ABSTRACT

Title of Dissertation: Non Traditional Solvent Effect
On Protein Behavior

Pei-Yin Lee, Doctor of Philosophy, 2022

Dissertation Directed by: Associate Professor Silvina Matysiak
Department of Bioengineering

Protein preservation has been a long lasting research topic due to its importance in many bio-pharmaceutical applications. A ”cold chain” is a commonplace solution to protein preservation, which stores biochemical products at a refrigerated temperature. A big advantage of cold chain is that the storing process is straightforward, without many further processes before the use of stored bio-products. However, it can also experience malfunction of the cooling system and results in economic lost and health care crisis. Ionic liquids (ILs), as a type of non traditional solvents, consist only of ions and are reported to be a potential candidate to replace the use of cold chain. The advantages of ILs include low flammability, high conductivity and less toxicity compared to some organic solvents. The most interesting feature of ILs is their extremely large number of cation-anion combinations, that can be tailored for specific use according to different needs. This thesis aims to investigate specific mechanism behind how ILs modulate protein behavior, specifically, how ILs affect protein stability, activity, and aggregation. We approach the research questions through the lens of molecular dynamics (MD) simulations and complement

with experimental findings.

In the first part of the thesis we first investigate the effects of two imidazolium based ILs (1-ethyl-3-methylimidazolium ethylsulfate, $[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$ and 1-ethyl-3-methylimidazolium diethylphosphate, $[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$) on lysozyme stability and activity. We collaborate with an experiment group at the University of Massachusetts (Bermudez lab) to complement our simulation results. Both ILs are found to destabilize lysozyme stability. In addition, both the cation and anions lower the stability of lysozyme, but in a different fashion. $[\mathrm{EMIM}]^+$ interacts with an Arg-Trp-Arg bridge that is critical in lysozyme stability through π–π and cation–π interactions, leading to a local induced destabilization. On the other hand, both anions interact with the whole protein surface through short-range electrostatic interactions, with $[\mathrm{Et_2PO_4}]^-$ having a stronger effect than $[\mathrm{EtSO_4}]^-$. Lysozyme activity is also reduced by the presence of the two ILs, but can be recovered after rehydration. It is found that the protein-ligand complex is less stable in the presence of ILs. In addition, a dense cloud of $[\mathrm{EMIM}]^+$ is found in the vicinity of the lysozyme active site residues, possibly leading to a competition with the sugar ligand. A fast leaving of these $[\mathrm{EMIM}]^+$ is observed after rehydration, which explains the reappearance of the active site and the recover of lysozyme activity.

Although classical all-atom MD simulations can provide us with a great deal of microscopic information, they are often limited by the temporal-spatial scale of the simulated systems. For example, systems with high viscosity solvents or systems involving large number of atoms will be difficult to reach convergence for all-atom MD. In this case, coarse grained (CG) MD can come into play to achieve the desired time- and length- scales. The faster sampling obtained from CG MD is achieved by reducing the degree of freedom of the system and by removing local energetic barriers. In CG MD, similar atoms are grouped to functional groups and thus the free energy landscape is smoothen. We develop a novel CG MD named "Protein Model with Polarizability and Transferability (*ProMPT*)". The novelty of this model is the inclusion of the charged dummies that can result in change of dipoles. These dipoles can reflect the change of environments and thus allow the model to respond to different environmental stimulus. We validate *ProMPT* with several benchmark proteins: Trp-cage, Trpzip4, villin, ww-domain, and β-α-β. *ProMPT* is able to simulate folding-unfolding and secondary structure transformation with minimal constraints, which is not feasible with previous CG models. In addition, *ProMPT* can also reproduce the experimental results for the dimerization of glycophorin A (GpA) with different point mutations. Here we demonstrate the ability of the model to capture the change of conformational space caused by point mutation.

In the last part of this thesis, we combine *ProMPT* and an in-house CG IL model to study the effects of $[\text{TEA}]^+[\text{Ms}]^-$ on amyloid beta 16-22 (Aβ_{16-22}) aggregation. Aβ_{16-22} is the hydrophobic core region and is the smallest fragment of Aβ that can fibrilize. Aβ has been extensively linked to the pathogenesis of the Alzheimer's disease. $[\text{TEA}]^+[\text{Ms}]^-$ is reported to suppress the formation of β-sheets and induce helices at high concentration. From our results, both β-sheet content and the aggregate size decrease with the increase of IL concentration, which are in agreement with experiments. Aggregates can form in both water and IL, but with different morphologies. In water, a nice hydrophobic core involving Phe-Phe interactions can form as well as intact β-sheet contacts. In addition, a cross β-sandwich structure is also observed, as seen from previous literature. However, the same hydrophobic core can not persist in the presence of IL. Aggregate structures in IL are not stable over time due to the $[\text{TEA}]^+$-Phe interaction. Helicity is also computed for Aβ_{16-22} in water and in IL at different concentrations and a positive correlation is found. The increase in helicity at high $[\text{TEA}]^+[\text{Ms}]^-$ concentration can be explained by the reduction of the inter-peptide contacts, which then increases the opportunity for the peptides to form helical structures. Single peptide studies also reveal that $[TEA]^+[Ms]^-$ increases the helicity, possibly through cation-induced dipole enhancement.

In this thesis, a series of detailed investigations on the effects of ILs on protein behavior is performed. Specific interactions between IL functional groups and protein local/global structures are examined. The mechanisms we studied here will help constructing a holistic view for the design of IL-protein pair applications. The construction of the new CG protein/IL model provides another tool for the scientific community to study secondary structure transformation, folding-unfolding, and other biochemical processes that are sensitive to the environment with CG MD.

NON TRADITIONAL SOLVENT EFFECT ON PROTEIN BEHAVIOR

by

Pei-Yin Lee

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
- Professor Silvina Matysiak, Advisor
- Professor Jeffery Klauda, Committee chair
- Professor John Weeks
- Professor Pratyush Tiwary
- Professor Amy Karlsson

ACKNOWLEDGMENT

Pei-Yin Lee acknowledges financial support from the National Science Foundation under grant CBET-1760879 and to computational resources at the University of Maryland.

I would like to first thank my advisor Silvina, who has always been encouraging and providing me with insightful feedback. My first two years in the graduate school had been difficult, if it were not you, I could not have stayed at the program and continued my PhD journey. You are not only an inspiring advisor, but also a thoughtful mentor who really cares about your students. Thank you for your advising both on academic work and life problems. I also want to thank my collaborators Dr. Harry Bermudez and Dr. Onkar Singh. From our collaboration I have learned a lot on how to effectively communicate with researchers that are not in the same fields. I would like to give a special thanks to Harry for all the recommendation letters. Your support means a lot to me. I also want to thank all my committee members (Dr. John Weeks, Dr. Pratyush Tiwary, Dr. Jeffery Klauda and Dr. Amy Karlsson) for their fruitful feedback during and after my defense. Especially I want to thank Dr. John Weeks for his support which helps me a lot when searching for a postdoctoral position. I also want to thank Dr. Weeks for his kind life advice. Thank you, Abhilash, for all the fruitful discussions. It was a pleasant experience to collaborate with you. Your hard work also inspires me to be a delicate and responsible researcher. I would also like to thank my fellow group mