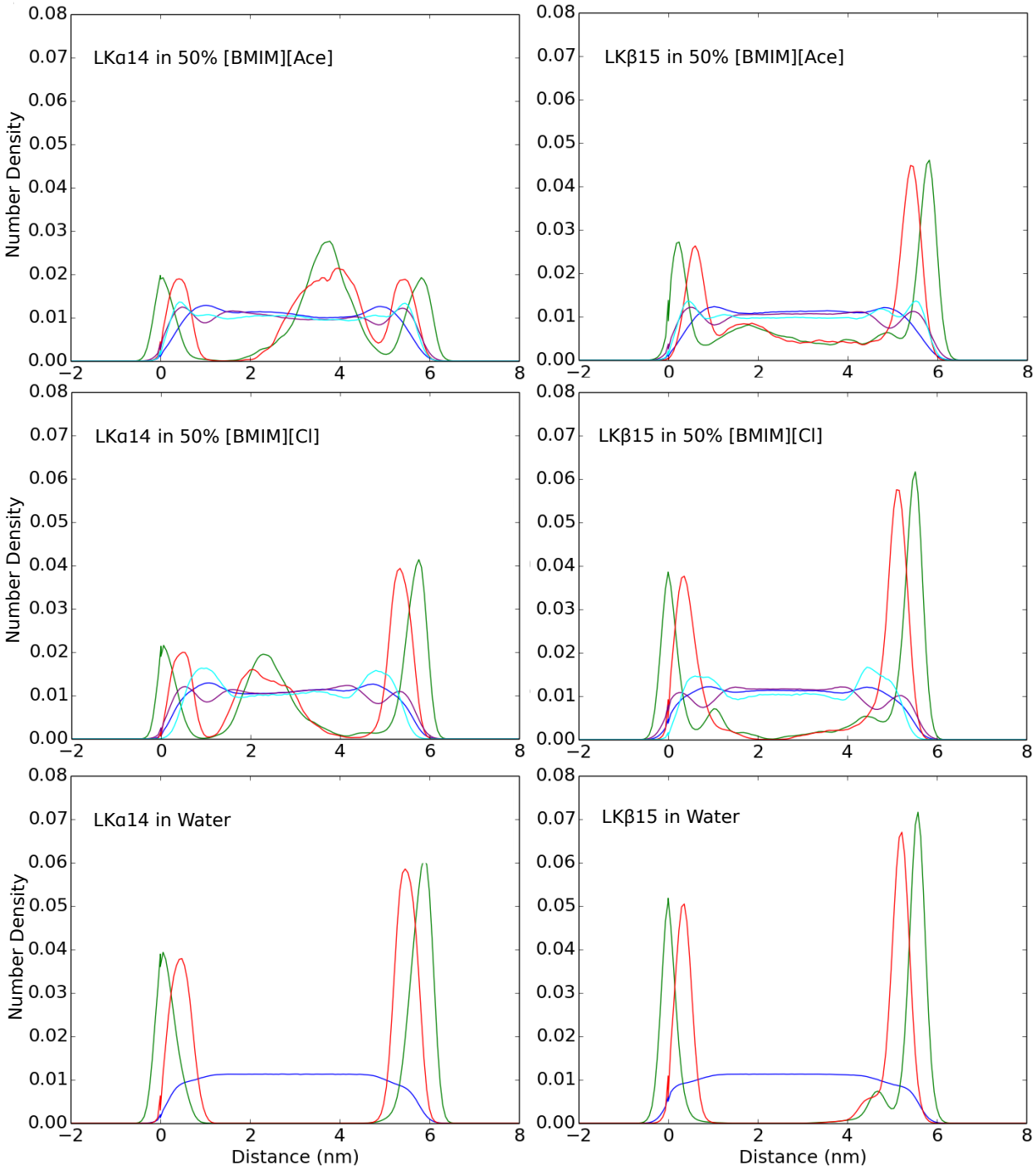

Figure 8.5. Atomic species densities for LKα14 and LKβ15 in aqueous solutions with 50% [BMIM][Ace] or [BMIM][Cl], and in pure water. Data for a single trial is shown of the three trials for each simulation. Lines are shown for water (blue), peptide backbone Cα carbons (red), peptide sidechain Cδ carbons (green), [BMIM] cations (purple), and either [ACE] or [Cl] anions (cyan). Note that (1) both types of carbons are included to show peptide orientation, (2) solvent is present between 0 and 6 nm, while the area below 0 nm and above 6 nm represents vacuum, and (3) that densities for different species are normalized only with respect to themselves for easier viewing.

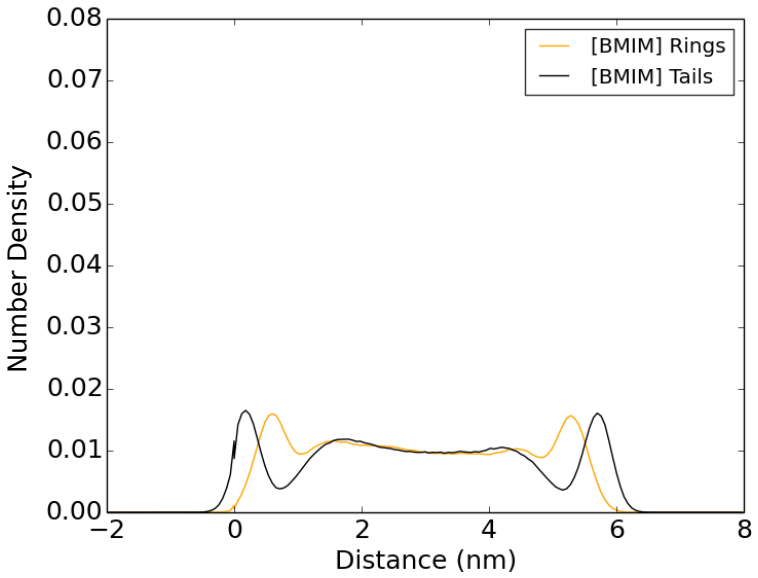

Figure 8.6. Atomic species densities for the [BMIM] cation in an interfacial system with LKα14 peptides. The black line represents the density of the last two carbons on the hydrophobic tail of the [BMIM] cation, whereas the orange line represents the density of the three carbons in the ring of the [BMIM] cation. The y-axis is the same as in Figure 8.5 for easy comparison across species.

8.3.2 Peptide Secondary Structure - LKα14

To determine the effects of ILs on the secondary structure of LKα14, we calculated with DSSP the average degree of α-helicity of the five peptides across the three replicate interfacial LKα14 simulations, both at the interface and in the bulk. As mentioned previously, only the second half of the trajectories were included in the calculations to avoid any artifacts from the initial starting configurations. Additionally, in obtaining the interfacial data, secondary structure was only calculated for peptides bound to the vacuum/solvent interface to avoid contaminating the results with desorbed peptides (see Methods for more detail). This screening was done dynamically; peptides that adsorbed and desorbed multiple times during this period were only included in the calculation during the times at which they were adsorbed. Bulk calculations (without any screening) were similarly performed on the second half of the three non-interfacial simulation trajectories for each of the five peptides.

Table 8.3. LKα14 degree of helicity as calculated by DSSP in IL solvents and water.

	Interface	Bulk
Water	66.0%[a] ± 2.0%[b]	51.9% ± 3.4%
[BMIM][Ace]	58.2% ± 8.1%	55.6% ± 1.9%
[BMIM][Cl]	58.2% ± 6.3%	64.2% ± 2.3%

[a]Averages calculated across the 2nd half of all individual peptide trajectories for the three replicate unbiased interfacial simulations.
[b]Errors represent the standard error of the mean, where the sample size (n=15) is the total number of individual peptide trajectories for the three replicate unbiased interfacial simulations.

The DSSP results, summarized in Table 8.3, show several interesting trends concerning α-helical behavior in the IL solutions: (1) ILs tend to decrease the amount of α-helical character for interfacially bound peptides compared to pure water, and (2) IL solutions tend to increase the amount of α-helical character for bulk-residing peptides compared to pure water. In the case of [BMIM][Cl], especially, the increase in α-helicity in the bulk IL solution compared to water is greater than 12%. One explanation for this increase is that hydrophobic interactions between IL cations and peptide leucine sidechains provide a stabilizing effect on LKα14's secondary structure. The high concentration of IL in our solutions (50% w/w) provides a hydrophobic buffer to the peptides, effectively shielding them from water and thus enabling LKα14 to form and sustain the intramolecular hydrogen bonds that make up the α-helix. This conclusion is supported by comparing the α-helicity of LKα14 in pure bulk water versus at the vacuum/water interface in Table 8.3: the significantly higher degree of α-helicity at the interface (66.0% versus 51.9% in bulk) can be explained by a driving force for LKα14's hydrophobic residues to co-localize away from the hydrophilic solvent, the same mechanism by which they are able to shield themselves from water using the hydrophobic tails of [BMIM] ions.

Figure 8.7 offers an indirect explanation for the nearly 8% decrease in interfacial α-helicity of LKα14 in either IL solution compared to water. Average sidechain tilt angles are calculated in Figure 8.7 for each of the eight leucine residues in LKα14 at the interface in each of the three solvents. While the values appear to be quite similar for the different solvents, the fluctuations in the tilt angles are notably higher in both IL solvents for nearly every leucine residue compared to water. Likely, the increased sidechain fluctuations lead to destabilization of backbone hydrogen bonds in the peptide, resulting in some loss of helicity as seen in the DSSP results of Table 8.3.

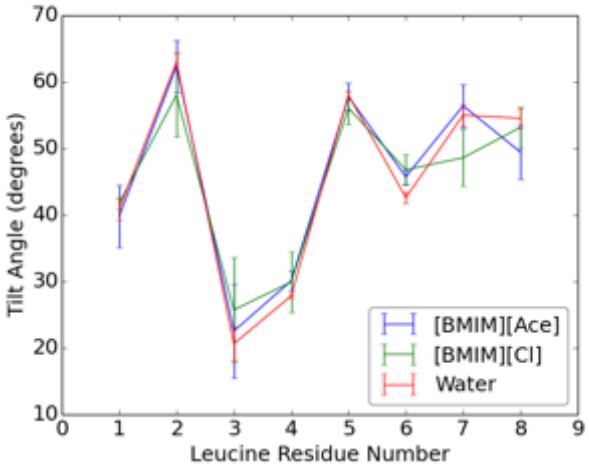

Figure 8.7. Tilt angles with respect to the vacuum/solvent interface of leucine sidechains for interfacially adsorbed LKα14 peptides in the three different solvents. Data represent averages and standard deviations across all peptides and replicate simulations.

8.3.3 Peptide Secondary Structure - LKβ15

A number of interesting patterns emerged when studying the secondary structure of LKβ15 in our simulations. Chief among these was the observation that upon adsorption to the interface,

LKβ15 peptides typically adopted one of two distinct conformations depending on the solvent. Simulation snapshots of these two conformations are shown in Figure 8.8.

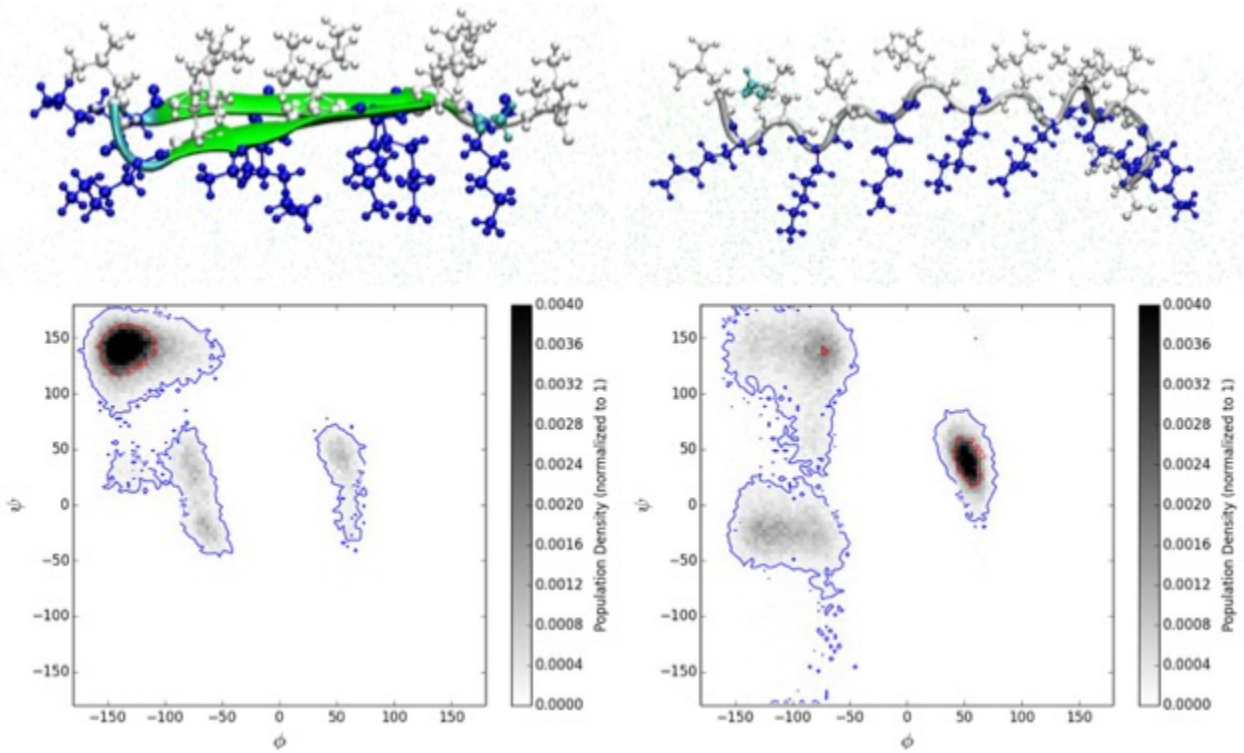

Figure 8.8. Comparison of conformations adopted by LKβ15 at the vacuum/water interface (left) and the interface of vacuum with 50% aqueous [BMIM][Ace] (right), along with corresponding Ramachandran plots. Angle data was calculated for a single peptide over the last 10 ns of a simulation – the representative snapshots were taken from a single trajectory frame from this range. Green, cyan, silver, white, and blue coloring indicate β-sheet character, turns, random coils, leucine, and lysine residues, respectively. Acetyl capping groups are shown in light blue. High-density regions are outlined in red; low density regions in blue.

Ramachandran plots are included to quantitatively highlight the differences in the secondary structures shown in Figure 8.8. We note that Ramachandran plots are used here because of the relative lack of hydrogen bonding in the LKβ15 structures we observed, rendering DSSP much less useful and elucidating than it was for LKα14. Figure 8.8 shows that at the vacuum/water interface, LKβ15 predictably folds into a β-hairpin structure, indicated by the high density of conformational states in the top left corner of the Ramachandran plot. In stark contrast, LKβ15 adopts a left-handed α-helical type structure at the interface of vacuum with both aqueous [BMIM][Ace] and [BMIM][Cl] (data not shown). This is indicated by a densely concentrated area on the Ramachandran plot that is nearly equally as populated as the β-hairpin region in the water Ramachandran plot. While the conformational snapshots of the two structures are highly dissimilar, both structures allow the peptide to partition its hydrophilic and hydrophobic residues to lead to energetically favorable conditions at the interface. Additionally, both structures can occur without interaction with other peptides, as confirmed in independent simulations containing a single LKβ15 peptide and by watching isolated peptides in the multi-peptide simulations.

Apart from this interesting structural shift, LKβ15 did not show other noteworthy structural motifs at the interface, and showed no structural definition in bulk solution, regardless of solvent. These results are summarized in Table 8.4, where DSSP was again used to calculate the degree of secondary structure of the peptide. The results were generated using data from the last half of three 200 ns simulations and averaged across each individual peptide in the simulations. Due to the larger structural variation of LKβ15 over LKα14, DSSP was used to calculate both the degree of β-sheet character as well as α-helicity of the peptide in the different solvents. In line with the Ramachandran plots, the DSSP analysis highlights the large reduction in β-sheet character of the peptide at the interface in both IL solvents compared to water. While Figure 8.8 indicates this is due to a population shift to more helical-like structures at the interface, these results are not reflected in Table 8.4 because the structure we observed does not contain any intramolecular hydrogen bonds, which DSSP uses to characterize secondary structure. Contrary to the results for LKα14, the results of Table 8.4 also show that the IL solutions slightly decrease the amount of secondary structure of the peptide in bulk compared to water (both for α-helicity and β-sheet character), though we note these losses in bulk IL solution are much smaller, indeed almost negligible, than those incurred upon interfacial adsorption in the IL solutions.

Table 8.4. Degree of secondary structure formation of LKβ15 in IL solvents and water, as calculated by DSSP. The novel helix structure described in the text is not shown, since it does not contain any hydrogen bonding and therefore cannot be detected by DSSP.

	α-helix (interface)	α-helix (bulk)	β-sheet (interface)	β-sheet (bulk)
Water	1.03%[a] ± 1.38%[b]	2.40% ± 2.29%	6.67% ± 2.49%	2.39% ± 0.82%
[BMIM][Ace]	0.00% ± 0.00%	0.00% ± 0.00%	0.31% ± 0.22%	0.50% ± 0.26%
[BMIM][Cl]	0.60% ± 1.22%	1.87% ± 2.65%	0.77% ± 0.99%	1.35% ± 0.84%

[a]Averages calculated across the 2nd half of all individual peptide trajectories for the three replicate unbiased interfacial simulations.
[b]Errors represent the standard error of the mean.

Figure 8.9 provides an overview of all LKβ15 backbone angles from our simulations in the form of additional Ramachandran plots, where the data has been averaged over all peptides in the last 10 ns of a single simulation for each plot. The results show both the β-sheet formation we observed in water (top left) and the helix-like structure we observed in ILs (top right). The results also show the solvents we tested did not have an appreciable impact on LKβ15's bulk structure, in line with the results reported in Table 8.4. The concentration of angles around ψ=150° and ϕ=75° corresponds to a very loosely structured, stretched out conformation that was easily shifted to and from other conformations in our simulations.

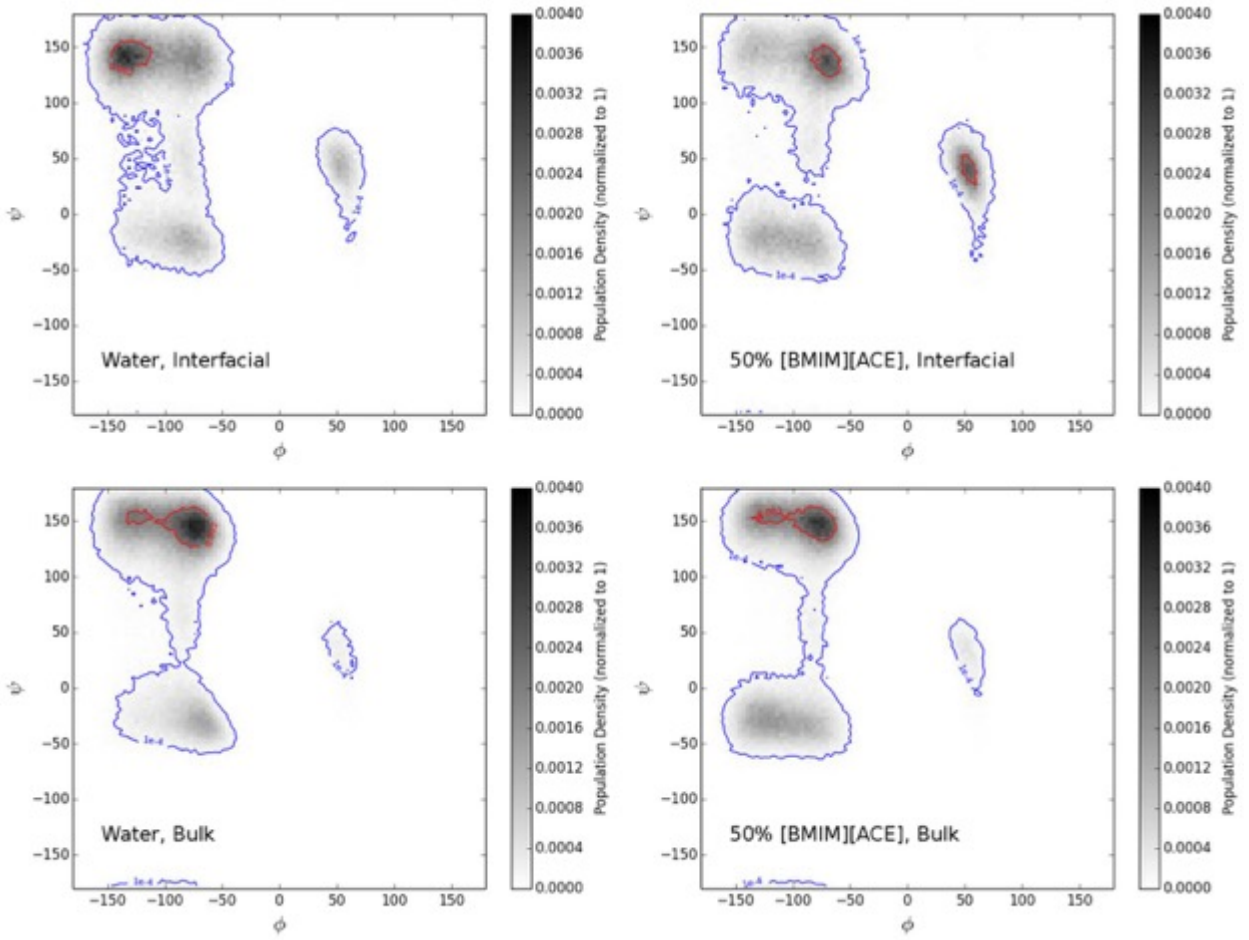

Figure 8.9. Comparison of secondary structures formed by LKβ15 both at the interface and in bulk in pure water and 50% [BMIM][Ace]. High-density regions are outlined in red; low density regions in blue. For each plot, the data was obtained by averaging angle data for all peptides over the last 10 ns of a single simulation.

As with LKα14, we also calculated the average leucine sidechain tilt angles for interfacially adsorbed LKβ15 peptides. In contrast to the results for LKα14 that showed the tilt angles were nearly identical in the different solvents, Figure 8.10 shows that for LKβ15 there is a significant difference in the tilt angles at the interface with pure water compared to the interface with the IL solutions, which show similar results between the two solvents. This difference reflects the added weight to the average tilt angles of structures with less β-sheet character and more helical character at the interface, as discussed previously in regards to Figures 8.8 and 8.9.

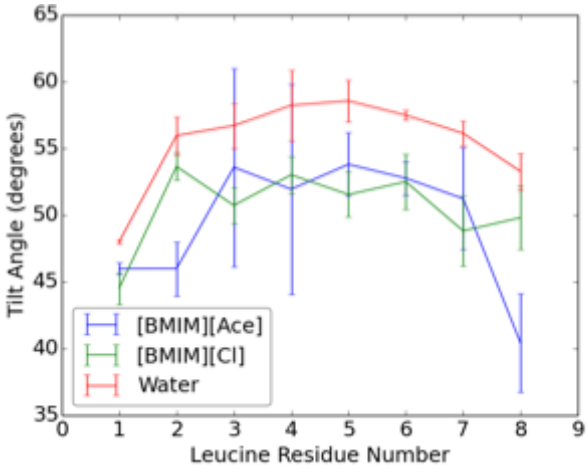

Figure 8.10. Tilt angles with respect to the vacuum/solvent interface of leucine sidechains for interfacially adsorbed LKβ15 peptides in the three different solvents. Data represent averages and standard deviations across all peptides and replicate simulations.

Conclusions

Using a combination of classical molecular dynamics and metadynamics simulations, we examined the effects of two aqueous IL solutions and water on two model peptides, LKα14 and LKβ15, both at an interface and in bulk solution. In particular, we evaluated the influence of IL on the diffusion behavior of the peptides and affinity for the vacuum/solvent interface, as well as the effects of IL on both the bulk and interfacial conformations of the peptides compared to in water.

In summary of our analysis of the diffusion and interfacial affinity characteristics of the peptides, it was determined that: (1) IL solutions slow the diffusion of peptides to the interface due to the increased solution viscosity, thus encouraging the bulk aggregation tendencies of the peptides, and (2) IL solutions reduce peptide affinity for the interface due to favorable interactions in the bulk between the hydrophobic tails of IL cations and hydrophobic peptide sidechains. From our analysis of IL effects on peptide secondary structure, it was determined that: (1) in the case of LKα14, the presence of IL led to increased sidechain fluctuations at the interface that disrupted helix formation and caused a reduction in the overall degree of secondary structure calculated by DSSP, and (2) in the case of LKβ15, the presence of IL both increased sidechain fluctuations to disrupt the formation of β-sheets or hairpins at the interface as determined by DSSP, and caused the peptides to sample a wider variety of conformations at the interface, leading to significant differences in the interfacial leucine tilt angles of the peptides compared to in water.

These results reveal several interesting effects of non-native solvent environments on the fundamental behavior of small peptides at interfaces, which is crucial for understanding the behavior of larger biomolecules in even more complex systems. There are also many novel applications for the self-assembly of peptides themselves at interfaces, for instance in the synthesis of new nanostructures with optical, energy, and healthcare-related applications, as discussed in a recent review by Kim et al[355]. This work highlights the effect of specific peptide/solvent interactions on peptide aggregation tendencies, or lack thereof in the case of LKβ15, which could

have broader applications for promoting or inhibiting this type of behavior in biomedical design fabrications, e.g. in preventing aggregation of biomolecules on implantable devices, or in preventing the role of β-sheet aggregation in diseases such as Alzheimer's or Parkinson's[356-358].

Additionally, there are numerous benefits to studying such types of model systems as LK peptides and simple interfaces like the air/water or vacuum/solvent interface. For example, the study of peptides adsorbed to interfaces offers unique opportunities to study aspects of peptide/protein behavior in confined, two-dimensional environments, such as in the 2D plane of a lipid bilayer where transitions to β-sheet-like structures can occur for some unstructured peptides[359]. Additionally, simulating model systems generally allows for easier direct comparison with experimental data and a more straightforward means to validate computational models.

Chapter 9

Determining Dominant Driving Forces Affecting Controlled Protein Release from Polymeric Nanoparticles[9]

Introduction

In recent years, proteins have played an increasingly important role as catalysts in the fields of biology and medicine, wherein protein delivery is of interest for therapeutic intervention in many diseases.[360] However, avoiding stability and activity loss of the protein in biological environments remains an active area of research. Several platforms have emerged around the concept of protein immobilization within a material to stabilize the proteins while still allowing for substrate diffusion to the protein through the pores of the material. Polymer nanoparticles (NPs) are particularly promising delivery vehicles for encapsulated proteins due to their high protein-loading capacity and porous nanostructure.[361-362] These properties are especially important for medical applications because delivering high drug payloads to sites of interest by polymer nanocarriers can reduce toxicity and off-site side effects. Additionally, therapeutic effects can only be achieved if the proteins are delivered in a stable and active form. Polymer NPs can promote stability through physical confinement of the protein and enable increased activity by serving as a pH buffer that provides proteins with an ideal chemical environment.[363-364]

Recent applications of protein-encapsulated polymer nanostructures include proteins in polymer nanoshells[365] or in polymer microspheres[366] for controlled drug release, leading to increases in protein activity and stability in both cases. In the latter study, protein stability was further increased with the addition of stabilizers like polyethylene glycol (PEG) to the polymer microspheres. Many variable protein/polymer frameworks exist due to the numerous factors that can influence the effect of encapsulation on the protein's stability and activity, which include: the nature of the protein/carrier interaction, the protein's ability to undergo conformational change within the carrier, the protein's surrounding microenvironment, the chemical and physical properties of the carrier, the properties of the protein/polymer linker, and the immobilization conditions of the experiments. The underlying mechanisms governing many of these relationships

[9]Reproduced in part with permission from K.G. Sprenger, Josh Smith, Rick Liao, Andrea Joseph, Elizabeth Nance, and J. Pfaendtner. Determining Dominant Driving Forces Affecting Controlled Protein Release from Polymeric Nanoparticles. Biointerphases, 12: 02D412, 2017. Copyright 2017 AVS.

are still unknown, and thus optimization of the performance of these protein/polymer structures is carried out through an exhaustive and experimental trial-and-error process.

Physics-based computational methods could provide needed fundamental insight into the interactions of proteins and polymers. Additionally, simulations could be used as a preliminary tool to parse the vast parameter design space of polymer nanostructures to determine optimal protein/polymer test candidates for given applications, thus providing a powerful guide for experiments. Despite this immense potential, the use of molecular simulations to study drug/polymer interactions has been limited by the computational expense of simulating these large-scale systems (i.e. proteins and lengthy polymer chains together),[367-368] which is exacerbated by the long relaxation times of these systems. To this end, computational studies to date have taken the approach of simulating the interactions of single or multiple drug molecules with a finite and relatively limited number of solvated polymer chains[369-370]. Another approach includes the use of coarse-grained models like dissipative particle dynamics to study larger polymer-based systems[371-373]. However, this approach sacrifices atomistic detail that is necessary to fully understand protein/polymer interactions and to be able to predict or fine-tune these interactions for new applications. Additionally, these methods are limited in their ability to quantitatively compare against experimental observations such as drug release kinetics. The need persists for a fundamental understanding of the major determinants of protein release and retention from synthetic biomaterials.

Herein, we use a combined experimental and computational approach to probe the relative contributions of various physical phenomena in determining protein release from polymeric nanoparticles. Experimental techniques – better suited to explore macroscopic phenomena such as transport effects and polymer matrix properties – are used to measure release profiles for model protein bovine serum albumin (BSA) from NPs that span a range of chemistries and sizes. These include poly(lactic-*co*-glycolic acid) (PLGA) copolymer, poly(styrene)-poly(lactic acid) (PS-PLA) copolymer, and poly(lactic acid) (PLA), from ~85 to 160 nm in diameter in NP form. Results rule out transport effects as a determining factor in release time, suggesting the need for computational study of nanoscale physical interactions. We have used classical, fully atomistic molecular dynamics (MD) simulations to investigate the relative importance of polymer-protein and polymer-polymer interactions as the basis for a predictive framework to design protein encapsulation systems for medical applications. Studies are performed on BSA exposed to aqueous solutions of trimeric PLGA, PLA, and PS-PLA. Additional simulations were performed with trimers of polystyrene (PS) as a control. Analysis of amino acid preferences and spatial distribution of polymer binding on the BSA surface suggest insufficient differentiation between PLA and PLGA to explain their different BSA release profiles. The aggregation tendency of PS in simulations with PS-PLA or PS alone supports the hypothesis from experiment that PS-PLA core-shell arrangement is responsible for similar release from PLA and PS-PLA NPs. Taken together, the combined study suggests that polymer matrix properties that tend toward self-interaction may outweigh polymer-protein interactions for controlled protein release. Simulations are extended to study trimers of poly(styrene)-acrylate (PSAC) copolymer, which may be tuned for the relative strength of self-interactions and protein interactions, providing a handle for future experimental optimization.

Experimental Methods

9.2.1 Materials

BSA conjugated with fluoro-isothiocyanate (BSA-FITC, or BSA) was purchased from Sigma Aldrich. PLGA copolymer (50%/50% molar ratio), PS-PLA copolymer (27%/75% molar ratio), and PLA were purchased from PolySciTech Akina, Inc. Poly(vinyl alcohol) (PVA) and bile salts (cholic acid, CHA), were obtained from Sigma Aldrich. Dulbecco's phosphate buffered saline (DPBS, or PBS) was obtained from Corning. Dichloromethane (DCM) was obtained from Fisher Chemical.

9.2.2 Nanoparticle Formulation

For polymer NP formulations, $w_1/o/w_2$ (water/oil/water) double emulsions were used, adapted from previously published protocols.[374] 50 mg polymer was dissolved in 1 mL DCM and added to 0.2 mL deionized (DI) water containing 5 mg dissolved BSA and 0.5 wt.% PVA. The solution was immediately sonicated, with 1s:1s on/off for 2 min. The first emulsion was then added to 4 mL water with 2% PVA, and immediately sonicated, following 1s:1s on/off sonication for 2 min. For PLA and PLGA NP formulations, the first emulsion was performed at 30% amplitude, and the second emulsion at 20% amplitude using an ultrasonic processor (QSonics). For PS-PLA NP formulations, the first and second emulsion amplitudes used were 40% and 30%, respectively. Immediately after the second emulsion, the sample was added to 25 mL of water containing 1% PVA and either 1% CHA for PLA and PS-PLA formulations, or 1x PBS for PLGA formulations. After 3 h of stirred incubation to evaporate off DCM, particles were collected via ultracentrifugation at 2000 relative centrifugal units (RCF) for 2 min to remove any large aggregates or undissolved polymer. Particles were then collected at 100k RCF for 60 min. The particles were then washed once with DI water in a 25-min centrifugation at 100k RCF prior to resuspension in 1 mL 1x PBS. For PS-PLA particles, the wash step was eliminated to avoid potential aggregation.

9.2.3 Nanoparticle Characterization

NP hydrodynamic diameter (size), polydispersity index (PDI), and zeta potential (ZP) were measured using dynamic light scattering (DLS, Malvern Zetasizer). Encapsulation efficiency (EE) was quantified using ultraviolet-visible light spectrometry (UV-Vis) compared to a standard calibration curve for BSA-FITC dissolved in 1xPBS. EE is defined as follows:

$$\%\ EE = \frac{weight\ of\ drug\ encapsulated\ in\ NP}{total\ weight\ of\ drug\ added\ into\ formulation} \times 100\% \tag{9.1}$$

9.2.4 BSA Release

BSA-FITC release studies were performed for all NP formulations by reconstituting and incubating NPs in 10 mL 1x PBS on a stir plate at 37°C. At specified time points, NPs were centrifuged at 100k RCF for 25 min, and the supernatant was collected. The amount of BSA in the supernatant at the time points of 1 h, 2 h, 4 h, 8 h, and 1, 2, 3, 4, 5, 6, and 7 days was analyzed using UV-Vis, as described above. % BSA release was calculated by dividing the concentration of BSA in the supernatant by the total initial encapsulated amount.

Computational Methods

Representative structures of all polymer surrogates and BSA are shown in Figure 9.1, with the addition of PS as a control. Additional simulations of other protein/polymer surrogate systems are described in the Results and Discussion. BSA was solvated in 2 wt.% of the polymer surrogates in water. Solutions with PLGA were simulated as trimers of alternating lactic acid (LA) and glycolic acid (GA) monomers (6 units total per molecule) to mimic experimental conditions of a PLGA random copolymer with an approximately 50%/50% molar ratio of LA and GA. Solutions with PS-PLA were simulated as individual trimers of PS and PLA in an approximately 25%/75% molar ratio (3 monomers per molecule of either PS or PLA; 2 wt.% total surrogate) to match experimental conditions of the diblock copolymer. Packmol[178] was used to generate cubic boxes of the protein/IL systems, resulting in approximately 120,000 particles per system. A single trial was conducted for each system. The low polymer wt.% in our solutions was motivated by the hypothesis that protein surface binding properties can be more accurately studied at low solute concentrations due to minimization of random surface collisions.

Polymer surrogates were parameterized with the general Amber force field (GAFF),[182] with atomic charges obtained via Gaussian[175] and the RESP[217] partial charge calculation method. Trimer charges were adjusted to more closely match monomer charge distributions (section IV-F). The Amber14SB[375] and TIP3P[181] force fields were used to model the protein and water, respectively, and topology for the polymer surrogates was built to mimic the intracellular pH of the brain that is maintained at approximately 7.2.[376] We are particularly interested in better understanding protein/polymer interactions in therapeutic NPs for delivery to the brain due to the growing body of literature demonstrating the impact polymer NP properties have on determining site specific delivery for brain diseases.[377-380] Simulations were carried out with the GROMACS 5.1.2[104] MD engine, using a time step of 2 fs and full periodic boundary conditions. For all MD simulations, the LINCS algorithm[110] was employed to constrain bonds between hydrogen and heavy atoms, and Particle mesh Ewald summations (PME[53]) were used to calculate long-range electrostatic interactions between the particles with a cutoff value of 10 Å. van der Waals forces were also calculated with a cutoff value of 10 Å and shifted to zero, and neighbor lists were updated every 10 steps with a cutoff of 10 Å.

Following a steepest descent energy minimization of the initial configurations of each system for 40,000 steps, a short 200 ps simulation in the NPT ensemble was run to equilibrate the systems at 298.15 K and 1 bar, with the Bussi-Donadio-Parrinello thermostat[54] and Berendsen barostat.[109] It is well known that the Berendsen barostat is useful for quickly equilibrating the box

size to the correct average pressure, but the fluctuations from the Berendsen barostat do not correctly resemble the NPT ensemble. Therefore, as is custom in the field, we used the Parrinello-Rahman barostat[222] for the production runs, with the same thermostat. Simulations were run for 200 ns, unless otherwise indicated. In solutions containing the highly hydrophobic PS surrogate (i.e., with either PS or PS-PLA), additional 50 ns simulations were run with the PLUMED 2.0 plugin[52] to prevent PS aggregation (see Results and Discussion, section A). Enforcing a maximum number of contacts allowed between each PS atom and all other PS atoms in the system prevented aggregation. The maximum number of contacts was set equal to the original number observed in the fully dispersed systems obtained from Packmol. Contacts were defined via a simple switching function that ranged from 0 to 1 with a reference bond length of 3.6 Å.

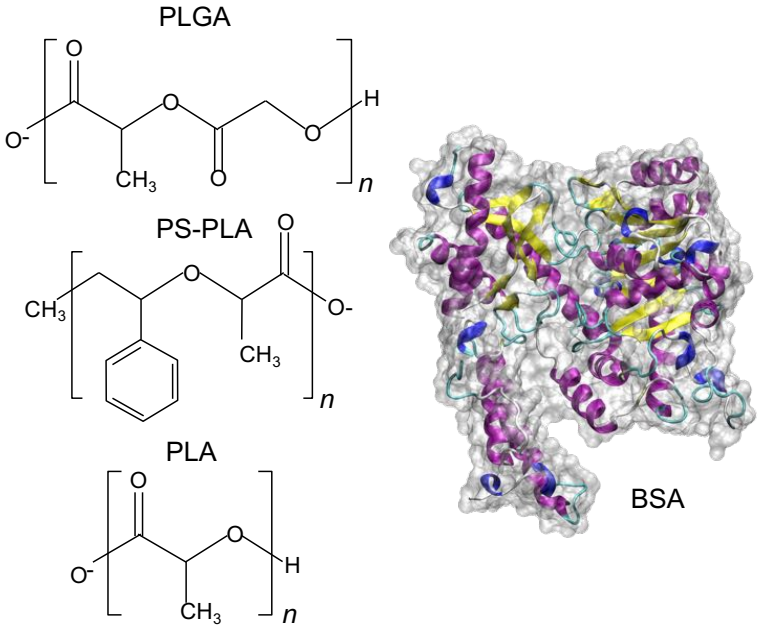

Figure 9.1. Protein and surrogates used to model polymer nanoparticles in MD simulations. Surrogates include trimers of PLGA, PS-PLA, and PLA. Purple, blue, yellow, cyan, and white coloring indicate an α-helix, α_{310}-helix, β-sheet, turn, and random coil, respectively, and the protein (BSA) is also shown in surface representation in gray.

Results and Discussion

9.4.1 Experimental Determination of Protein Release Rates from Polymer Nanoparticles

Using a single emulsion method, we formulated monodisperse PLGA NPs of size 85.3 nm, PLA NPs of size 102.5 nm, and PS-PLA NPs of size 160.3 nm, all with PDIs below 0.2 (Table 9.1).

Table 9.1. Nanoparticle characterization before and after BSA release experiments (n=3 batches for all measurements).

Polymer Type (MW in Da)	Mean Diameter ± SEM (nm)	PDI ± SEM	Mean ZP ± SEM (mV)
PLGA (10-15k)	85.3 ± 7.6	0.15 ± 0.01	-67.0 ± 2.2
PLA (10-15k)	102.5 ± 4.9	0.15 ± 0.01	-62.1 ± 11.3
PS-PLA (5k:5k)	160.3 ± 21	0.11 ± 0.01	-38.2 ± 1.6

Due to the inability of accurately measuring BSA content in PS-PLA NPs because of absorbance/fluorescence interference from styrene π stacking,[381] we estimated that PS-PLA NPs contained the same initial BSA content as PLA NPs of 0.823 mg. We expect this approximation to be along the right magnitude or potentially higher than the actual value, considering the hydrophobic styrene groups would discourage greater BSA encapsulation. PLGA had an initial BSA content of 0.536 mg. We calculated the total amount of BSA released for PLA (0.286 mg BSA) to equal the difference between BSA content of PLA NPs before (0.823 mg BSA, 16.39% EE) and after release (0.537 mg BSA, 10.70% EE), and scaled the raw fluorescence of supernatant BSA values to this maximum release. We then scaled the raw fluorescence to corresponding BSA release concentrations for PLGA and PS-PLA. After a 7 day incubation period, PLGA, PLA, and PS-PLA achieved 8.12%, 34.74%, and 42.66% BSA release, respectively (Figure 9.2).

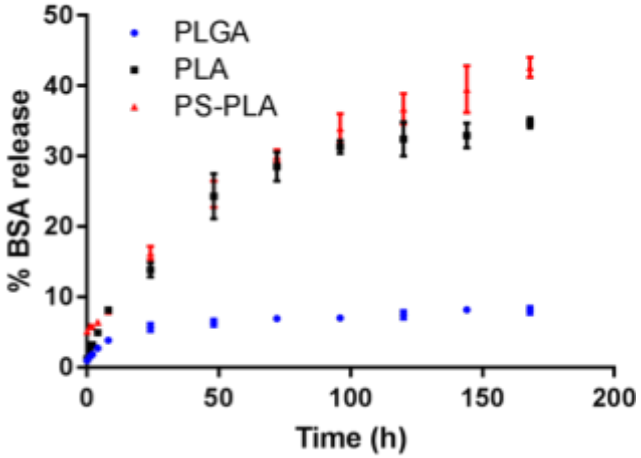

Figure 9.2. Percent BSA release profile of BSA released against total BSA initially encapsulated for PLA (black), PS-PLA (red), and PLGA (blue) nanoparticles (n=3 batches for all samples).

PS-PLA had the greatest % BSA release and PLGA the least % BSA release after 7 days of incubation. Release is driven by water adsorption into and interaction with the polymer matrix that drives polymer dissolution.[382] Interestingly, PLA and PS-PLA exhibit a nearly identical release profile up to 3 days, then a greater rate of BSA release is seen in PS-PLA from 3 to 7 days. Experimentally, it is possible PS-PLA forms a PS core with a PLA outer layer, whereas PLA NPs are uniform in composition throughout the particle matrix. The initial pattern of release (up to 3 days) seen in PLA and PS-PLA suggests that BSA is evenly distributed among the PLA components of each formulation. However, BSA is not strongly associated with PS in the core of

the PS-PLA NP, lending towards faster release from PS-PLA at later time points, as water adsorbs into the polymer core. These results would be further supported with future analysis of spatial distribution of BSA within the polymer NP matrix, and with precise determination of BSA content in PS-PLA NPs.

It is also important to note that transport-limited release is refuted given that there is no obvious correlation between nanoparticle size and % release. In fact, the qualitative correlation here is opposite of that expected in the context of diffusion limitations – the smallest particle has the slowest release. These observations suggest that computational modeling can provide useful insight into the role of nanoscale physical interactions in explaining differences in empirical protein release from chemically distinct NPs.

9.4.2 Aggregation Tendencies of Polymer Surrogates

As demonstrated in Figure 9.3A, aggregation was observed in our MD simulations to occur relatively quickly between PS molecules in aqueous solution with either PS-PLA or PS alone, clearly driven by strong hydrophobic and aromatic interactions. Aggregation occurred more slowly in the solution with PS-PLA, where the presence of free PLA trimers could act to partially shield hydrophobic interactions between nearby PS molecules. After 50 ns, we concluded protein/PS interactions were no longer changing in the simulations due to significant PS aggregation. Based on our observations, PS self-interactions can out-compete interactions between the protein and PS surrogate molecules. We hypothesized this would manifest in faster release of BSA from PS-PLA particles than from PLGA or PLA NPs, that per our simulations should not suffer from polymer self-interactions and instead should have stronger overall binding with the encapsulated protein. This was verified by our experiments that, of the three NP formulations, PS-PLA showed the greatest % BSA release after 7 days of incubation (Figure 9.2).

The high level of aggregation in our PS-based simulations prevented a detailed study of the molecular scale interactions between hydrophobic polymer substituents and protein surfaces. It is crucial to study these interactions because not all polymers with hydrophobic moieties will aggregate, as we demonstrate in a later section, which could provide new insight into protein release kinetics from polymer NPs. Furthermore, the possibility of preparing enzyme-nanoparticle formulations closer to the theta point of the polymer suggests the need to more carefully probe the specific protein/polymer interactions. To probe these interactions, we performed additional simulations of BSA in aqueous solutions with PS-PLA and PS alone with restraints in place to prevent aggregation between PS molecules (see Computational Methods for details). As shown in Figure 9.3B, even after 200 ns of simulation, the PS remained completely dispersed in solution, leading to increased interactions between the protein and polymer surrogates, discussed in detail in the following section.

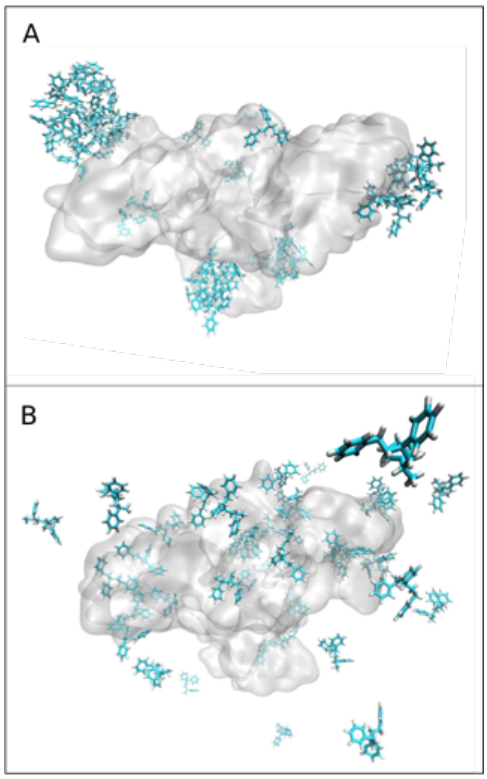

Figure 9.3. Final snapshots from simulation trajectories of BSA in aqueous solutions with 2 wt.% PS, without (A) and with (B) restraints between PS atoms in neighboring molecules to prevent self-aggregation. BSA is shown in surface representation in transparent gray coloring to show PS binding to all sides of the surface. PS molecules are shown with carbon and hydrogen atoms colored cyan and white, respectively. Water and neutralizing counterions are not pictured.

9.4.3 Amino Acid Binding Preferences of Polymer Surrogates

To assess the amino acid binding preferences of the polymer surrogates, the fraction of contacts over the course of the MD trajectories was calculated between each polymer surrogate and the different types of amino acid residues on the surface of BSA. Fractional contacts were calculated by dividing the number of frames in which surface amino acid residues were in contact with a polymer surrogate by the total number of frames in the trajectory, and then summing over all residues of each type of amino acid. Contacts were defined between polymer surrogate molecules and protein surface residues within 4 Å of each other. The first 100 ns of each simulation trajectory was excluded to ensure the calculations were only performed over converged portions of the data. Figure A5.1 demonstrates that with few exceptions, the fraction of contacts with each amino acid type converged to within a few percent by approximately 100 ns with each polymer surrogate; extension of the simulation of BSA in the solution with PLGA to 500 ns further demonstrates that 100-200 ns is an appropriate timeframe over which to calculate fractional contacts of small polymer surrogates to protein surface residues (Figure A5.2). We also note that in quantifying binding trends, we did not consider BSA conformational changes in our calculations; the root mean squared deviation (RMSD) of the protein's Cα atoms from the crystal structure over

the course of the simulations shows the structural integrity of BSA remained intact in all solutions, with similar fluctuations in the polymer surrogate solutions as in pure water (Figure A5.3).

The results for the fraction of surface amino acid types contacted by the polymer surrogates are shown in Figure 9.4. The data have been normalized by the protein surface fraction of each amino acid residue type contacted by each polymer surrogate over the portion of the simulation trajectories from 100-200 ns, and error bars reflect the average standard deviation of calculations performed over four 25 ns portions in this timeframe. The results illustrate the specificity of protein/surrogate binding, in that the unique functional groups on the polymer surrogates lead to different amino acid binding preferences. The slightly more hydrophilic nature of PLGA is reflected in a small decrease in binding to hydrophobic residues and increase in binding to negatively-charged residues compared to PLA; however, binding to most amino acids types is statistically indistinguishable for PLGA and PLA. Considering the earlier results of Figure 9.2 that showed PLGA NPs had a significantly lower % BSA release compared to PLA NPs, we draw the conclusion that molecular scale interactions between the protein and surrounding host environment are of less importance than materials-level properties (e.g., polymer flexibility and hydration properties) in determining and controlling protein release kinetics from polymer NPs.

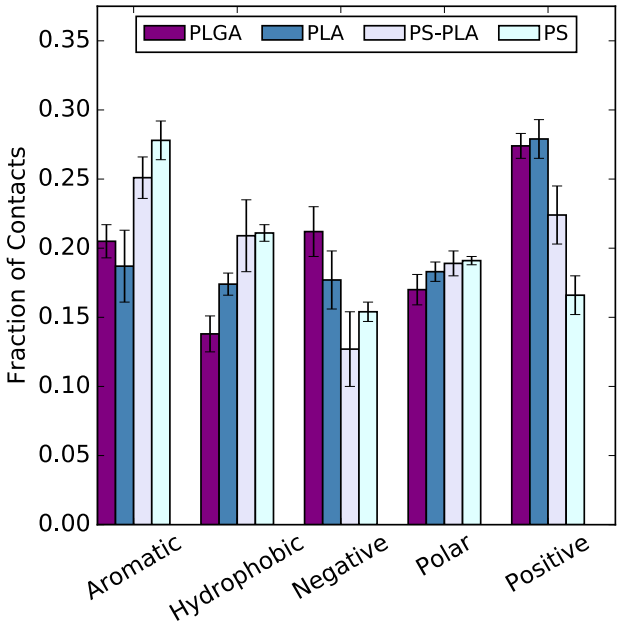

Figure 9.4. Fraction of contacts to the different amino acid types on the surface of BSA by polymer surrogates. Error bars represent the average standard deviation of calculations performed on consecutive 25 ns portions over the second half of each 200 ns simulation trajectory (i.e., n=4).

Comparing the results in Figure 9.4 for the PS-PLA solution versus the solutions with PLGA or PLA shows very different binding behavior for PS-PLA. Whereas PLGA and PLA are seen to bind more frequently to surface residues through strong electrostatic interactions (i.e., to both negatively and positively charged residues), PS-PLA binds more frequently to hydrophobic and aromatic residues. These trends are intuitive due to the predominantly hydrophilic nature of PLGA and PLA, and the partially hydrophobic nature of benzene-containing PS-PLA. Comparing instead the results for PS-PLA with either PLA or PS alone shows the fraction of BSA surface

contacts targeted by PS-PLA falls in between that of PLA and PS for nearly every amino acid type. This finding implies that binding can be modulated by altering the relative surrogate compositions, supporting the notion that changes in the NP functionality will lead to differential protein binding and controlled protein release rates from NPs formulated using these polymers. Future experimental studies are planned to explore the validity of this hypothesis. Figure 9.4 shows that the fractional occupancy values for PS-PLA tend to be slightly closer to the values for PS over PLA, despite the approximately 25%/75% molar ratio of PS to PLA in the solution. Therefore, we infer that the nature of the relationship between the composition of the NP substituents and the resultant characteristics of protein/polymer binding is non-linear and complex, and may follow a "winner takes all" approach, as we investigate in the following section. Additionally, while the results of Figure 9.4 tell us the frequency of protein/polymer surrogate binding, they do not tell us the relative strength of binding of the different polymer surrogates to protein surface residues, which will also be addressed in the following section.

9.4.4 Spatiotemporal Resolution of Polymer Surrogate Binding to Protein Surface Residues

It is necessary to further explore the spatiotemporal binding characteristics of each polymer on the BSA surface to confirm or rule out the central importance of polymer-protein interactions for controlled release. Differences in protein surface coverage and/or polymer residence times could provide the necessary explanation for differences in experimentally observed release times. For each MD trajectory, the fractional occupancy of each BSA residue was calculated, followed by the generation of occupancy heat maps to illustrate the spatial distribution of binding. Occupancy heat maps were created in Python 2.7[383] from the preprocessed MD trajectory data, and the fractional occupancy was calculated for each residue by dividing the number of frames contacting a polymer substituent by the total number of frames (50 frames per 1 ns). Average polymer residence times for each residue were calculated with an in-house Python script analyzing the same preprocessed data. We also tracked highly occupied residues and trimers with high residence time in VMD[47] to resolve inconsistencies between isolated PS/PLA molecules and mixed binding in PS-PLA. Figure 9.5 includes heat maps for each independent polymer simulation as well as a component-wise breakdown of occupancy in the PS-PLA simulation.

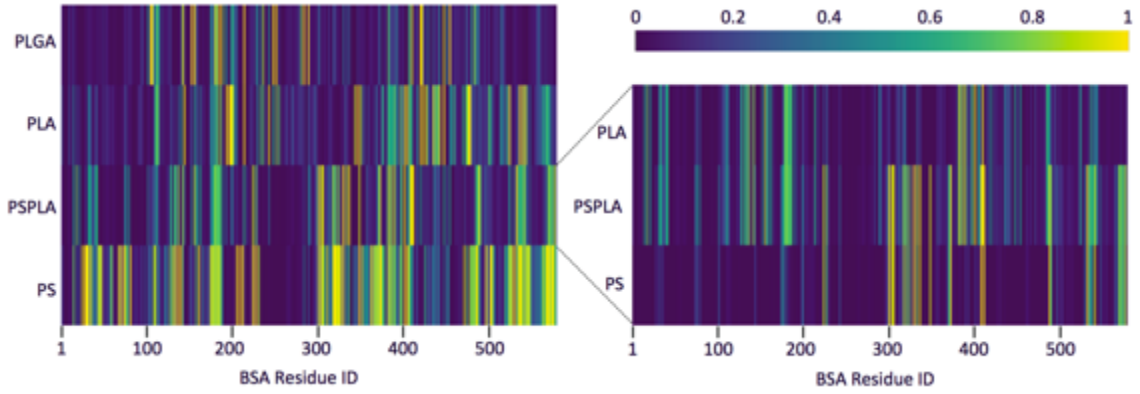

Figure 9.5. Occupancy heat maps for BSA residues: (left) BSA residue occupancy for PLGA, PLA, PS-PLA, and PS simulations. PS simulations include a coordination potential to discourage aggregation. Occupancy was calculated as the fraction of surrogate contacts for each residue over the last 100 ns; (right) PLA, PS-PLA, and PS occupancy from the PS-PLA simulation.

Although PLGA and PLA have similar amino acid type binding preferences (Figure 9.4), the occupancy heat maps suggest the spatial distribution of their binding is different (Figure 9.5, top left). The high-occupancy band from H285 to E291 is unique to PLGA, whereas the thick band from V421 to Y449 is present only for PLA. Overall, the surface coverage for PLA and PLGA is similar, especially compared to the prolific binding of PS in the coordination-constrained simulation. Table 9.2 includes average occupancy per BSA residue for each simulation. The average occupancy is similarly nondiscriminatory for PLGA and PLA, predicting more surface binding for PLA than PLGA. The slight differences in physical interactions do not appear to capture the differences in experimental binding.

Table 9.2. Trimer mean occupancy per BSA residue.

Polymer Type	Mean Occupancy ± Std Dev
PLGA	0.16 ± 0.021
PLA	0.23 ± 0.005
PS-PLA	0.23 ± 0.016
PS	0.38 ± 0.031

The occupancy results above also provide useful information regarding the experimental hypothesis that PS-PLA forms a core-shell particle. With the added coordination potential to prevent aggregation, PS binds persistently to the majority of BSA surface residues. Without the coordination potential, PS aggregates and interacts sparingly with the protein (Figure A5.4). Thus, while PS has a strong preference for the protein over water, it has a much higher affinity for self-interaction. The PS-PLA simulation with PS interaction constraints (represented by the right side of Figure 9.5) exhibits a "winner take all" behavior, where only the polymer substituent with the greatest affinity for a given residue binds to that residue. The PS-PLA simulation without the coordination constraint on PS is more representative of the experimental hypothesis, where PS aggregates and the binding characteristics are dominated by PLA. It may be necessary to use a random copolymer instead of a block copolymer such as PS-PLA to tune aggregation and protein/polymer surrogate binding. The following section includes an extension of this computational approach to such a polymer.

9.4.5 Tuning Polymer Self-Interactions

Building on our finding that the self-aggregation tendencies of the polymer surrogates may play an important role in determining protein release kinetics, we performed MD simulations of BSA in solution with a polymer surrogate that we hypothesize would allow for more control over the protein encapsulation process. Figure 9.6 shows the structure of this new polymer surrogate, poly(styrene)-acrylate (PSAC), which we simulated as trimers of alternating PS and acrylate monomers (6 units total per molecule) to mimic experimental conditions of a PSAC random copolymer with an approximately 50%/50% molar ratio of PS and acrylate. Like PS-PLA, PSAC has strong hydrophobic moieties, though in contrast to PS-PLA, PSAC has no tendency to self-aggregate in solution at the intercellular pH of the brain (~7.2), as determined in our MD simulations. Figure 9.6 shows that even after the full 200 ns of simulation time, the PSAC

molecules are fully distributed around the protein surface, indicating protein/PSAC interactions can outcompete PSAC-PSAC interactions. The difference in aggregation tendencies of the polymer surrogates observed in our simulations is likely due to the different ways in which the monomers are joined in each polymer surrogate chain (i.e., as separate trimers in PS-PLA and as alternating monomers in PSAC). In PSAC, PS-PS interactions were largely inhibited by the unavoidable presence of the neighboring, hydrophilic acrylate monomers.

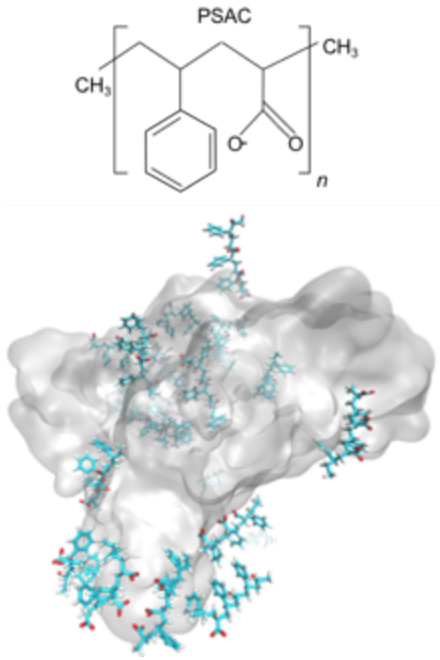

Figure 9.6. Top: Representative structure of poly(styrene)-acrylate (PSAC) copolymer. Bottom: Snapshot from the end of a 200 ns simulation trajectory of BSA in an aqueous solution with 2 wt.% PSAC, with no restraints to discourage PSAC self-aggregation.

Figure 9.7 shows the fraction of amino acid types on the surface of BSA contacted by PSAC compared to PS-PLA and PLGA, where the data and error bars have been calculated and treated as described earlier. Like with previously studied systems, Figure A5.5 shows the fraction of contacts with each amino acid type converges in approximately 100 ns for the simulation of BSA in the solution with PSAC. Additionally, Figure A5.6 shows the structure of BSA remains stable throughout the 200 ns simulation in the solution with PSAC. The results of Figure 9.7 show the binding characteristics of PSAC are similar to PS-PLA when compared to PLGA, in that PSAC binds more often to hydrophobic and aromatic residues and less often to positive and negatively charged residues than does PLGA. However, the magnitude of the changes in binding to the different amino acids types between PSAC and PLGA is generally smaller than between PS-PLA and PLGA, which arise from the separately simulated trimers of PS and PLA in PS-PLA that results in more definitive binding to specific amino acid types. These findings highlight the sensitivity of our approach to both the individual chemical functionalities in a given polymer, as well as the overall polymer chemistry that is driven by the spatial proximity of individual functional groups in the polymer chain. Based on our observations, we would expect slower protein release rates from PSAC NPs compared to PS-PLA NPs due to decreased polymer self-

interactions leading to increased protein/polymer interactions. However, we still anticipate PSAC NPs would lead to faster release compared to PLGA or PLA NPs due to weaker overall binding with the largely hydrophilic BSA protein.

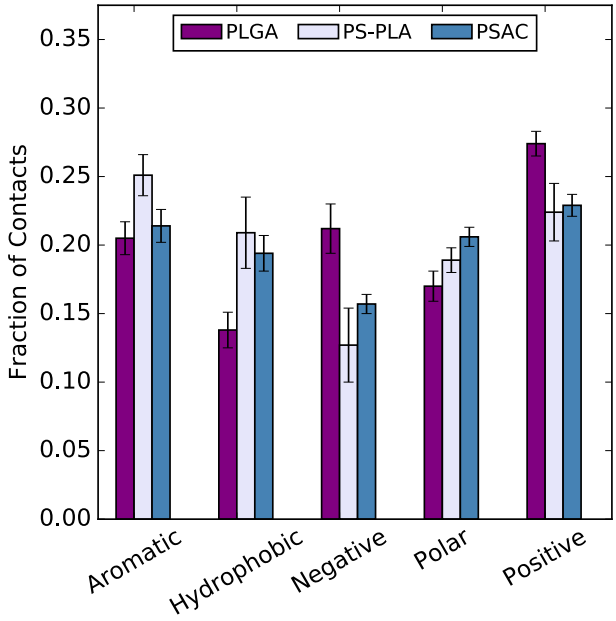

Figure 9.7. Fraction of contacts for different amino acid types on the surface of BSA in an aqueous solution with PSAC compared to with PS-PLA or PLGA. Error bars represent the average standard deviation of calculations performed on consecutive 25 ns portions over the second half of each 200 ns simulation trajectory (i.e., sample size of n=4).

At a neutral pH of 7.2, acrylate side chains in PSAC molecules would be almost completely deprotonated, as they were in our simulations. Interestingly, additional simulations we ran with protonated acrylate side chains showed rapid aggregation of PSAC molecules around the surface of BSA (Figure A5.7). This finding highlights that a small change in the chemical functionality of the NP could dramatically shift the balance of polymer self-interactions versus protein/polymer interactions, thereby affecting protein release rates. As previously hypothesized, some NPs such as those made with PS-PLA copolymers may form a tight core with polymer that is not exposed to physiological conditions. Thus, changes in NP formulation methods may lead to tunable core properties to control the degree of aggregation and thus protein release kinetics.

9.4.6 Effect of Surrogate Length on Binding Convergence and Characteristics

By exploring the preference of polymer surrogates of varying lengths for specific amino acid types on the protein surface, we can assess the validity of using small polymer molecules as proxies for more realistic systems. To stringently test this idea, we performed additional MD simulations of BSA in aqueous solutions with 2 wt.% of the monomeric forms of hydrophilic LGA and hydrophobic SAC. Figure A5.8 shows the RMSD of the Cα atoms from the crystal structure of BSA in these new solutions is very similar to the RMSD values in solutions with the trimeric

forms of these NP substituents (i.e., PLGA and PSAC, respectively), discussed in earlier sections of the manuscript. A comparison of the fraction of amino acid types on the surface of BSA contacted by the monomeric and trimeric polymer surrogates is shown in Figure 9.8, and Figure A5.9 shows these contacts converge in similar timeframes for the monomeric and trimeric polymer surrogates (~75-100 ns). The results of Figure 9.8 indicate that in general, the binding characteristics of monomeric LGA and SAC closely match those of trimeric PLGA and PSAC, respectively, for all five amino acid types on the surface of BSA. For example, the amino acids to which the surrogates bind most strongly and weakly are the same for each monomer/trimer pair: LGA/PLGA bind most strongly to positively charged residues and most weakly to hydrophobic residues, and SAC/PSAC bind most strongly to positively charged residues and most weakly to negatively charged residues. Interestingly, Figure 9.8 shows similar binding to aromatic residues for both the hydrophilic and hydrophobic polymer surrogates. This is likely due to the close proximity of the benzene ring and carboxylate ion on the SAC/PSAC backbone that limits the frequency and strength of aromatic interactions with surface residues. This was discussed earlier as the main reason for why PSAC does not aggregate in solution like PS-PLA with similar hydrophobic moieties. Overall, these findings suggest that qualitatively, the use of small polymer surrogates, even with the greatest simplification of simulating long-chained polymers as monomeric units, can be used to decipher and predict governing principles of protein/polymer interactions.

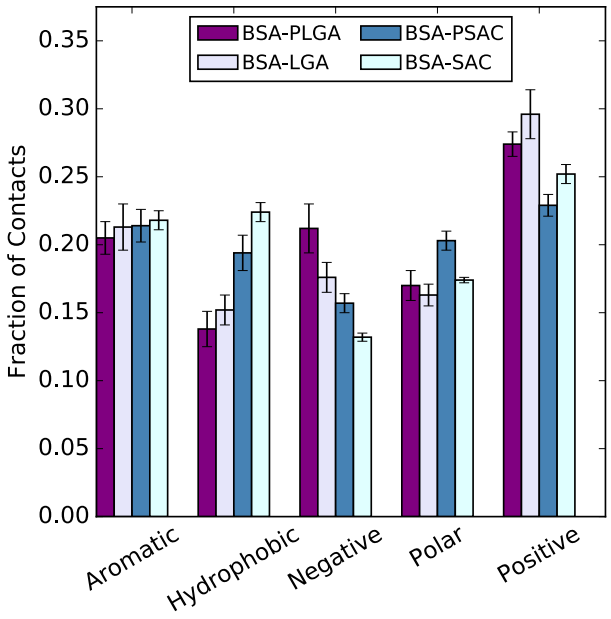

Figure 9.8. Fraction of contacts for different amino acid types on the surface of BSA in solutions with monomeric or trimeric LGA or PLGA copolymer, or SAC or PSAC copolymer, respectively. Error bars represent the average standard deviation of calculations performed on consecutive 25 ns portions over the second half of each 200 ns simulation trajectory (i.e., sample size of n=4).

Summary and Conclusions

In this work, we demonstrate a combined experimental and computational approach to gain insight into the driving forces of protein release from polymeric nanoparticles. This approach provides a holistic picture of BSA release from NPs by describing physical interactions with macroscopic and microscopic detail.

Experimental results suggest transport limitations do not significantly restrict protein diffusion from larger NPs. The similarity between measured BSA release rates from PLA and PS-PLA NPs, and support from MD simulations, imply the diblock copolymer PS-PLA assumes a core-shell structure. The simplified system emphasizes the importance of physical interactions between a protein and trimeric polymer surrogates and selectively ignores polymer conformation and transport limitations to protein release. Although rapid screening of physical interactions between polymer constituents and a target protein is computationally tractable, it is unlikely that these physical interactions are sufficient descriptors for protein retention in the complex NP environment. The relative affinity of a polymer for self-interaction over protein interaction seems to be an important factor in release. PS aggregates in normal MD solutions, corresponding to exclusion or rapid clearance of protein in practice. A more complete model of the microscopic environment is difficult to clearly define, but is perhaps necessary for an accurate description of the atomic scale interactions of encapsulated proteins.

Spatial binding resolution for protein-nanoparticle systems is of interest for a wide range of biomedical applications. Beyond controlled release, the potential for design of synthetic particles that specifically interact in biological signaling and protein interaction pathways depends on the precision with which we can target protein surface patches. This work demonstrates the feasibility of progressing from chemical intuition to experimental and computational observation of specific protein-biomaterial interactions.

Coda

Impact

Our work modeling ILs and investigating protein behavior in IL solutions has received attention from the scientific community both domestically and internationally, resulting in 30+ citations to date in the past two years. Our publications have been cited in computational studies investigating: the influence of ILs on the structure of Trpcage mini-protein[384] and red fluorescent protein mCherry[385], the solvation of model cellulose compounds in ILs for biomass conversion[386-387], CO_2 vibrational frequencies in ILs for industrial gas separation[388], and ILs as catalysts for CO_2 fixation[389]. In addition, our investigation of the stability of newly-discovered LPMO enzymes in ILs (Chapter 6) was discussed in a recent experimental study by Tanghe et al. to increase the thermostability of an LPMO enzyme through the introduction of additional disulfide bridges[390]. Furthermore, our efforts have paved the way for additional studies within our own research group at the University of Washington, for example an investigation into the kinetics of IL-induced protein unfolding[331].

Our work on biomolecular behavior at the protein/surface interface has also inspired both computational and experimental research, here in the US and abroad. A recent experimental study by Raigoza and coworkers investigated peptide adsorption on surfaces with different chemical functionalities, to aid in the design of surfaces with biological/inorganic functions through "click" chemistry[391]; our findings from simulations performed on similar systems (Chapter 1), were used to help to explain the authors' own findings. In addition, simulations citing our work have been performed to investigate the influence of surface-bound water molecules on the orientation and conformation of adsorbed peptides[392], as well as peptide adsorption to a gold surface.[393]

Overall, this research has led to new fundamental insights into the structure and function of biomolecules at interfaces, whether it be the protein/liquid interface or the protein/surface interface. Our findings are expected to have impacts across many fields and processes, including fouling, biomineralization, medicine, biocatalysis, and numerous other applications where engineered host environments (e.g., gels, polymers, and colloids) have been shown to play a large role in determining the structure and function of biomolecules.

Future Work

Collectively, this work represents a step forward in our understanding of protein interfacial behavior. However, there remain many aspects of protein adsorption that require further exploration and study. At the onset of this research, for example, it was commonplace in the field – and still is to some extent – to perform adsorption simulations with an idealized surface that is mostly, or even completely, frozen. The studies described in this document utilized recently developed force fields (e.g., for the flexible silica surface in Chapter 2) that were specifically designed for interfacial simulations by taking into account relevant surface chemistries like pH,

common defects, and cleavage planes.[108] It is only in the past few years that these types of models have been developed, and they have as yet only become available for a handful of biological surfaces. As novel experimental techniques for surface analysis develop, providing more data – and new types of data – on complex interfacial systems, interfacial models for a diverse range of surfaces should be developed, and future protein adsorption simulations should utilize them. Additionally, building off of our work that showed surface defects can have large impacts on protein interfacial structure (Chapter 3), future studies should further investigate the role of surface defects on the adsorption process to ensure simulations represent experimental conditions as accurately as possible. Furthermore, as computational processing power increases with developments in computer hardware/software, it will become increasingly feasible to simulate larger and more complex systems. This will enable future simulations of interfacial processes to investigate the role of multiprotein interactions, which are likely to have significant effects on both the thermodynamics of adsorption and the structure of adsorbed proteins.

Related to the protein/liquid interface, Chapter 7 presents the most detailed study to date of the validity of AMBER-family force fields for studying protein/IL interactions (to our knowledge). However, many other force fields have been used to study protein/IL simulations, and their accuracy should be similarly tested in the future. The simulations described in Chapter 7 featured a single lipase-family enzyme, providing insight into specific interactions of IL ions with different residue types on the protein's surface (e.g., hydrophobic vs. charged residues). However, to truly determine governing features of protein/IL systems to effectively guide future experimental efforts, our approach should be repeated with a wide range of proteins or enzymes from multiple different classes or families. We recommend the use of automation/workflow tools to carry out these simulations in a systematic fashion, and the application of machine learning or other data science tools to predict dominant surface binding features of protein/IL systems. We hope that our approach for simulating protein/IL systems and extracting quantitative information from them for direct comparison against experimental data will inspire future simulations of proteins in other types of media for a variety of applications.

Appendix I

Table A1.1. Setup of PTMetaD-WTE Simulations.

System	Peptide	SAM	Number of Replicas	Temperature Range (K)	Biasfactor
I	LKα14	COO/H	12	300-450	10
II	LKα14	COO/H	12	300-450	10
III[a]	LKα14	CH_3	25	300-500	10
IV	LKβ15	COO/H	12	300-450	10
V	LKβ15	COO/H	12	300-450	10
VI[a]	LKβ15	CH_3	25	300-500	10
VII	none	COO/H	12	300-450	10
VIII	none	CH_3	12	300-450	10
IX[b]	Trpcage	COO/H	20	300-450	12
X[b]	Trpcage	CH_3	20	300-450	12

[a]Data from Deighan et al.[29]
[b]Data from Levine et al.[46]

A Slight Tweak to the PTMetaD-WTE Protocol

In this study, the production runs were performed in a slightly different manner than previous applications of the method by us and others. Previously, we have included the WTE potential energy biases made in the WTE setup simulation as static biasing potentials acting on each individual replica during the production run, allowing for the biasing of other MetaD collective variables (CVs). In preliminary simulations we noticed that the lengthy simulation times, high driving forces from the protein adsorption, and the large WTE biasfactor (50) that was applied could lead to some spurious effects unless very lengthy initial 'setup' times were used in order to obtain sufficiently smoothed WTE static bias potentials. Therefore, we used the following modification to our sampling protocol. After the initial PE-only bias was applied (over the 10 ns/replica PTMetaD simulation), MetaD "hills" were concurrently deposited on two separate bias potentials. The peptide/surface CVs were biased (in a 2D bias potential) according to the description in the main text (i.e., the original PTMetaD-WTE algorithm) and the 1D WTE bias potential continued to grow with a very slow hills deposition rate (we increased the spacing between hill additions 5-fold during the production run).

Since the transient period of the PE bias was essentially over during the initial setup simulation and the PE of the explicitly solvated system is essentially uncorrelated to the biased peptide CVs, this additional step only affects the simulation in that it beneficially smoothes out the PE bias potentials, especially at the tails of the distribution where spurious effects might be seen if the initial setup were not performed for sufficiently long simulation times. This whole process

is trivially implemented with the PLUMED plugin and suggested for future applications of PTMetaD-WTE.

Evaluating the Temperature Dependence of the Potential Energy

Defining the change in Helmholtz free-energy on binding as $\Delta A_{ads}=\Delta U_{ads}-T\Delta S_{ads}$ assumes that the change in energy upon adsorption (ΔU_{ads}) is constant with temperature. To test this assumption, we calculated the temperature dependence of the change in potential energy upon adsorption of LKα14 to the COO/COOH SAM as a test case. The potential energy of each replica was monitored with PLUMED during the metadynamics (MetaD) simulation. Following the protocol of Fleming et al., a reweighting algorithm[75] was used to calculate the unbiased probability distribution of the potential energy from the MetaD simulation.[185] A two-dimensional free-energy surface F(CV1,U) was constructed for each replica, where CV1 was the distance between LKα14 and the surface (peptide-surface separation distance or SSD) and U was the total potential energy of the system. The ensemble average value of the potential energy at each value of CV1 (n=51 data points) was then obtained for each replica by taking a Boltzmann-weighted average over all potential energy values at each point in CV1 space:

$$\langle U(CV1)\rangle = \sum U e^{-\beta F(CV1,U)} \tag{A1.1}$$

where F is the free-energy and β is the Boltzmann constant multiplied by the replica temperature. These results are shown in Figure A1.1. In regards to this analysis, the data was calculated assuming the bias added to the potential energy during the MetaD simulation did not change the overall distribution of the potential energy.[55] The average potential energy change upon adsorption was then calculated by subtracting the value of U of the solvated state (average SSD in the range of 1-3.5 nm) from the value of U of the adsorbed state (SSD of ~0.7 nm). To avoid any influence on the potential energy from the restraining force of the wall that began acting on the peptide at an SSD of 4.5 nm (see System Setup, main text), the region over which the potential energy of the solvated state was calculated was restricted to an upper SSD value of 3.5 nm. Figure A1.2 shows the results for the average potential energy change upon adsorption as a function of temperature. A color map is included to clarify the temperature of each replica, and error bars represent the standard deviation in the potential energy over a range in SSD of 1-3.5 nm. Before drawing any conclusions from the results, it is important to note that the quantity of ΔU_{ads} calculated in this analysis is a function of the entire potential energy of all degrees of freedom, whereas the quantity of ΔU_{ads} obtained from a linear plot of ΔA_{ads} versus temperature (see Figure 1.2, main text) represents the interactions that are dominant in the definition of CV1. However, a linear fit of the data in Figure A1.2 shows that no dominant temperature trend exists, which implies that our model of ΔU_{ads} is a sufficient approximation of ΔU_{ads}(CV1).

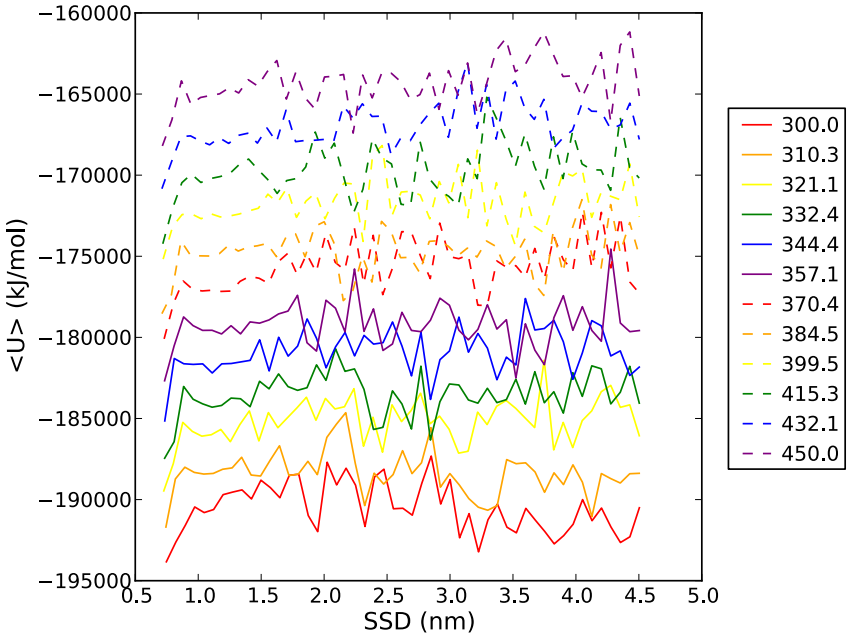

Figure A1.1. Reweighted potential energies averaged over discrete peptide-surface separation distances (SSD) for LKα14 on the COO/COOH SAM (n=51 slices over the range of x-axis values shown above) as a function of temperature.

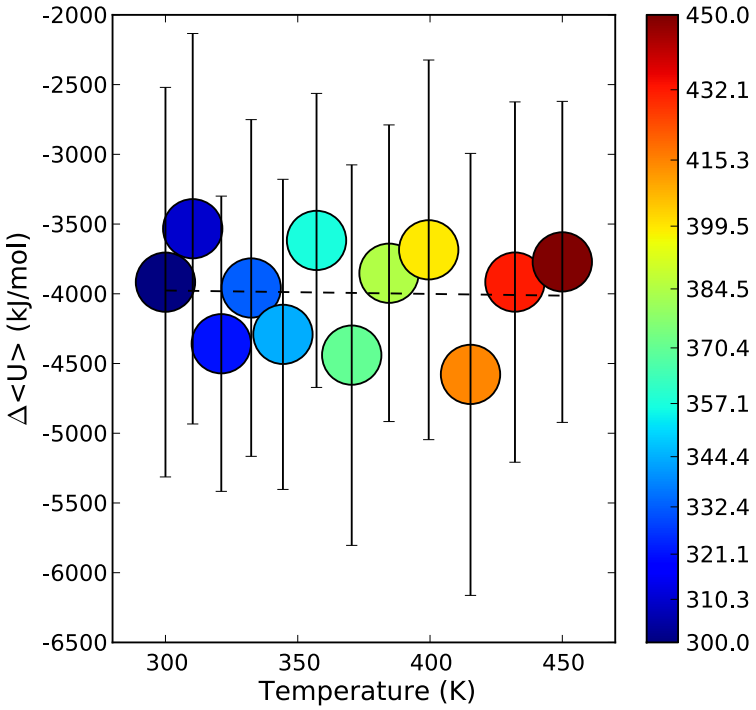

Figure A1.2. Calculated change in potential energy upon adsorption (adsorbed - solvated state) of LKα14 to the COO/COOH SAM as a function of temperature. Adsorbed potential energies correspond to an SSD value of ~0.7 nm (c.f. Figure A1.1) and solvated potential energies were calculated as averages over a range in SSD of 1-3.5 nm. Error bars represent the standard deviation in potential energy over a range in SSD of 1-3.5 nm. A linear fit of the data is shown as a dotted black line.

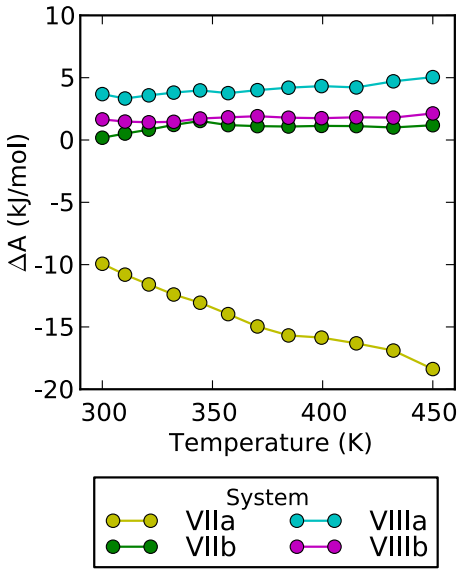

Figure A1.3. Calculated binding free energies versus replica temperature for systems VIIa/b and VIIIa/b, corresponding to systems VII (COOH/COO SAM) and VIII (CH_3 SAM) listed in Table 1 (main text), respectively, where a single ion/water molecule was biased. The lines represent binding free energies of step C of the thermodynamic cycle of adsorption shown in Figure 1.3 (main text).

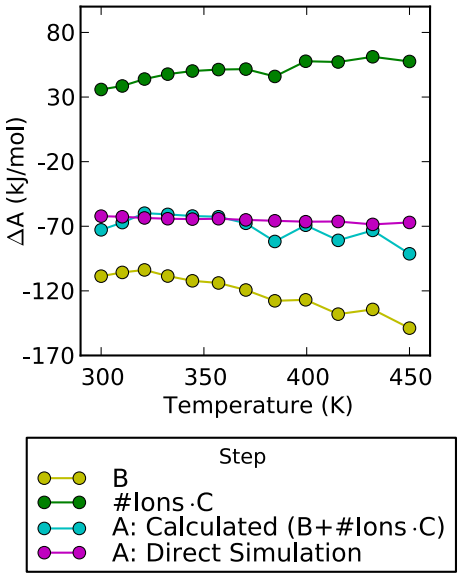

Figure A1.4. Calculated binding free energies for the three steps of the thermodynamic cycle for LKα14 on COOH/COO, using a value of *x,* or the average number of ions desorbing from the surface, unique to each replica temperature. Yellow and green lines represent steps B (system I, corrected for ion desorption) and #Ions·C (system VIIa) of the mechanism proposed in Figure 1.3 (main text), respectively, whereas purple and blue lines represent the results from the direct simulation of step A (system II) and the calculated results for step A based on adding the yellow and green lines (B+#Ions·C).

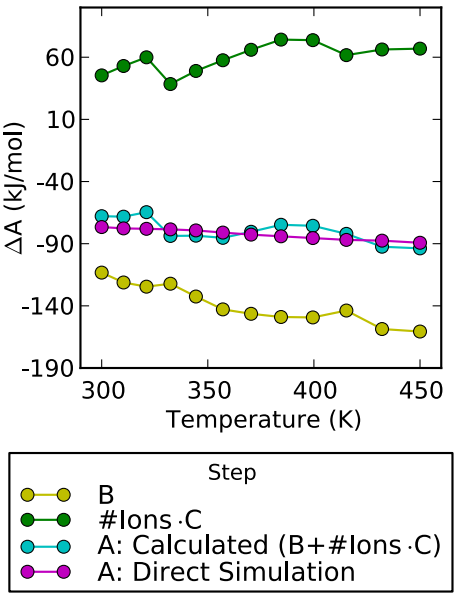

Figure A1.5. Calculated binding free energies for the three steps of the thermodynamic cycle for LKβ15 on COOH/COO, using a value of *x,* or the reweighted number of ions desorbing from the surface, unique to each replica temperature. Yellow and green lines represent steps B (system IV, corrected for ion desorption) and #Ions·C (system VIIa) of the proposed mechanism, respectively, whereas purple and blue lines represent the results from the direct simulation of step A (system V) and the calculated results for step A based on adding the yellow and green lines (B+#Ions·C).

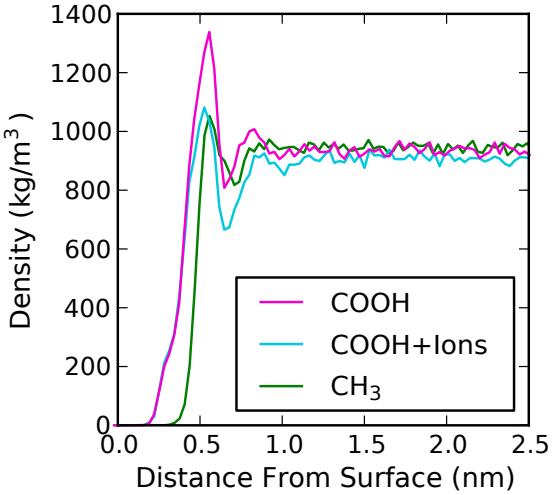

Figure A1.6. Unidirectional density of water above: 1) the COOH/COO surface (pink curve, system VII); 2) the COOH/COO surface with 100 additional sodium and chlorine ions each (blue curve); and 3) the CH_3 surface (green curve, system VIII). The z-component of the topmost frozen carbon atom (C10) in each surface was used as the point of reference.

Appendix II

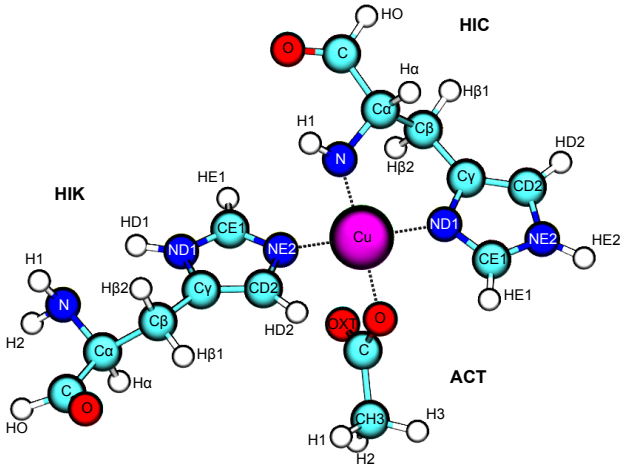

Figure A2.1. Atoms used in quantum mechanics calculations of the copper-histidine coordination complex in enzymes *Sc*LPMO10B and *Sc*LPMO10C.

Table A2.1. Calculated partial charges for the atoms in Figure A2.1.

Residue	Atom	Charge	Residue	Atom	Charge
HIC	N	-1.070	HIK	N	-0.920
	Cα	0.478		Cα	0.473
	C	0.477		C	0.358
	O	-0.476		O	-0.447
	Cβ	-0.529		Cβ	-0.454
	Cγ	0.292		Cγ	0.272
	ND1	-0.198		ND1	-0.237
	CD2	-0.296		CD2	-0.336
	CE1	0.024		CE1	0.086
	NE2	-0.219		NE2	-0.179
	H1	0.365		H1	0.361
	Hα	0.040		H2	0.342
	HO	0.004		Hα	0.005
	Hβ1	0.158		HO	0.019
	Hβ2	0.206		Hβ1	0.167
	HD2	0.254		Hβ2	0.154
	HE1	0.213		HD1	0.308
	HE2	0.379		HD2	0.243
ACT	C	0.993		HE1	0.216
	O	-0.762	Cu	Cu	1.105
	OXT	-0.697			
	CH3	-0.563			
	H1	0.133			
	H2	0.125			
	H3	0.131			

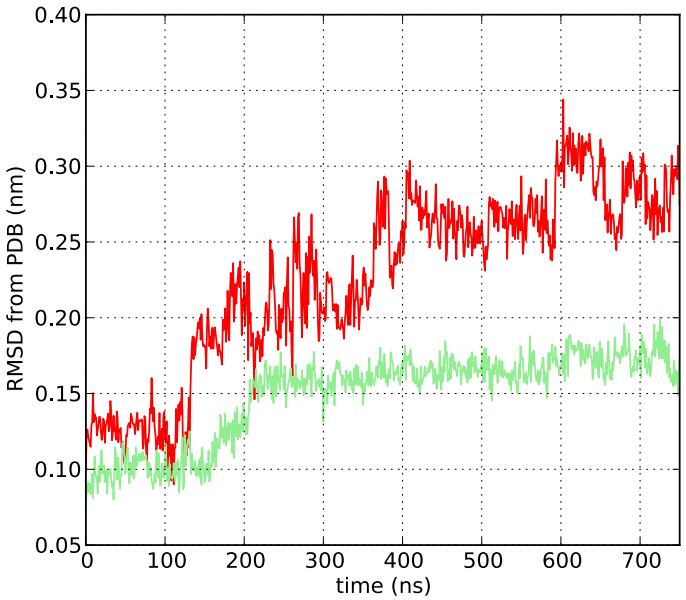

Figure A2.2. Time series of the Cα RMSD from the crystal structure for extended systems *Sc*LPMO10B/20%[BMIM][Cl]-trial 1 (green) and *Sc*LPMO10C/10%[AMIM][Cl]-trial 1 (red).

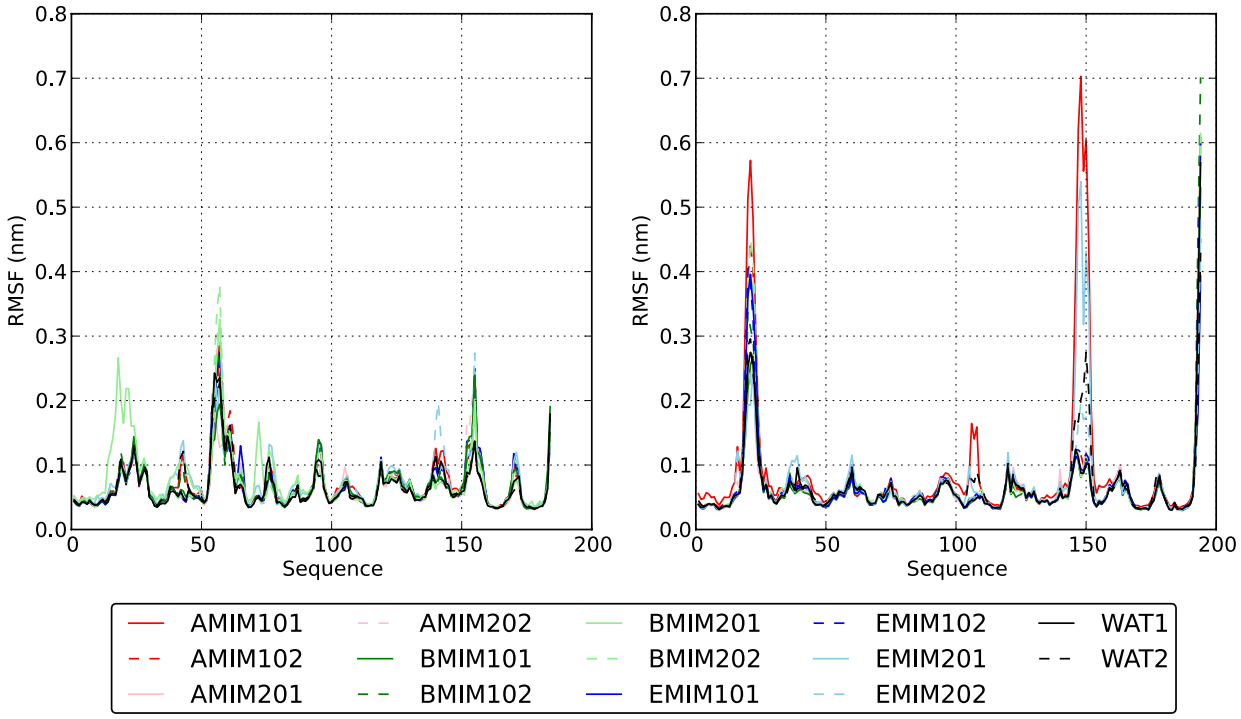

Figure A2.3. Side-by-side comparison of the time series of the Cα RMSF of *Sc*LPMO10B (left) and *Sc*LPMO10C (right) from the crystal structure for each enzyme/solvent combination. Legend entries indicate the IL, wt% in water, and trial number.

Appendix III

Table A3.1. Characterization of cation binding sites on the surface of WT-lipA as determined through a combined experimental and computational approach.

Binding Site	Residue Numbers*	Amino Acids	Site Color
1	27, 33	VAL, ARG	Blue
2	89, 90	ASN, LEU	Red
3	152	HIS	Purple
4	157, 161	ILE, TYR	Dark Pink
5	44, 45	LYS, THR	Yellow
6	12, 13, 108	ILE, GLY, LEU	Cyan
7	107, 142, 144	ARG, ARG, ASP	Green
8	42,49	TRP, TYR	Light Pink
9	119	PRO	Violet
10	17	PHE	Black
11	37	TYR	Orange
12	134,135, 137,139	MET, ILE, MET, TYR	Gray

*Consistent with numbering in PDB code 5CRI

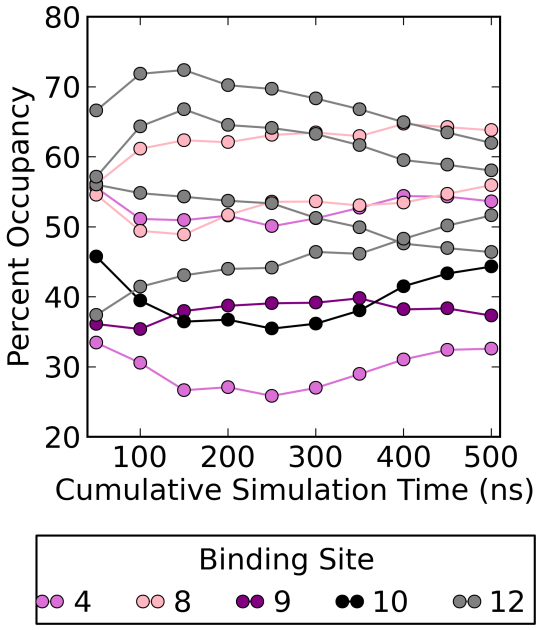

Figure A3.1. Percent occupancy of WT-lipA residues by IL cations, calculated over increasing periods of simulation time, for the residues in the top five cation-binding sites identified from the experimental crystal structures.

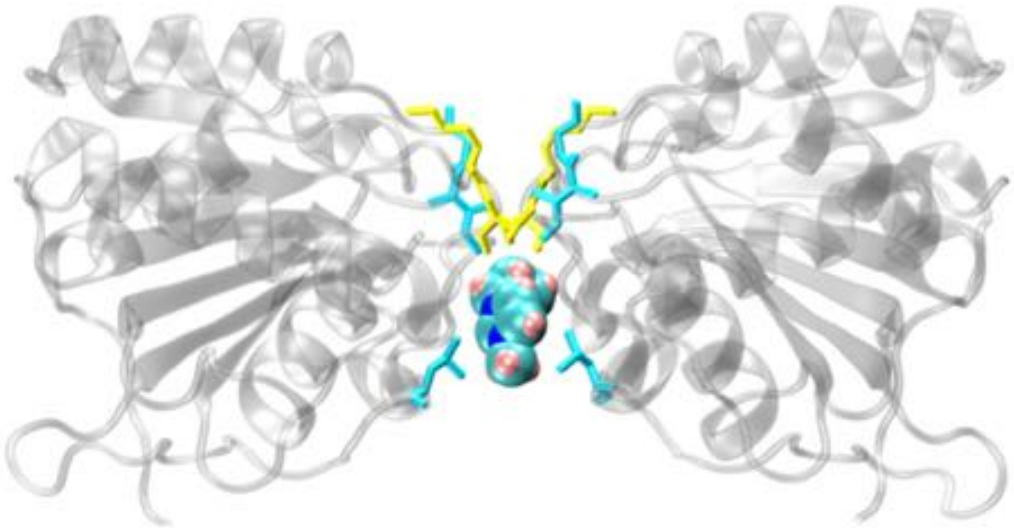

Figure A3.2. Snapshot of an IL cation stabilized between two proteins in the crystal structure of lipA in 20% aqueous IL solution (PDB code 5CT6). Residues in binding sites 5 (yellow) and 6 (cyan) are shown in licorice representation. The IL cation is shown in van der Waals representation with carbon, nitrogen, and hydrogen atoms shown in cyan, blue, and pink, respectively.

Additional Metrics for Binding Site Clarification

The residence time distribution ($rtd(\Delta\tau_k)$) for an IL cation ([BMIM]) around a given residue in lipase was calculated via the following correlation function:

$$rtd(\tau_k) = \sum_{j=1}^{N_{t\max}-k} \sum_{i=1}^{N_{[BMIM]}} v_i(t_j)v_i(t_j+\tau_k), \qquad k = 0,1,...,(N_{t\max}-1), \tag{A3.1}$$

where τ_k is the k^{th} time interval ($\tau_k = k\Delta\tau$ and $\Delta\tau = 20$ ps). In Equation 1, $v_i(t)$ acquires a value of 0 or 1 depending if any atom within the i^{th} cation molecule is outside or inside, respectively, a 4Å radius shell surrounding the enzyme at time t. Figure A3.3 shows the IL cation residence time around each residue in WT-lipA for the different IL concentrations. We note that the trends of Figure A3.3 are not dependent upon the size of the time window ($\Delta\tau$) we chose to perform the calculations; results are similar over the range in window size from 20-200 ps (data not shown), thus ensuring that the re-migration of a cation molecule into the hydration shell after vacating it for some time does not influence the results.

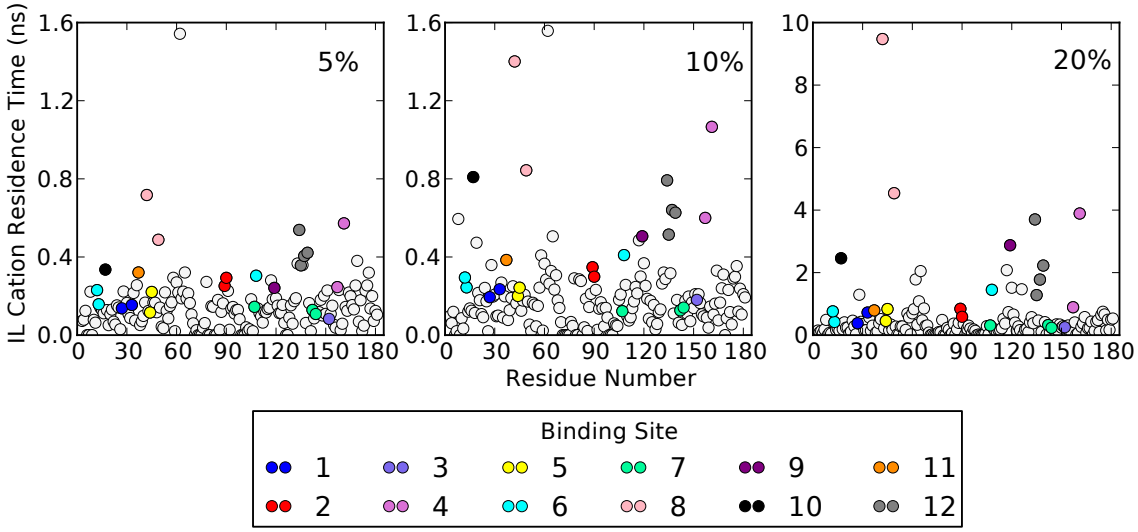

Figure A3.3. Residence time of IL cation binding to residues in WT-lipA as a function of IL concentration in water. Colored circles represent experimentally identified binding sites, and uncolored circles represent all other protein residues.

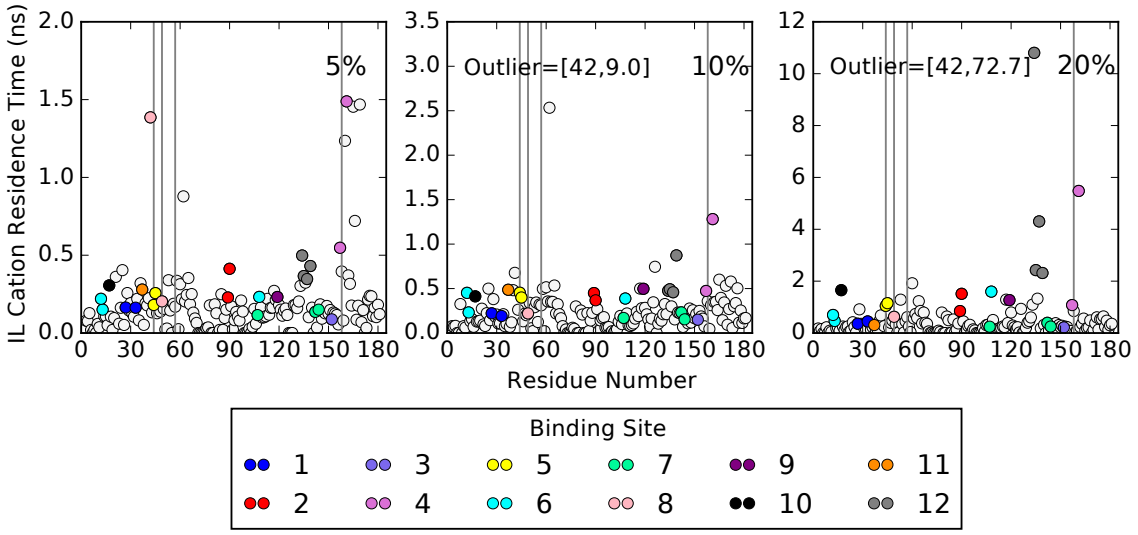

Figure A3.4. Residence time of IL cation binding to residues in QM-lipA as a function of IL concentration in water. Colored circles represent experimentally identified binding sites, and uncolored circles represent all other protein residues. Vertical gray lines correspond to residues mutated in mutagenesis experiments.

In general, the results for residence time are less consistent across the different IL concentrations than the occupancy results of Figure 7.2 (main text). Our hypothesis is that the significantly higher residence times for the 20% IL solution likely reflect a change from a dilute salt solution (5/10% IL) where water can effectively shield the electrostatic effects of ILs, to a concentrated salt solution where electrostatics dominate and lead to significant reductions in diffusion of IL ions to and from the enzyme surface. Another factor that could contribute is the possibility of cations swapping places near a given lipase residue from one time to the next, which

our script would not be able to distinguish and this would be much more likely to occur in the concentrated 20% IL solution.

Figure A3.3 provides some new information in regards to residue V62, which shows a high cation residence time at all IL concentrations, indeed the highest residence time of all residues in lipase for the 5 and 10% IL solutions. This is especially interesting because the occupancy values in Figure 7.2 (main text), while of significance, do not indicate V62 to be nearly as important as Figure A3.3 does. The contrast between these two metrics can be understood by considering the stability and location of V62 on the enzyme's surface: its exposure to solvent is limited by its location on the underside of a helix (Figure A3.4), hindering cation diffusion to the site, however this same location on a structured region of the enzyme (as opposed to a flexible loop region), combined with a short sidechain length, leads to strong interactions with cations once they are bound. The above analysis provides a case for using multiple metrics to gain a complete picture of IL binding to the surface of an enzyme.

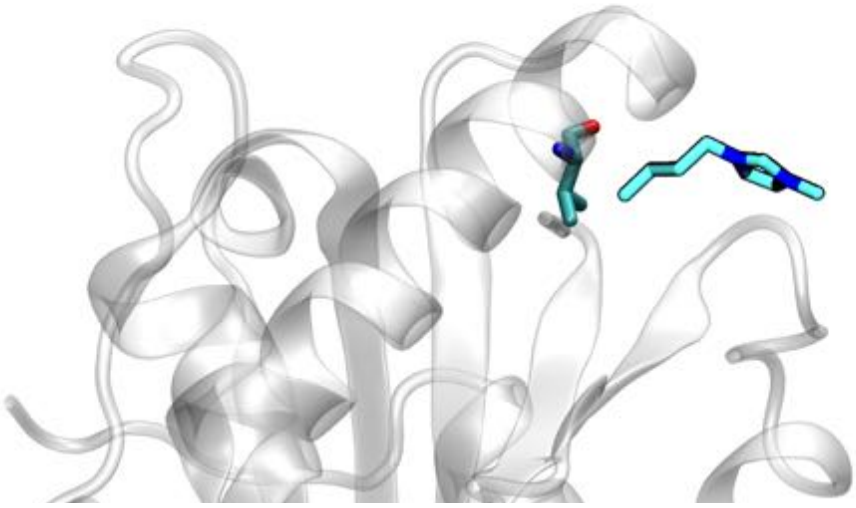

Figure A3.5. Simulation snapshot from the 5% IL solution simulation highlighting the location of residue 62 (valine) on the surface of lipase (in gray) and its interaction with an IL cation. Hydrogen atoms not pictured.

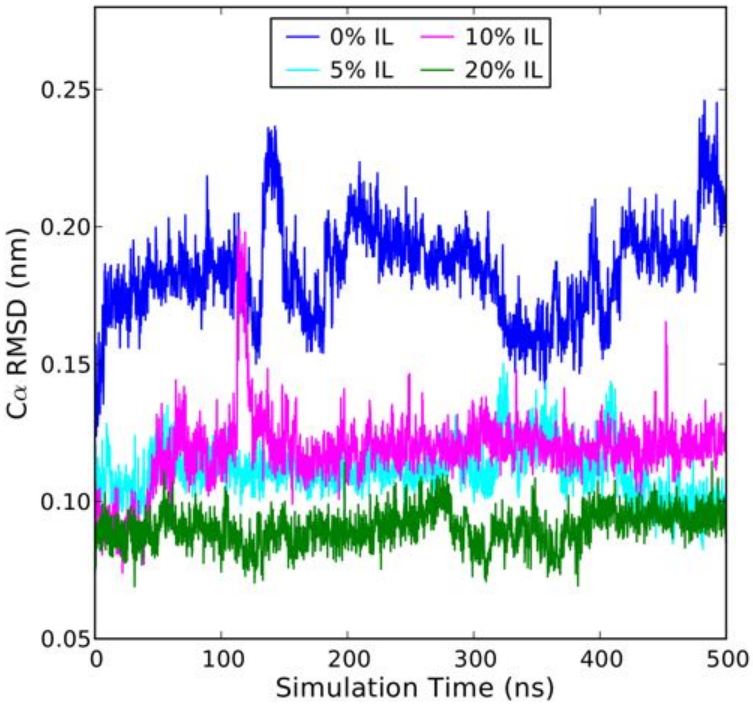

Figure A3.6. Time series of the root-mean-squared deviation (RMSD) of distances between Cα atoms in WT-lipA for each of the different IL concentrations studied. Data was collected from the trajectory every 20 ps and plotted every 10th point.

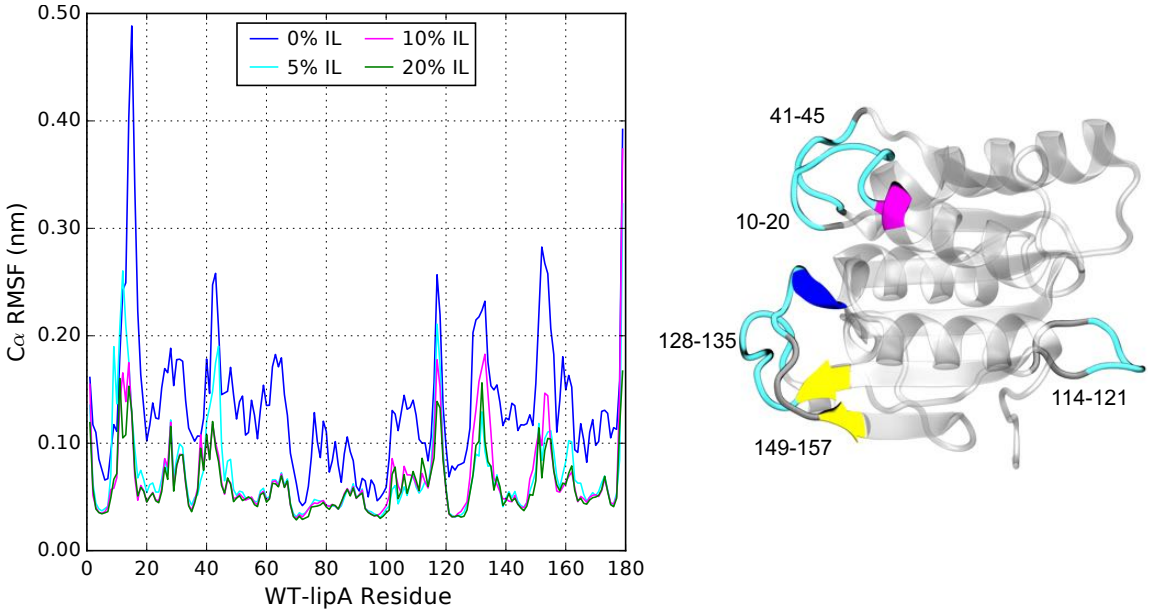

Figure A3.7. (left) Time series of the root-mean-squared fluctuation (RMSF) of Cα atoms (sequence) in WT-lipA for each of the different IL concentrations studied. (right) crystal structure of WT-lipA, highlighting regions of high RMSF values. Number ranges next to highlighted areas indicate the corresponding sequence range. Magenta and blue coloring indicate α- and 3_{10}-helices, respectively, and yellow, cyan, and dark gray coloring indicate β-sheets, turns, and random coils, respectively.

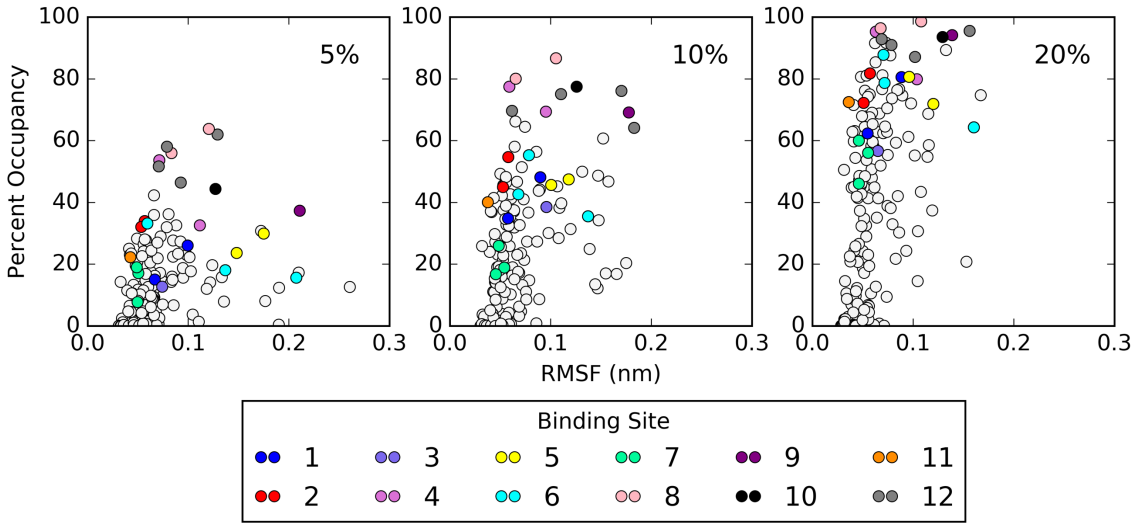

Figure A3.8. RMSF vs. percent occupancy of residues in WT-lipA by IL cations as a function of IL concentration in water. Colored circles represent experimentally identified binding sites, and uncolored circles represent all other protein residues.

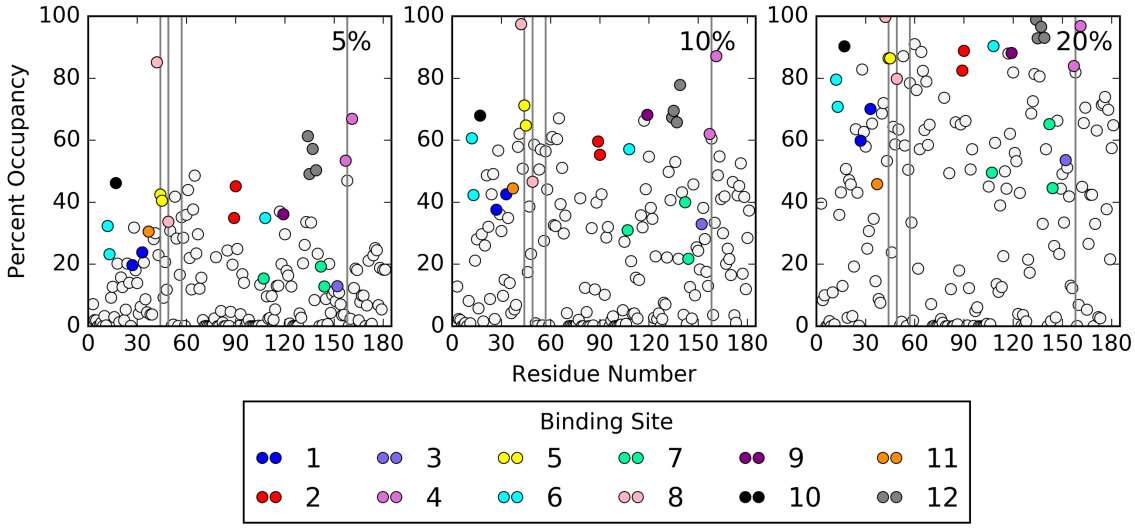

Figure A3.9. Percent occupancy of residues in QM-lipA by IL cations as a function of IL concentration in water. Colored circles represent experimentally identified binding sites, and uncolored circles represent all other protein residues. Vertical gray lines correspond to residues mutated in mutagenesis experiments.

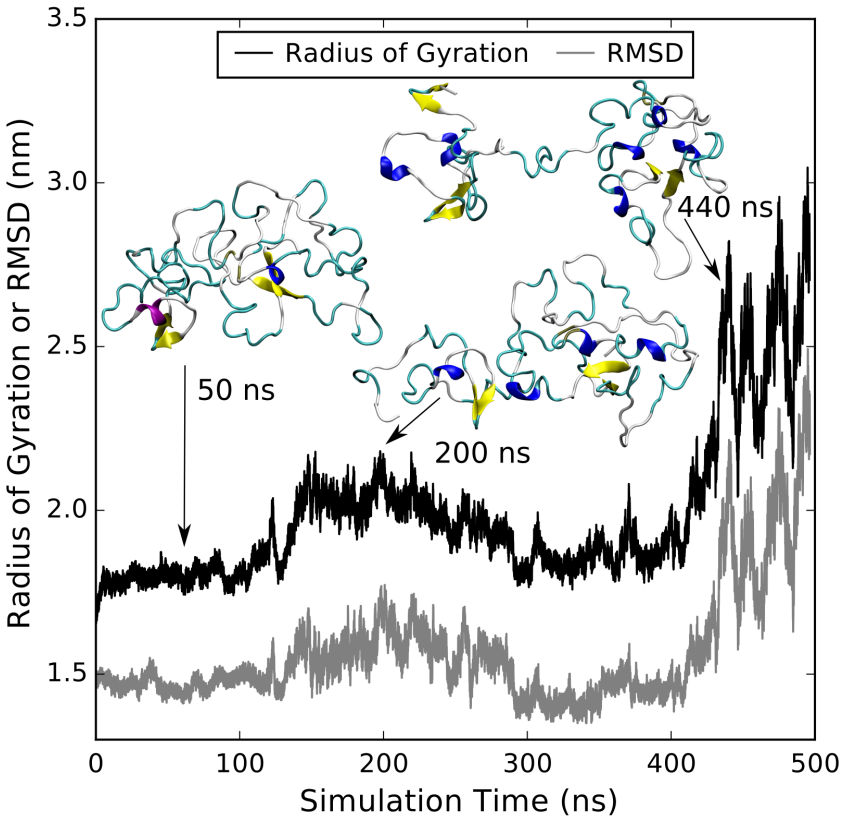

Figure A3.10. Radius of gyration and RMSD (relative to the crystal structure) of denatured WT-lipA versus time with corresponding simulation snapshots (insets).

Appendix IV

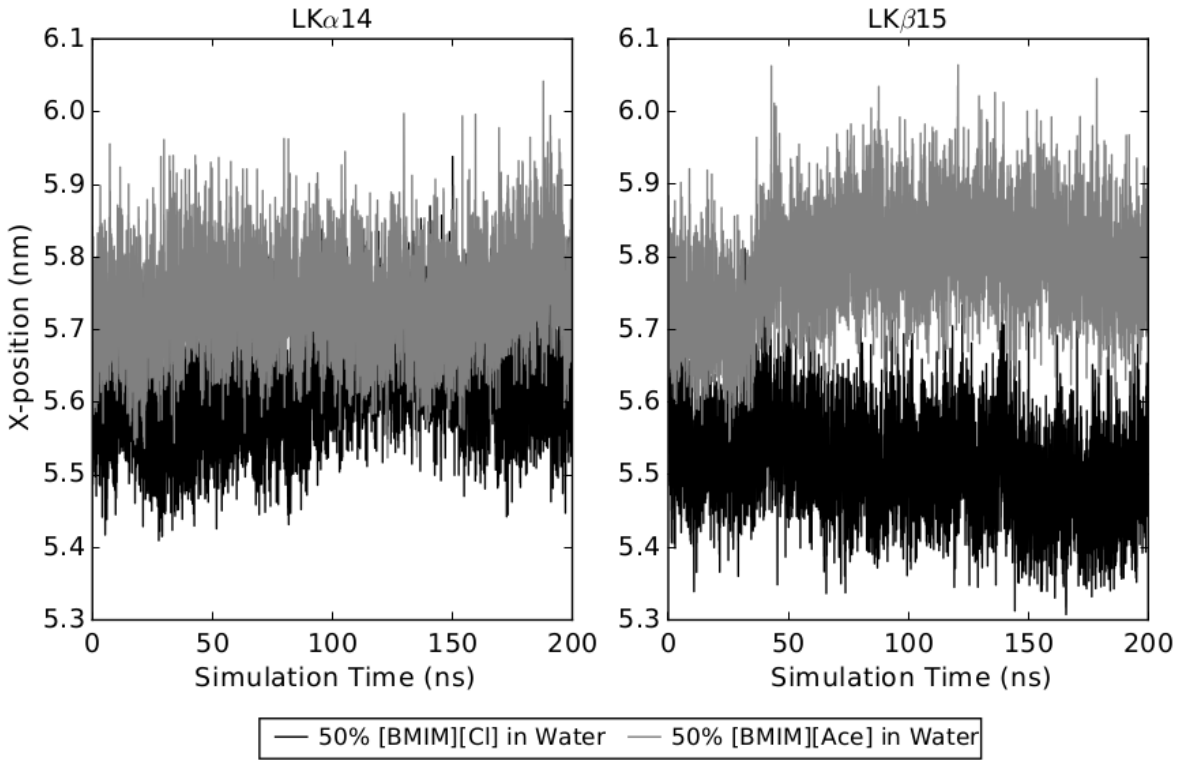

Figure A4.1. Average x-position of the five highest central ring carbons in [BMIM] (from the bottom of the simulation box) as a function of simulation time, for the simulation of (left) LKα14 and (right) LKβ15 in the IL solutions.

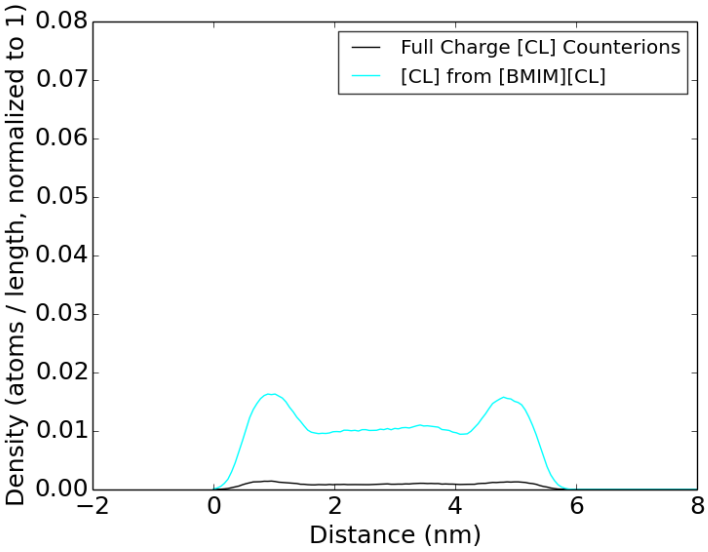

Figure A4.2. Atomic species densities for LKα14 in 50% aqueous [BMIM][Cl]. Data for a single trial is shown of the three total trials. Note that solvent is present between 0 and 6 nm, while the area below 0 nm and above 6 nm represents vacuum. The y-axis is the same as in Figure 8.5 (main text) for easy comparison across species.

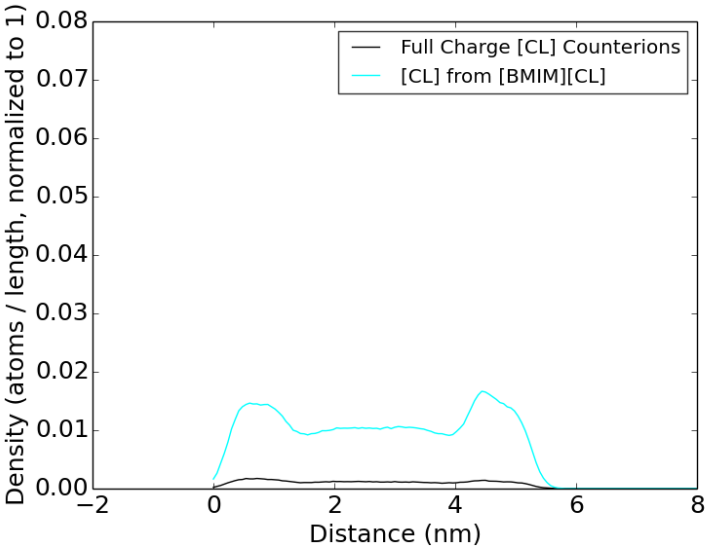

Figure A4.3. Atomic species densities for LKβ15 in 50% aqueous [BMIM][Cl]. Data for a single trial is shown of the three total trials. Note that solvent is present between 0 and 6 nm, while the area below 0 nm and above 6 nm represents vacuum. The y-axis is the same as in Figure 8.5 (main text) for easy comparison across species.

Appendix V

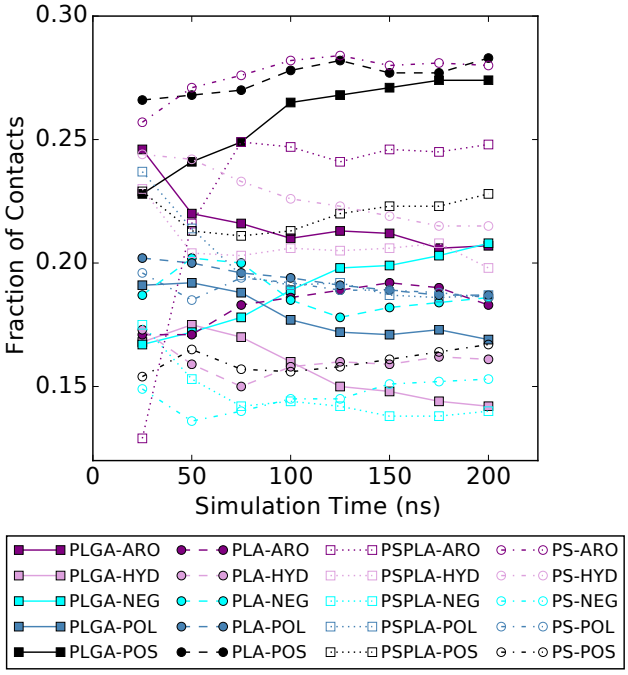

Figure A5.1. Convergence of amino acid contacts on BSA in polymer surrogate solutions.

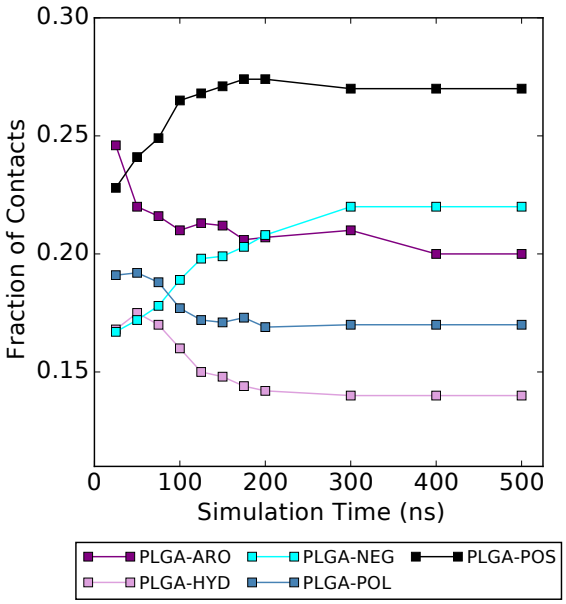

Figure A5.2. Convergence of amino acid contacts on BSA in a 2 wt.% PLGA solution over an extended period of 500 ns.

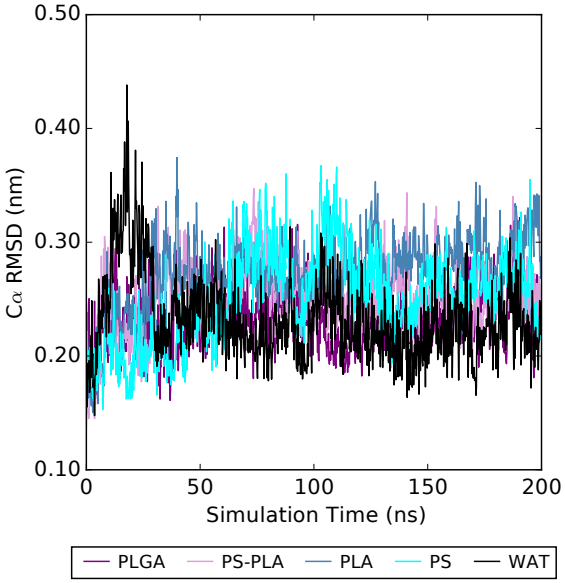

Figure A5.3. Root mean squared deviation (RMSD) of α-carbons (Cα) in BSA compared to the crystal structure in the different aqueous polymer solutions and in pure water.

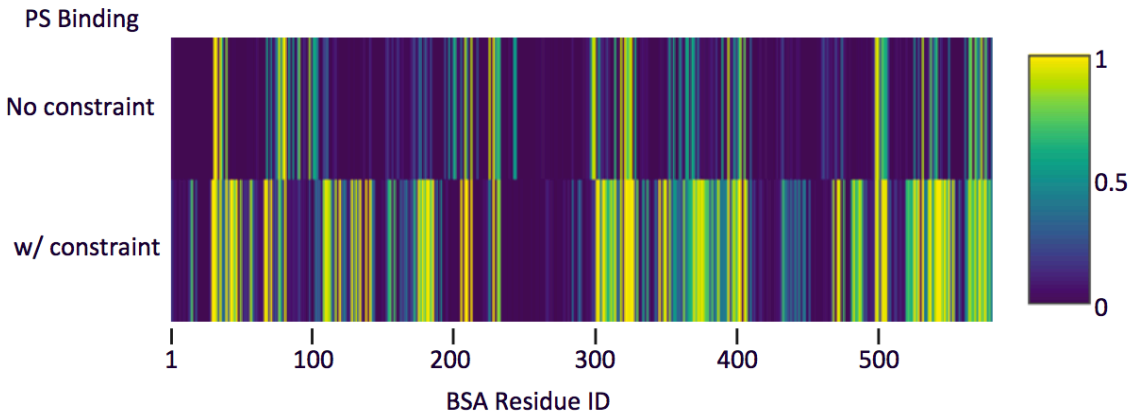

Figure A5.4. Occupancy heat maps for PS simulation with and without coordination constraint. Aggregation in unconstrained simulation (top) results in relatively sparse binding.

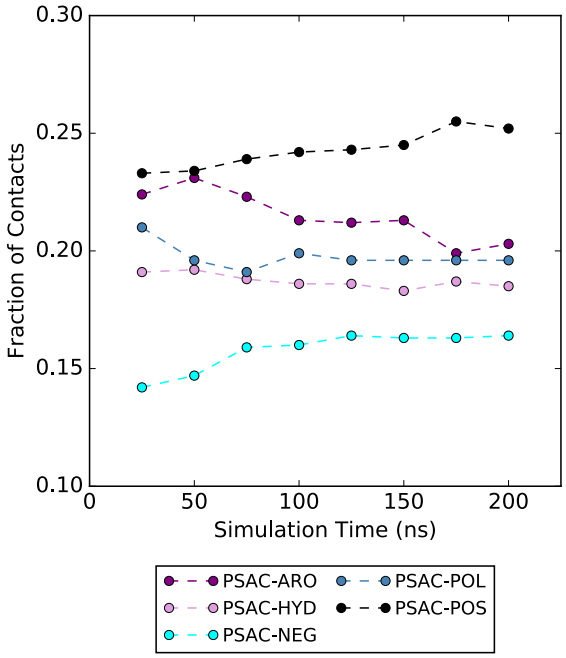

Figure A5.5. Convergence of amino acid contacts on the surface of BSA in an aqueous solution with 2 wt.% PSAC.

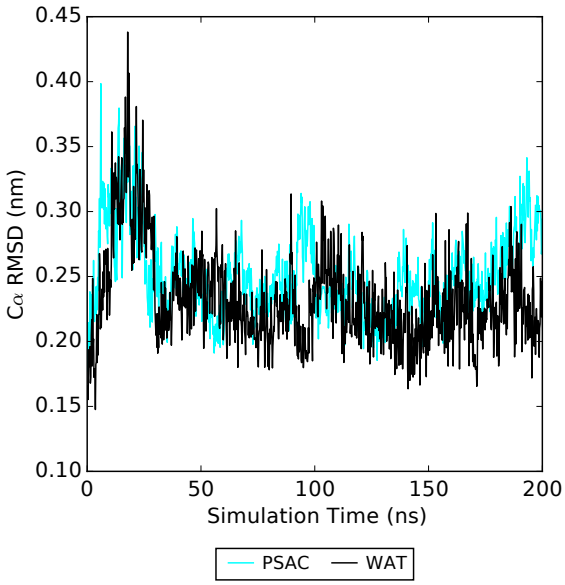

Figure A5.6. Root mean squared deviation (RMSD) of Cα atoms in BSA compared to the crystal structures in PSAC and in pure water.

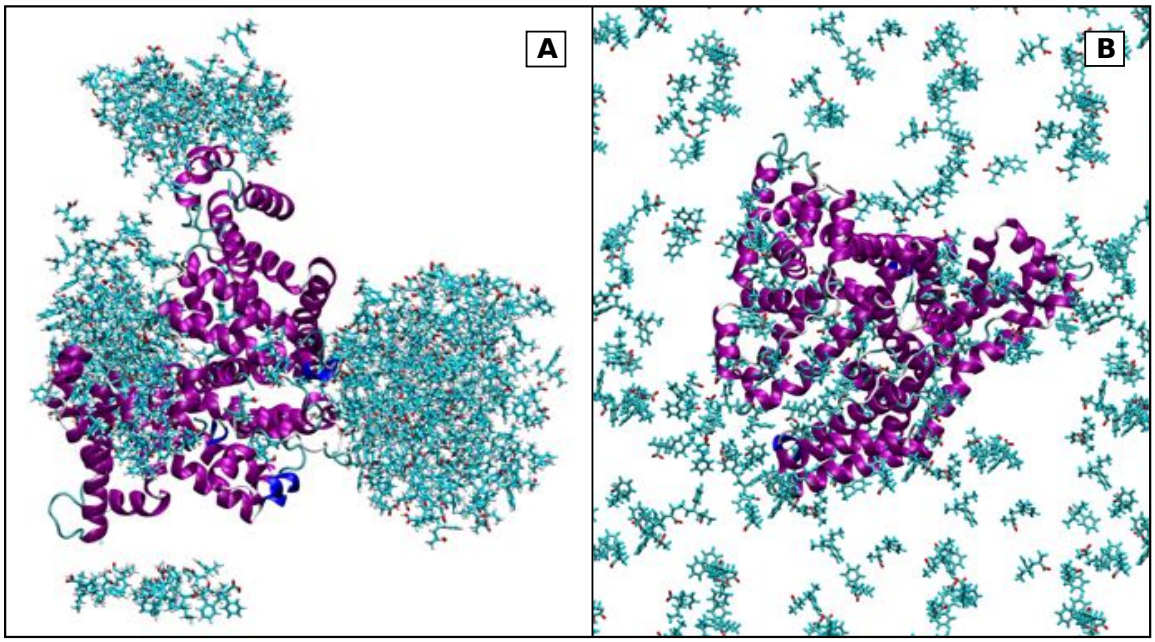

Figure A5.7. Snapshots from simulation trajectories of BSA in aqueous solutions with 2 wt.% PSAC, with: (A) PSAC molecules with protonated acrylate side chains, leading to self-aggregation, and (B) deprotonated acrylate side chains, approximating a solution at the intracellular pH of the brain of around 7.2.

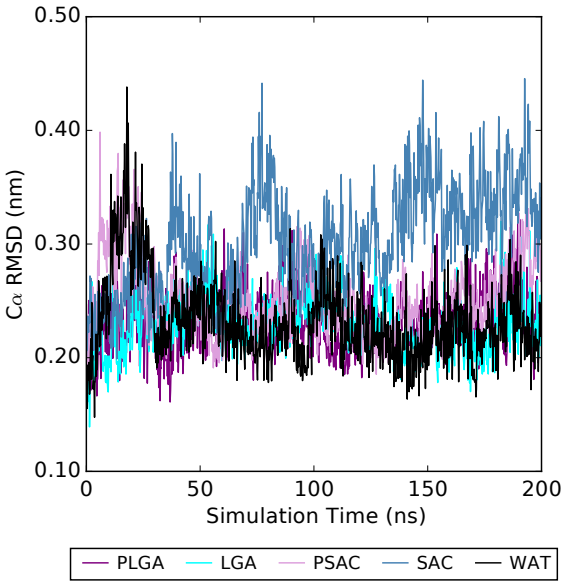

Figure A5.8. Root mean squared deviation (RMSD) of Cα atoms in BSA compared to the crystal structure in monomeric vs. trimeric polymer surrogate solutions and in pure water.

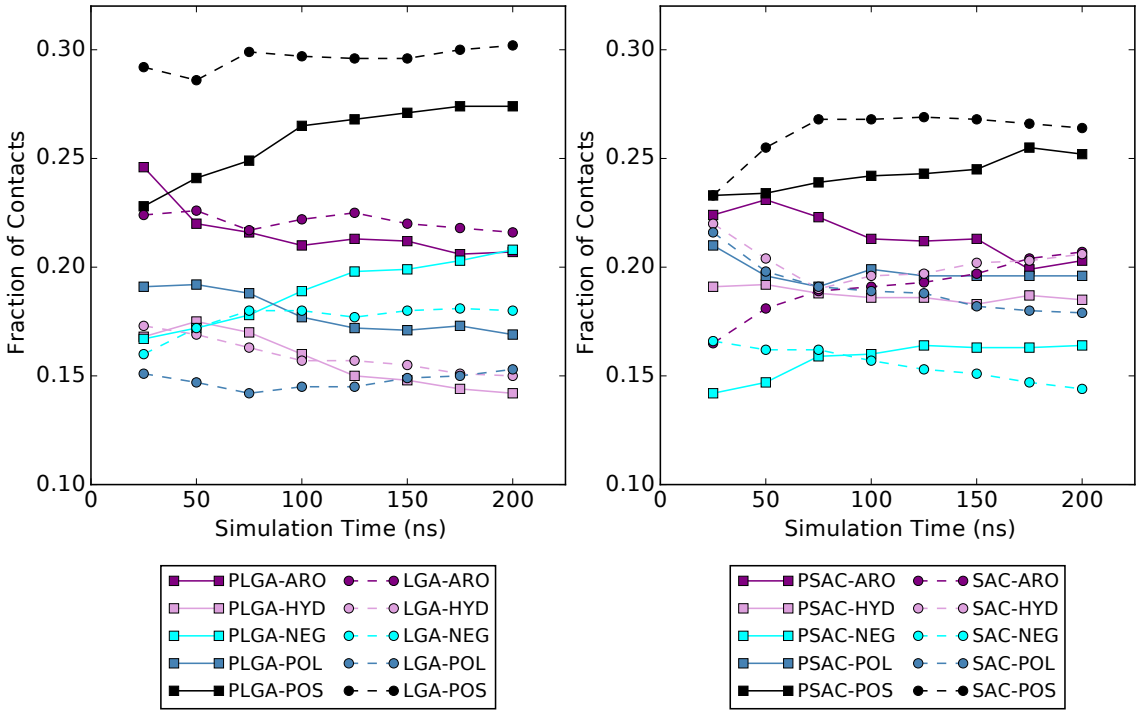

Figure A5.9. Convergence of amino acid contacts on BSA in (left) monomeric vs. trimeric LGA or PLGA copolymer, and (right) monomeric vs. trimeric SAC or PSAC copolymer, respectively.

Bibliography

1 Romei, F., *Leonardo Da Vinci*. The Oliver Press: 2008.
2 Benyus, J. M., *Biomimicry: Innovation Inspired by Nature*. HarperCollins: 1997.
3 Cogdell, R. J., Gardiner, A. T., Molina, P. I., et al. *Phil. Trans. R. Soc. A.*, 2013, **371**, A.N.: 20110603
4 Baio, J. E., Zane, A., Jaeger, V., et al. *JACS*, 2014, **136**, 15134.
5 Tanford, C., and Kirkwood, J. G. *JACS*, 1957, **79**, 5333.
6 Kauzmann, W. *Adv. Protein Chem.*, 1959, **14**, 1.
7 Prabhu, N., and Sharp, K. *Chem. Rev.*, 2006, **106**, 1616.
8 Nordwald, E. M., Plaks, J. G., Snell, J. R., et al. *ChemBioChem*, 2015, **16**, 2456.
9 Prime, K., and Whitesides, G. *Science*, 1991, **252**, 1164.
10 Latour, R. A. *Biointerphases*, 2008, **3**, FC2.
11 DeGrado, W. F., and Lear, J. D. *JACS*, 1985, **107**, 7684.
12 Long, J. R., Oyler, N., Drobny, G. P., et al. *J Am Chem Soc*, 2002, **124**, 6297.
13 Weidner, T., Samuel, N. T., McCrea, K., et al. *Biointerphases*, 2010, **5**, 9.
14 Weidner, T., Apte, J. S., Gamble, L. J., et al. *Langmuir*, 2010, **26**, 3433.
15 Mermut, O., Phillips, D. C., York, R. L., et al. *JACS*, 2006, **128**, 3598.
16 York, R. L., Browne, W. K., Geissler, P. L., et al. *Isr. J. Chem.*, 2007, **47**, 51.
17 York, R. L., Mermut, O., Phillips, D. C., et al. *J. Phys. Chem. C*, 2007, **111**, 8866.
18 Phillips, D. C., York, R. L., Mermut, O., et al. *J. Phys. Chem. C*, 2007, **111**, 255.
19 Apte, J. S., Collier, G., Latour, R. A., et al. *Langmuir* 2010, **26**, 3423.
20 Guiffo-Soh, G., Hernandez, B., Coic, Y. M., et al. *J. Phys. Chem. B*, 2007, **111**, 12563.
21 Zane, A. C., Michelet, C., Roehrich, A., et al. *Langmuir*, 2014, **30**, 7152.
22 Chouchane, K., Vendrely, C., Amari, M., et al. *J. Phys. Chem. B*, 2015, **119**, 10543.
23 Apte, J. S., Gamble, L. J., Castner, D. G., et al. *Biointerphases*, 2010, **5**, 97.
24 Goobes, R., Goobes, G., Campbell, C. T., et al. *Biochemistry*, 2006, **45**, 5576.
25 Wei, Y., Thyparambil, A. A., and Latour, R. A. *Chemphyschem*, 2012, **13**, 3782.
26 Thyparambil, A. A., Wei, Y., and Latour, R. A. *Langmuir*, 2012, **28**, 5687.
27 Wei, Y., and Latour, R. A. *Langmuir*, 2009, **25**, 5637.
28 Latour, R. A. *Colloids Surf. B. Biointerfaces*, 2014, **124**, 25.
29 Deighan, M., and Pfaendtner, J. *Langmuir*, 2013, **29**, 7999.
30 Marino, K. A., and Bolhuis, P. G. *J. Phys. Chem. B*, 2012, **116**, 11872.
31 O'Brien, C. P., Stuart, S. J., Bruce, D. A., et al. *Langmuir*, 2008, **24**, 14115.
32 Collier, G., Vellore, N. A., Yancey, J. A., et al. *Biointerphases*, 2012, **7**, 24.
33 Wright, L. B., and Walsh, T. R. *PCCP*, 2013, **15**, 4715.
34 Nimlos, M. R., Beckham, G. T., Matthews, J. F., et al. *J. Biol. Chem.*, 2012, **287**, 20603.
35 Battle, K., Alan Salter, E., Wesley Edmunds, R., et al. *J. Cryst. Growth*, 2010, **312**, 1257.
36 Grasso, G., Deriu, M. A., Prat, M., et al. *J. Phys. Chem. B*, 2015, **119**, 8239.
37 Brandt, E. G., and Lyubartsev, A. P. *J. Phys. Chem. C*, 2015, **119**, 18126.
38 Shen, J.-W., Wu, T., Wang, Q., et al. *Biomaterials*, 2008, **29**, 513.
39 Mijajlovic, M., Penna, M. J., and Biggs, M. J. *Langmuir*, 2013, **29**, 2919.
40 Schneider, J., and Colombi Ciacchi, L. *JACS*, 2012, **134**, 2407.
41 Nategholeslam, M., Gray, C. G., and Tomberli, B. *J. Phys. Chem. B*, 2014, **118**, 14203.
42 Laio, A., and Parrinello, M. *PNAS*, 2002, **99**, 12562.

43 Bussi, G., Gervasio, F. L., Laio, A., et al. *JACS*, 2006, **128**, 13435.
44 Bonomi, M., and Parrinello, M. *Phys. Rev. Lett.*, 2010, **104**, 190601.
45 Elder, R. M., Pfaendtner, J., and Jayaraman, A. *Biomacromolecules*, 2015, **16**, 1862.
46 Levine, Z. A., Fischer, S. A., Shea, J.-E., et al. *J. Phys. Chem. B*, 2015, **119**,10417.
47 Humphrey, W., Dalke, A., and Schulten, K. *J. Mol. Graphics*, 1996, **14**, 33.
48 Ulman, A., Eilers, J. E., and Tillman, N. *Langmuir*, 1989, **5**, 1147.
49 Fears, K. P., Creager, S. E., and Latour, R. A. *Langmuir*, 2008, **24**, 837.
50 Hess, B., Kutzner, C., van der Spoel, D., et al. *JCTC*, 2008, **4**, 435.
51 Lindorff-Larsen, K., Piana, S., Palmo, K., et al. *Proteins*, 2010, **78**, 1950.
52 Tribello, G. A., Bonomi, M., Branduardi, D., et al. *Comput. Phys. Commun.*, 2014, **185**, 604.
53 Essmann, U., Perera, L., Berkowitz, M. L., et al. *J. Chem. Phys.*, 1995, **103**, 8577.
54 Bussi, G., Donadio, D., and Parrinello, M. *J. Chem. Phys.*, 2007, **126**, 014101.
55 Deighan, M., Bonomi, M., and Pfaendtner, J. *JCTC*, 2012, **8**, 2189.
56 Sugita, Y., and Okamoto, Y. *Chem. Phys. Lett.*, 1999, **314**, 141.
57 Prakash, M. K., Barducci, A., and Parrinello, M. *JCTC*, 2011, **7**, 2025.
58 Barducci, A., Bussi, G., and Parrinello, M. *Phys. Rev. Lett.*, 2008, **100**, 020603.
59 Alonso, D. V., and Daggett, V. *Protein Sci.*, 1998, **7**, 860.
60 Bond, C. J., Wong, K.-B., Clarke, J., et al. *PNAS*, 1997, **94**, 13409.
61 Li, A., and Daggett, V. *PNAS*, 1994, **91**, 10430.
62 Daggett, V., and Levitt, M. *J. Mol. Biol.*, 1993, **232**, 600.
63 Tirado-Rives, J., and Jorgensen, W. L. *Biochemistry*, 1993, **32**, 4175.
64 Mark, A. E., and Van Gunsteren, W. F. *Biochemistry*, 1992, **31**, 7745.
65 Lindorff-Larsen, K., Piana, S., Dror, R. O., et al. *Science*, 2011, **334**, 517.
66 Piana, S., Lindorff-Larsen, K., and Shaw, D. E. *PNAS*, 2013, **110**, 5915.
67 Levy, Y., Jortner, J., and Becker, O. M. *PNAS*, 2001, **98**, 2188.
68 García, A. E., and Onuchic, J. N. *PNAS*, 2003, **100**, 13898.
69 Lindorff-Larsen, K., Maragakis, P., Piana, S., et al. *PLoS ONE*, 2012, **7**, e32131.
70 Piana, S., Lindorff-Larsen, K., and Shaw, David E. *Biophys. J.*, **100**, L47.
71 Zhai, Y., Laio, A., Tosatti, E., et al. *JACS*, 2011, **133**, 2535.
72 Deshapriya, I. K., and Kumar, C. V. *Langmuir*, 2013, **29**, 14001.
73 DeHaseth, P. L., Lohman, T. M., and Record, M. T. *Biochemistry*, 1977, **16**, 4783.
74 Vivcharuk, V., and Kaznessis, Y. *J. Phys. Chem. B*, 2010, **114**, 2790.
75 Bonomi, M., Barducci, A., and Parrinello, M. *J. Comput. Chem.*, 2009, **30**, 1615.
76 Sprenger, K. G., He, Y., and Pfaendtner, J., Probing How Defects in Self-Assembled Monolayers Affect Peptide Adsorption with Molecular Simulation. In *Foundations of Molecular Modeling and Simulation: Select Papers from Fomms 2015*, Snurr, R. Q.; Adjiman, C. S.; Kofke, D. A., Eds. Springer Singapore: Singapore, 2016; pp 21.
77 Carlsson, J., and Åqvist, J. *J. Phys. Chem. B*, 2005, **109**, 6448.
78 Xia, N., May, C. J., McArthur, S. L., et al. *Langmuir*, 2002, **18**, 4090.
79 Kollman, P. *Chem. Rev.*, 1993, **93**, 2395.
80 Dama, J. F., Parrinello, M., and Voth, G. A. *Phys. Rev. Lett.*, 2014, **112**, 240602.
81 Valsson, O., Tiwary, P., and Parrinello, M. *Annu. Rev. Phys. Chem.*, 2016, **67**, 159.
82 Branduardi, D., Bussi, G., and Parrinello, M. *JCTC*, 2012, **8**, 2247.
83 Raiteri, P., Alessandro, L., Luigi, G. F., et al. *J. Phys. Chem. B*, 2006, **110**, 3533.
84 Piana, S., and Laio, A. *J. Phys. Chem. B*, 2007, **111**, 4553.

85 Pfaendtner, J., and Massimiliano, B. *JCTC*, 2015, **11**, 5062.
86 Barducci, A., Bonomi, M., and Parrinello, M. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 826.
87 Limongelli, V., Bonomi, M., and Parrinello, M. *PNAS*, 2013, **110**, 6358.
88 Söderhjelm, P., Tribello, G. A., and Parrinello, M. *PNAS*, 2012, **109**, 5170.
89 Gervasio, F. L., Laio, A., and Parrinello, M. *JACS*, 2005, **127**, 2600.
90 Biarnés, X., Bongarzone, S., Vargiu, A. V., et al. *J. Comput. Aided Mol. Des.*, 2011, **25**, 395.
91 Sprenger, K. G., and Jim, P. *Langmuir*, 2016, **32**, 5690.
92 Wright, L. B., Palafox-Hernandez, J. P., Rodger, P. M., et al. *Chem. Sci.*, 2015, **6**, 5204.
93 Palafox-Hernandez, J. P., Tang, Z., Hughes, Z. E., et al. *Chem. Mater.*, 2014, **26**, 4960.
94 Wright, L. B., Freeman, C. L., and Walsh, T. R. *Mol. Simul.*, 2013, **39**, 1093.
95 Sultan, A. M., Hughes, Z. E., and Walsh, T. R. *Langmuir*, 2014, **30**, 13321.
96 Sultan, A. M., Westcott, Z. C., Hughes, Z. E., et al. *ACS Appl. Mater. Interfaces*, 2016, **8**, 18620.
97 Meißner, R. H., Schneider, J., Schiffels, P., et al. *Langmuir*, 2014, **30**, 3487.
98 Meißner, R. H., Wei, G., and Ciacchi, L. C. *Soft Matter*, 2015, **11**, 6254.
99 Hughes, Z. E., and Walsh, T. R. *J. Mater. Chem. B*, 2015, **3**, 3211.
100 Wu, P., Hu, X., and Yang, W. *J. Phys. Chem. Lett.*, 2011, **2**, 2099.
101 Mori, T., Hamers, R. J., Pedersen, J. A., et al. *JCTC*, 2013, **9**, 5059.
102 Quigley, D., Freeman, C. L., Harding, J. H., et al. *J. Chem. Phys.*, 2011, **134**, 044703.
103 Wu, C., Skelton, A. A., Chen, M., et al. *Langmuir*, 2012, **28**, 2799.
104 Abraham, M. J., Murtola, T., Schulz, R., et al. *SoftwareX*, 2015, **1–2**, 19.
105 Heinz, H. a. L. T.-J. a. *Langmuir*, 2013, **29**, 1754.
106 Huang, J., and MacKerell, A. D. *J. Comput. Chem.*, 2013, **34**, 2135.
107 Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., et al., Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*, Pullman, B., Ed. Springer Netherlands: Dordrecht, 1981; pp 331.
108 Emami, F. S., Puddu, V., Berry, R. J., et al. *Chem. Mater.*, 2014, **26**, 2647.
109 Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., et al. *J. Chem. Phys.*, 1984, **81**, 3684.
110 Hess, B., Bekker, H., Berendsen, H. J. C., et al. *J. Comput. Chem.*, 1997, **18**, 1463.
111 Darden, T., Darrin, Y., and Lee, P. *J. Chem. Phys.*, 1993, **98**, 10089.
112 Bussi, G., and Branduardi, D. *Rev. Comput. Chem.*, 2015, **28**, 1.
113 Tiwary, P., and Michele, P. *J. Phys. Chem. B*, 2015, **119**, 736.
114 Collins, K. D., and W., W. M. *Q. Rev. Biophys.*, 1985, **18**, 323.
115 Israelachvili, J. N., *Intermolecular and Surface Forces*. Academic press 2015.
116 Sumper, M., and Kroger, N. *J. Mater. Chem.*, 2004, **14**, 2059.
117 McDermott, C. A., McDermott, M. T., Green, J.-B., et al. *J. Phys. Chem.*, 1995, **99**, 13257.
118 Noh, J., and Hara, M. *Langmuir*, 2001, **17**, 7280.
119 Godin, M., Williams, P. J., Tabard-Cossa, V., et al. *Langmuir*, 2004, **20**, 7090.
120 Gannon, G., Greer, J. C., Larsson, J. A., et al. *ACS Nano*, 2010, **4**, 921.
121 Vemparala, S., Karki, B. B., Kalia, R. K., et al. *J. Chem. Phys.*, 2004, **121**, 4323.

122 Prathima, N., Harini, M., Rai, N., et al. *Langmuir*, 2005, **21**, 2364.
123 Jiang, L., Sangeeth, C. S. S., Yuan, L., et al. *Nano Lett.*, 2015.
124 O'Mahony, S., O'Dwyer, C., Nijhuis, C. A., et al. *Langmuir*, 2013, **29**, 7271.
125 Ahn, Y., Saha, J. K., Schatz, G. C., et al. *J. Phys. Chem. C*, 2011, **115**, 10668.
126 Barducci, A., Pfaendtner, J., and Bonomi, M. *Methods Mol Biol*, 2015, **1215**, 151.
127 Hansmann, U. H. E. *Chem. Phys. Lett.*, 1997, **281**, 140.
128 Daura, X., Gademann, K., Jaun, B., et al. *Angew. Chem. Int. Ed.*, 1999, **38**, 236.
129 Torrie, G. M., and Valleau, J. P. *J. Comput. Phys.*, 1977, **23**, 187.
130 Benedetto, A., and Ballone, P. *ACS Sustainable Chem. Eng.*, 2015, **4**, 392.
131 Fort, D. A., Remsing, R. C., Swatloski, R. P., et al. *Green Chem.*, 2007, **9**, 63.
132 Ha, S. H., Lan, M. N., Lee, S. H., et al. *Enzyme Microb. Technol.*, 2007, **41**, 480.
133 Guo, Z., and Xu, X. *Green Chem.*, 2006, **8**, 54.
134 Moniruzzaman, M., Nakashima, K., Kamiya, N., et al. *Biochem. Eng. J.*, 2010, **48**, 295.
135 Curto, V. F., Scheuermann, S., Owens, R. M., et al. *Phys. Chem. Chem. Phys.*, 2014, **16**, 1841.
136 Ghaedizadeh, S., Emamzadeh, R., Nazari, M., et al. *Biochem. Eng. J.*, 2016, **105,** 505.
137 Dabirmanesh, B., Daneshjou, S., Sepahi, A. A., et al. *Int. J. Biol. Macromol.*, 2011, **48**, 93.
138 Ghosh, S., Parui, S., Jana, B., et al. *J. Chem. Phys.*, 2015, **143**, 125103.
139 Ajloo, D., Sangian, M., Ghadamgahi, M., et al. *Int. J. Biol. Macromol.*, 2013, **55**, 47.
140 Jaeger, V. W., and Pfaendtner, J. *ACS Chem. Biol.*, 2013, **8**, 1179.
141 Nordwald, E. M., Armstrong, G. S., and Kaar, J. L. *ACS Catal.*, 2014, **4**, 4057.
142 Nordwald, E. M., and Kaar, J. L. *Biotechnol. Bioeng.*, 2013, **110**, 2352.
143 Nordwald, E. M., and Kaar, J. L. *J. Phys. Chem. B*, 2013, **117**, 8977.
144 Mai, N. L., and Koo, Y.-M. *Biochem. Eng. J.*, 2014, **87**, 33.
145 Kim, H. S., Ha, S. H., Sethaphong, L., et al. *Phys. Chem. Chem. Phys.*, 2014, **16**, 2944.
146 Liu, H., Zhu, L., Bocola, M., et al. *Green Chem.*, 2013, **15**, 1348.
147 Chen, Z., Pereira, J. H., Liu, H., et al. *PLoS ONE*, 2013, **8**, e79725.
148 Tee, K. L., Roccatano, D., Stolte, S., et al. *Green Chem.*, 2008, **10**, 117.
149 Carter, J. L. L., Bekhouche, M., Noiriel, A., et al. *ChemBioChem*, 2014, **15**, 2710.
150 Wolski, P. W., Dana, C. M., Clark, D. S., et al. *Protein Eng. Des. Sel.*, 2016, **29**, 117.
151 Micaêlo, N. M., and Soares, C. M. *J. Phys. Chem. B*, 2008, **112**, 2566.
152 Klähn, M., Lim, G. S., Seduraman, A., et al. *PCCP*, 2011, **13**, 1649.
153 Klähn, M., Lim, G. S., and Wu, P. *PCCP*, 2011, **13**, 18647.
154 Burney, P. R., Nordwald, E. M., Hickman, K., et al. *Proteins*, 2015, **83**, 670.
155 Jaeger, V., Burney, P., and Pfaendtner, J. *Biophys. J.*, 2015, **108**, 880.
156 Li, H., Kankaanpää, A., Xiong, H., et al. *Enzyme Microb. Technol.*, 2013, **53**, 414.
157 Burney, P. R., and Pfaendtner, J. *J. Phys. Chem. B*, 2013, **117**, 2662.
158 Latif, M., Alif, M., Micaelo, N. M., et al. *RSC Advances*, 2014, **4**, 48202.
159 Klähn, M., Stüber, C., Seduraman, A., et al. *J. Phys. Chem. B*, 2010, **114**, 2856.
160 Malde, A. K., Zuo, L., Breeze, M., et al. *JCTC*, 2011, **7**, 4026.
161 Schuttelkopf, A. W., and van Aalten, D. M. F. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, 2004, **60**, 1355.
162 Canongia Lopes, J. N., and Pádua, A. A. H. *J. Phys. Chem. B*, 2004, **108**, 16893.
163 Canongia Lopes, J. N., and Pádua, A. A. H. *J. Phys. Chem. B*, 2006, **110**, 19586.

164 Canongia Lopes, J. N., Pádua, A. A. H., and Shimizu, K. *J. Phys. Chem. B*, 2008, **112**, 5039.
165 Sambasivarao, S. V., and Acevedo, O. *JCTC*, 2009, **5**, 1038.
166 Sprenger, K. G., Jaeger, V. W., and Pfaendtner, J. *J. Phys. Chem. B*, 2015, **119**, 5882.
167 Hornak, V., Abel, R., Okur, A., et al. *Proteins*, 2006, **65**, 712.
168 Kirschner, K. N., Yongye, A. B., Tschampel, S. M., et al. *J. Comput. Chem.*, 2008, **29**, 622.
169 Jarin, Z., and Pfaendtner, J. *JCTC*, 2014, **10**, 507.
170 Vanommeslaeghe, K., and MacKerell, A. D. *J. Chem. Inf. Model.*, 2012, **52**, 3144.
171 Vanommeslaeghe, K., Raman, E. P., and MacKerell, A. D. *J. Chem. Inf. Model.*, 2012, **52**, 3155.
172 Zhang, Y., and Maginn, E. J. *J. Phys. Chem. B*, 2012, **116**, 10036.
173 Hurisso, B. B., Lovelock, K. R. J., and Licence, P. *PCCP*, 2011, **13**, 17737.
174 Men, S., Lovelock, K. R. J., and Licence, P. *PCCP*, 2011, **13**, 15244.
175 Frisch, M. J., Trucks, G. W., Schlegel, H. B., et al. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.
176 Wang, J., Wang, W., Kollman, P. A., et al. *J. Mol. Graphics Modell.*, 2006, **25**, 247.
177 Case, D. A., Berryman, J. T., Betz, R. M., et al. *Amber 2015*, University of California, San Francisco, 2015.
178 Martinez, L., Andrade, R., Birgin, E. G., et al. *J. Comput. Chem.*, 2009, **30**, 2157.
179 Sousa da Silva, A. W., and Vranken, W. F. *BMC Res. Notes*, 2012, **5**, 1.
180 Case, D. A., Cheatham, T. E., 3rd, Darden, T., et al. *J. Comput. Chem.*, 2005, **26**, 1668.
181 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., et al. *J. Chem. Phys.*, 1983, **79**, 926.
182 Wang, J. M., Wolf, R. M., Caldwell, J. W., et al. *J. Comput. Chem.*, 2004, **25**, 1157.
183 Tiwary, P., and Parrinello, M. *Phys. Rev. Lett.*, 2013, **111**, 230602.
184 Salvalaglio, M., Tiwary, P., and Parrinello, M. *JCTC*, 2014, **10**, 1420.
185 Fleming, K. L., Tiwary, P., and Pfaendtner, J. *J. Phys. Chem. A*, 2015.
186 Youngs, T. G. A., Holbrey, J. D., Mullan, C. L., et al. *Chem. Sci.*, 2011, **2**, 1594.
187 Cho, H. M., Gross, A. S., and Chu, J.-W. *JACS*, 2011, **133**, 14033.
188 Gupta, K. M., Hu, Z., and Jiang, J. *Polymer*, 2011, **52**, 5904.
189 Gross, A. S., Bell, A. T., and Chu, J.-W. *J. Phys. Chem. B*, 2011, **115**, 13433.
190 Mostofian, B., Smith, J., and Cheng, X. *Interdiscip. Sci.*, 2011, **3**, 308.
191 Youngs, T. G. A., Hardacre, C., and Holbrey, J. D. *J. Phys. Chem. B*, 2007, **111**, 13765.
192 Gross, A. S., Bell, A. T., and Chu, J.-W. *PCCP*, 2012, **14**, 8425.
193 Rabideau, B. D., Agarwal, A., and Ismail, A. E. *J. Phys. Chem. B*, 2013, **117**, 3469.
194 Rabideau, B. D., and Ismail, A. E. *PCCP*, 2015, **17**, 5767.
195 Zhao, Y., Liu, X., Wang, J., et al. *Carbohydr. Polym.*, 2013, **94**, 723.
196 Li, Y., Liu, X., Zhang, S., et al. *PCCP*, 2015, **17**, 17894.
197 Best, R. B., Buchete, N.-V., and Hummer, G. *Biophys. J.*, 2008, **95**, L07.
198 Shaw, D. E., Deneroff, M. M., Dror, R. O., et al. *Commun. ACM*, 2008, **51**, 91.
199 Sarkar, S., Majumder, T., Kalyanaraman, A., et al., Hardware Accelerators for Biocomputing: A Survey. In *ISCAS,* 2010; pp 3789.
200 Keskin, S., Kayrak-Talay, D., Akman, U., et al. *J. Supercrit. Fluids*, 2007, **43**, 150.
201 Mora-Pale, M., Meli, L., Doherty, T. V., et al. *Biotechnol. Bioeng.*, 2011, **108**, 1229.
202 Jorgensen, W. L., Maxwell, D. S., and TiradoRives, J. *JACS*, 1996, **118**, 11225.
203 Vanommeslaeghe, K., Hatcher, E., Acharya, C., et al. *J. Comput. Chem.*, 2010, **31**, 671.

204 Rappe, A. K., Casewit, C. J., Colwell, K. S., et al. *JACS*, 1992, **114**, 10024.
205 Liu, H., Sale, K. L., Simmons, B. A., et al. *J. Phys. Chem. B*, 2011, **115**, 10251.
206 Liu, H., and Maginn, E. *J. Chem. Phys.*, 2011, **135**, 124507.
207 Jiang, W., Yan, T., Wang, Y., et al. *J. Phys. Chem. B*, 2008, **112**, 3121.
208 Liu, H., Sale, K. L., Holmes, B. M., et al. *J. Phys. Chem. B*, 2010, **114**, 4293.
209 Youngs, T. G. A., and Hardacre, C. *Chemphyschem*, 2008, **9**, 1548.
210 Bhargava, B. L., and Balasubramanian, S. *J. Chem. Phys.*, 2007, **127**, 114510.
211 Sieffert, N., and Wipff, G. *J. Phys. Chem. B*, 2006, **110**, 13076.
212 Zhao, W., Eslami, H., Cavalcanti Welchy, L., et al. *Z. Phys. Chem.*, 2007, **221**, 1647.
213 Fiebig, O. C., Mancini, E., Caputo, G., et al. *J. Phys. Chem. B*, 2013, **118**, 406.
214 Yu, C.-Y., Wei, P., Li, X.-F., et al. *Ind. Eng. Chem. Res.*, 2014, **53**, 7923.
215 Frisch, M. J., Trucks, G. W., Schlegel, H. B., et al., Gaussian 09, Revision B.01. Wallingford CT, 2009.
216 Wang, J. W., W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.*, 2006, **25**.
217 Cornell, W. D., Cieplak, P., Bayly, C. I., et al. *JACS*, 1993, **115**, 9620.
218 Heaps, C. G., Long-Range Energy Alternatives Planning (Leap) System [Software Version 2014.0.1.20]. Stockholm Environment Institute: 2012.
219 Tenney, C. M., Massel, M., Mayes, J. M., et al. *J. Chem. Eng. Data*, 2014, **59**, 391.
220 Berendsen, H. J. C. v. d. S., D.; van Drunen, R. *Comp. Phys. Comm.* , 1995, **91**, 43.
221 Lindahl, E. H., B.; van der Spoel, D. *J. Mol. Mod.* , 2001, **7**, 306.
222 Parrinello, M., and Rahman, A. *J. Appl. Phys.*, 1981, **52**, 7182.
223 Gosling, E. M., McDonald, I. R., and Singer, K. *Mol. Phys.*, 1973, **26**, 1475.
224 Vega, C., and Abascal, J. L. F. *PCCP*, 2011, **13**, 19663.
225 Greaves, T. L., Weerawardena, A., Fong, C., et al. *J. Phys. Chem. B*, 2006, **110**, 22479.
226 Almeida, H. F. D., Passos, H., Lopes-da-Silva, J. A., et al. *J. Chem. Eng. Data*, 2012, **57**, 3005.
227 Araújo, J. M. M., Pereiro, A. B., Alves, F., et al. *J. Chem. Thermodyn.*, 2013, **57**, 1.
228 Shiflett, M. B., Harmer, M. A., Junk, C. P., et al. *J. Chem. Eng. Data*, 2006, **51**, 483.
229 Pereiro, A. B., Verdía, P., Tojo, E., et al. *J. Chem. Eng. Data*, 2007, **52**, 377.
230 Singh, T., and Kumar, A. *J. Solution Chem.*, 2009, **38**, 1043.
231 García-Miaja, G., Troncoso, J., and Romaní, L. *Fluid Phase Equilib.*, 2008, **274**, 59.
232 Domańska, U., Pobudkowska, A., and Wiśniewska, A. *J. Solution Chem.*, 2006, **35**, 311.
233 Iglesias-Otero, M. A., Troncoso, J., Carballo, E., et al. *J. Chem. Thermodyn.*, 2008, **40**, 949.
234 Harris, K. R., Kanakubo, M., and Woolf, L. A. *J. Chem. Eng. Data*, 2007, **52**, 1080.
235 Nieto de Castro, C. A., Langa, E., Morais, A. L., et al. *Fluid Phase Equilib.*, 2010, **294**, 157.
236 Krummen, M., Wasserscheid, P., and Gmehling, J. *J. Chem. Eng. Data*, 2002, **47**, 1411.
237 Vranes, M., Dozic, S., Djeric, V., et al. *J. Chem. Eng. Data*, 2012, **57**, 1072.
238 Troncoso, J., Cerdeiriña, C. A., Sanmamed, Y. A., et al. *J. Chem. Eng. Data*, 2006, **51**, 1856.
239 Freire, M. G., Teles, A. R. R., Rocha, M. A. A., et al. *J. Chem. Eng. Data*, 2011, **56**, 4813.
240 Fröba, A. P., Rausch, M. H., Krzeminski, K., et al. *Int. J. Thermophys.*, 2010, **31**, 2059.
241 Rosenboom, J.-G., Afzal, W., and Prausnitz, J. M. *J. Chem. Thermodyn.*, 2012, **47**, 320.
242 Schmidt, H., Stephan, M., Safarov, J., et al. *J. Chem. Thermodyn.*, 2012, **47**, 68.

243 Gaciño, F. M., Regueira, T., Lugo, L., et al. *J. Chem. Eng. Data*, 2011, **56**, 4984.
244 Seki, S., Tsuzuki, S., Hayamizu, K., et al. *J. Chem. Eng. Data*, 2012, **57**, 2211.
245 Gómez, E., González, B., Calvar, N., et al. *J. Chem. Eng. Data*, 2006, **51**, 2096.
246 Vercher, E., Llopis, F. J., González-Alfaro, V., et al. *J. Chem. Eng. Data*, 2011, **56**, 4499.
247 Olivier, E., Letcher, T. M., Naidoo, P., et al. *J. Chem. Thermodyn.*, 2010, **42**, 78.
248 Dzida, M., Chorążewski, M., Geppert-Rybczyńska, M., et al. *J. Chem. Eng. Data*, 2013, **58**, 1571.
249 González, B., Calvar, N., Gómez, E., et al. *J. Chem. Eng. Data*, 2008, **53**, 1824.
250 Calvar, N., Gómez, E., González, B., et al. *J. Chem. Eng. Data*, 2009, **54**, 2229.
251 González, E. J., Calvar, N., González, B., et al. *J. Chem. Thermodyn.*, 2009, **41**, 1215.
252 González, B., Corderí, S., and Santamaría, A. G. *J. Chem. Thermodyn.*, 2013, **60**, 9.
253 Gómez, E., González, B., Domínguez, Á., et al. *J. Chem. Eng. Data*, 2006, **51**, 696.
254 He, R.-H., Long, B.-W., Lu, Y.-Z., et al. *J. Chem. Eng. Data*, 2012, **57**, 2936.
255 Kenneth, R. S., Annegret, S., and María-José, T., Viscosity and Density of 1-Alkyl-3-Methylimidazolium Ionic Liquids. In *Clean Solvents*, American Chemical Society: Washington, DC, 2002.
256 Huddleston, J. G., Visser, A. E., Reichert, W. M., et al. *Green Chem.*, 2001, **3**, 156.
257 Widegren, J. A., and Magee, J. W. *J. Chem. Eng. Data*, 2007, **52**, 2331.
258 Kato, R., and Gmehling, J. *J. Chem. Thermodyn.*, 2005, **37**, 603.
259 Muhammad, A., Abdul Mutalib, M. I., Wilfred, C. D., et al. *J. Chem. Thermodyn.*, 2008, **40**, 1433.
260 McHale, G., Hardacre, C., Ge, R., et al. *Anal. Chem.*, 2008, **80**, 5806.
261 Oliveira, F. S., Freire, M. G., Carvalho, P. J., et al. *J. Chem. Eng. Data*, 2010, **55**, 4514.
262 Liu, Q.-S., Yang, M., Li, P.-P., et al. *J. Chem. Eng. Data*, 2011, **56**, 4094.
263 Arce, A., Rodríguez, O., and Soto, A. *J. Chem. Eng. Data*, 2004, **49**, 514.
264 David, W., Letcher, T. M., Ramjugernath, D., et al. *J. Chem. Thermodyn.*, 2003, **35**, 1335.
265 Nebig, S., and Gmehling, J. *Fluid Phase Equilib.*, 2010, **294**, 206.
266 Papaiconomou, N., Yakelis, N., Salminen, J., et al. *J. Chem. Eng. Data*, 2006, **51**, 1389.
267 Alonso, L., Arce, A., Francisco, M., et al. *J. Chem. Thermodyn.*, 2008, **40**, 265.
268 Tokuda, H., Tsuzuki, S., Susan, M. A. B. H., et al. *J. Phys. Chem. B*, 2006, **110**, 19593.
269 Liu, Q.-S., Yang, M., Yan, P.-F., et al. *J. Chem. Eng. Data*, 2010, **55**, 4928.
270 Damm, W., Frontera, A., Tirado–Rives, J., et al. *J. Comput. Chem.*, 1997, **18**, 1955.
271 Strechan, A. A., Paulechka, Y. U., Blokhin, A. V., et al. *J. Chem. Thermodyn.*, 2008, **40**, 632.
272 Rocha, M. A. A., Bastos, M., Coutinho, J. A. P., et al. *J. Chem. Thermodyn.*, 2012, **53**, 140.
273 Shimizu, Y., Ohte, Y., Yamamura, Y., et al. *Chem. Lett.*, 2007, **36**, 1484.
274 Zhang, Z.-H., Tan, Z.-C., Sun, L.-X., et al. *Thermochim. Acta*, 2006, **447**, 141.
275 Ge, R., Hardacre, C., Jacquemin, J., et al. *J. Chem. Eng. Data*, 2008, **53**, 2148.
276 García-Miaja, G., Troncoso, J., and Romaní, L. *J. Chem. Thermodyn.*, 2009, **41**, 161.
277 Paulechka, Y. U., Blokhin, A. V., Kabo, G. J., et al. *J. Chem. Thermodyn.*, 2007, **39**, 866.
278 Waliszewski, D., Stępniak, I., Piekarski, H., et al. *Thermochim. Acta*, 2005, **433**, 149.
279 Crosthwaite, J. M., Muldoon, M. J., Dixon, J. K., et al. *J. Chem. Thermodyn.*, 2005, **37**, 559.
280 Calvar, N., Gómez, E., Macedo, E. A., et al. *Thermochim. Acta*, 2013, **565**, 178.

281 Verevkin, S. P., Zaitsau, D. H., Emel'yanenko, V. N., et al. *Thermochim. Acta*, 2013, **562**, 84.
282 Blokhin, A. V., Paulechka, Y. U., and Kabo, G. J. *J. Chem. Eng. Data*, 2006, **51**, 1377.
283 Shimizu, Y., Ohte, Y., Yamamura, Y., et al. *J. Phys. Chem. B*, 2006, **110**, 13970.
284 Maginn, E. J. *Acc. Chem. Res.*, 2007, **40**, 1200.
285 Cadena, C., Zhao, Q., Snurr, R. Q., et al. *J. Phys. Chem. B*, 2006, **110**, 2821.
286 Earle, M. J., Esperanca, J. M. S. S., Gilea, M. A., et al. *Nature*, 2006, **439**, 831.
287 Santos, L. M. N. B. F., Canongia Lopes, J. N., Coutinho, J. A. P., et al. *JACS*, 2006, **129**, 284.
288 Kelkar, M. S., and Maginn, E. J. *J. Phys. Chem. B*, 2007, **111**, 9424.
289 Armstrong, J. P., Hurst, C., Jones, R. G., et al. *PCCP*, 2007, **9**, 982.
290 Zaitsau, D. H., Kabo, G. J., Strechan, A. A., et al. *J. Phys. Chem. A*, 2006, **110**, 7303.
291 Zaitsau, D., Yermalayeu, A., Emel'yanenko, V., et al. *Sci. China Chem.*, 2012, **55**, 1525.
292 Köddermann, T., Paschek, D., and Ludwig, R. *Chemphyschem*, 2007, **8**, 2464.
293 Tokuda, H., Hayamizu, K., Ishii, K., et al. *J. Phys. Chem. B*, 2005, **109**, 6103.
294 Canongia Lopes, J. N., Deschamps, J., and Pádua, A. A. H. *J. Phys. Chem. B*, 2004, **108**, 2038.
295 Tsuzuki, S., Shinoda, W., Saito, H., et al. *J. Phys. Chem. B*, 2009, **113**, 10641.
296 Hess, B. *J. Chem. Phys.*, 2002, **116**, 209.
297 Hu, Z., and Margulis, C. J. *J. Phys. Chem. B*, 2007, **111**, 4705.
298 Fendt, S., Padmanabhan, S., Blanch, H. W., et al. *J. Chem. Eng. Data*, 2011, **56**, 31.
299 Miran Beigi, A. A., Abdouss, M., Yousefi, M., et al. *J. Mol. Liq.*, 2013, **177**, 361.
300 Andreatta, A. E., Francisco, M., Rodil, E., et al. *Fluid Phase Equilib.*, 2011, **300**, 162.
301 Quijada-Maldonado, E., van der Boogaart, S., Lijbers, J. H., et al. *J. Chem. Thermodyn.*, 2012, **51**, 51.
302 Arce, A., Rodriguez, H., and Soto, A. *Green Chem.*, 2007, **9**, 247.
303 Calvar, N., Gómez, E., González, B., et al. *J. Chem. Eng. Data*, 2007, **52**, 2529.
304 González, B., Calvar, N., González, E., et al. *J. Chem. Eng. Data*, 2008, **53**, 881.
305 Rodríguez, H., and Brennecke, J. F. *J. Chem. Eng. Data*, 2006, **51**, 2145.
306 Schreiner, C., Zugmann, S., Hartl, R., et al. *J. Chem. Eng. Data*, 2009, **55**, 1784.
307 Yao, H., Zhang, S., Wang, J., et al. *J. Chem. Eng. Data*, 2012, **57**, 875.
308 Kandil, M. E., Marsh, K. N., and Goodwin, A. R. H. *J. Chem. Eng. Data*, 2007, **52**, 2382.
309 Ahosseini, A., and Scurto, A. *Int. J. Thermophys.*, 2008, **29**, 1222.
310 González, E. J., Alonso, L., and Domínguez, Á. *J. Chem. Eng. Data*, 2006, **51**, 1446.
311 Alonso, L., Arce, A., Francisco, M., et al. *J. Chem. Eng. Data*, 2008, **53**, 1750.
312 Andreatta, A. E., Arce, A., Rodil, E., et al. *J. Chem. Eng. Data*, 2009, **54**, 1022.
313 Vermaas, J. V., Crowley, M. F., Beckham, G. T., et al. *J. Phys. Chem. B*, 2015, **119**, 6129.
314 Beeson, W. T., Vu, V. V., Span, E. A., et al. *Annu. Rev. Biochem*, 2015, **84**, 923.
315 Vaaje-Kolstad, G., Westereng, B., Horn, S. J., et al. *Science*, 2010, **330**, 219.
316 Quinlan, R. J., Sweeney, M. D., Leggio, L. L., et al. *PNAS*, 2011, **108**, 15079.
317 Beeson, W. T., Phillips, C. M., Cate, J. H., et al. *JACS*, 2011, **134**, 890.
318 Kim, S., Ståhlberg, J., Sandgren, M., et al. *PNAS*, 2014, **111**, 149.
319 Hu, J., Chandra, R., Arantes, V., et al. *Bioresour. Technol.*, 2015, **186**, 149.
320 Arfi, Y., Shamshoum, M., Rogachev, I., et al. *PNAS*, 2014, **111**, 9109.
321 Cannella, D., Hsieh, C.-W., Felby, C., et al. *Biotechnol. Biofuels*, 2012, **5**, 26.

322 Dadi, A. P., Varanasi, S., and Schall, C. A. *Biotechnol. Bioeng.*, 2006, **95**, 904.
323 Heinze, T., Schwikal, K., and Barthel, S. *Macromol. Biosci.*, 2005, **5**, 520.
324 Sheldon, R. A., Lau, R. M., Sorgedrager, M. J., et al. *Green Chem.*, 2002, **4**, 147.
325 Nordwald, E. M., Brunecky, R., Himmel, M. E., et al. *Biotechnol. Bioeng.*, 2014, **111**, 1541.
326 Forsberg, Z., Mackenzie, A. K., Sørlie, M., et al. *PNAS*, 2014, **111**, 8446.
327 Wu, M., Beckham, G. T., Larsson, A. M., et al. *J. Biol. Chem.*, 2013, **288**, 12828.
328 Wise, O., and Coskuner, O. *J. Comput. Chem.*, 2014, **35**, 1278.
329 Sprenger, K. G., Choudhury, A., Kaar, J. L., et al. *J. Phys. Chem. B*, 2016, **120**, 3863.
330 Warner, L., Gjersing, E., Follett, S. E., et al. *Biochem. Biophys. Rep.*, 2016, **8**, 75.
331 Tung, H.-J., and Pfaendtner, J. *Mol. Syst. Des. Eng.*, 2016, **1**, 382.
332 Schrödinger, L. *The Pymol Molecular Graphics System, Version 1.8.*, 2015.
333 Sprenger, K. G., and Pfaendtner, J., Chapter Sixteen - Using Molecular Simulation to Study Biocatalysis in Ionic Liquids. In *Methods Enzymol.*, Gregory, A. V., Ed. Academic Press: 2016; Vol. Volume 577, pp 419.
334 Kaar, J. L., Jesionowski, A. M., Berberich, J. A., et al. *J Am Chem Soc*, 2003, **125**, 4125.
335 Hayes, R., Warr, G. G., and Atkin, R. *Chem. Rev.*, 2015, **115**, 6357.
336 Lozano, P., de Diego, T., Guegan, J. P., et al. *Biotechnol. Bioeng.*, 2001, **75**, 563.
337 Laszlo, J. A., and Compton, D. L. *Biotechnol. Bioeng.*, 2001, **75**, 181.
338 Lin Huang, J., Noss, M. E., Schmidt, K. M., et al. *Chem. Commun.*, 2011, **47**, 8007.
339 Attri, P., and Venkatesu, P. *PCCP*, 2011, **13**, 6566.
340 Zabet-Moghaddam, M., Krüger, R., Heinzle, E., et al. *J. Mass Spectrom.*, 2004, **39**, 1494.
341 Plaquevent, J.-C., Levillain, J., Guillen, F., et al. *Chem. Rev.*, 2008, **108**, 5035.
342 Tietze, A. A., Heimer, P., Stark, A., et al. *Molecules*, 2012, **17**, 4158.
343 Lesch, V., Heuer, A., Tatsis, V. A., et al. *PCCP*, 2015, **17**, 26049.
344 Yu, Y., Wang, J., Shao, Q., et al. *Sci. Rep.*, 2016, **6**, 19500.
345 Dalgicdir, C., Globisch, C., Peter, C., et al. *PLoS Comput. Biol.*, 2015, **11**, e1004328.
346 Dalgicdir, C., and Sayar, M. *J. Phys. Chem. B*, 2015, **119**, 15164.
347 Lutz, H., Jaeger, V., Berger, R., et al. *Adv. Mater. Interfaces*, 2015, **2**, 1500282.
348 Fu, L., Liu, J., and Yan, E. C. Y. *JACS*, 2011, **133**, 8094.
349 Pronk, S., Pall, S., Schulz, R., et al. *Bioinformatics*, 2013, **29**, 845.
350 Kabsch, W., and Sander, C. *Biopolymers*, 1983, **22**, 2577.
351 Donovan, M. A., Yimer, Y. Y., Pfaendtner, J., et al. *JACS*, 2016, **138**, 5226.
352 Einstein, A. *Annalen der Physik*, 1905, **322**, 549.
353 Bowers, J., Vergara-Gutierrez, M. C., and Webster, J. R. P. *Langmuir*, 2004, **20**, 309.
354 Smith, J., and Rick, S. W. *Condens. Matter Phys.*, 2016, **19**, 23002.
355 Kim, S., Kim, J. H., Lee, J. S., et al. *Small*, 2015, **11**, 3623.
356 Irvine, G. B., El-Agnaf, O. M., Shankar, G. M., et al. *Mol. Med.*, 2008, **14**, 451.
357 Wadhwani, P., Bürck, J., Strandberg, E., et al. *JACS*, 2008, **130**, 16515.
358 Wadhwani, P., Strandberg, E., Heidenreich, N., et al. *JACS*, 2012, **134**, 6512.
359 Fanghänel, S., Wadhwani, P., Strandberg, E., et al. *PLoS ONE*, 2014, **9**, e99653.
360 Sawyer, A. J., and Kyriakides, T. R. *Adv. Drug Deliv. Rev.*, 2016, **97**, 56.
361 Cohen, S., Yoshioka, T., Lucarelli, M., et al. *Pharm. Res.*, 1991, **8**, 713.
362 Peer, D., Karp, J. M., Hong, S., et al. *Nat. Nanotechnol.*, 2007, **2**, 751.
363 Moore, C. M., Akers, N. L., Hill, A. D., et al. *Biomacromolecules*, 2004, **5**, 1241.

364 González-Guerrero, M. J., del Campo, F. J., Esquivel, J. P., et al. *J. Power Sources*, 2016, **326**, 410.
365 Liu, Y., Du, J., Yan, M., et al. *Nat. Nanotechnol.*, 2013, **8**, 187.
366 Montalvo-Ortiz, B. L., Sosa, B., and Griebenow, K. *AAPS PharmSciTech*, 2012, **13**, 632.
367 Srinivas, G., Discher, D. E., and Klein, M. L. *Nat. Mater.*, 2004, **3**, 638.
368 Fukunaga, H., Takimoto, J., and Doi, M. *J. Chem. Phys.*, 2002, **116**, 8183.
369 Samanta, S., and Roccatano, D. *J. Phys. Chem. B*, 2013, **117**, 3250.
370 Ahmad, S., Johnston, B. F., Mackay, S. P., et al. *J. R. Soc. Interface*, 2010, **7**, S423.
371 Loverde, S. M., Ortiz, V., Kamien, R. D., et al. *Soft Matter*, 2010, **6**, 1419.
372 Moeendarbary, E., Ng, T. Y., and Zangeneh, M. *Int. J. Appl. Mech.*, 2010, **2**, 161.
373 Ortiz, V., Nielsen, S. O., Discher, D. E., et al. *J. Phys. Chem. B*, 2005, **109**, 17708.
374 Zambaux, M. F., Bonneaux, F., Gref, R., et al. *J. Controlled Release*, 1998, **50**, 31.
375 Maier, J. A., Martinez, C., Kasavajhala, K., et al. *JCTC*, 2015, **11**, 3696.
376 Casey, J. R., Grinstein, S., and Orlowski, J. *Nat. Rev. Mol. Cell Biol.*, 2010, **11**, 50.
377 Curtis, C., Zhang, M., Liao, R., et al. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.*, 2017, **9**.
378 Nance, E., Zhang, C., Shih, T. Y., et al. *ACS Nano*, 2014, **8**, 10655.
379 Nance, E., Zhang, F., Mishra, M. K., et al. *Biomaterials*, 2016, **101**, 96.
380 Nance, E. A., Woodworth, G. F., Sailor, K. A., et al. *Sci. Transl. Med.*, 2012, **4**, 149ra119.
381 Kuo, A. *CheM*, 2011, **1**, 80.
382 Liechty, W. B., Kryscio, D. R., Slaughter, B. V., et al. *Annu. Rev. Chem. Biomol. Eng.*, 2010, **1**, 149.
383 Van Rossum, G. In *Python Programming Language*, USENIX Annual Technical Conference, 2007; p 36.
384 Baker, J. L., Furbish, J., and Lindberg, G. E. *J. Mol. Graphics Modell.*, 2015, **62**, 202.
385 Borrell, K. L., Cancglin, C., Stinger, B. L., et al. *J. Phys. Chem. B*, 2017, **121**, 4823.
386 Schutt, T. C., Hegde, G. A., Bharadwaj, V. S., et al. *J. Phys. Chem. B*, 2017, **121**, 843.
387 Schutt, T. C., Bharadwaj, V. S., Hegde, G. A., et al. *PCCP*, 2016, **18**, 23715.
388 Daly, C. A., Berquist, E. J., Brinzer, T., et al. *J. Phys. Chem. B*, 2016, **120**, 12633.
389 Yang, H., Wang, X., Ma, Y., et al. *Catal. Sci. Tech.*, 2016, **6**, 7773.
390 Tanghe, M., Danneels, B., Last, M., et al. *Protein Eng. Des. Sel.*, 2017, **30**, 401.
391 Raigoza, A. F., Onyirioha, K., and Webb, L. J. *Appl. Surf. Sci.*, 2017, **396**, 1831.
392 Krause, K. D., Roy, S., and Hore, D. K. *Biointerphases*, 2017, **12**, 02D407.
393 Hughes, Z. E., Kochandra, R., and Walsh, T. R. *Langmuir*, 2017, **33**, 3742.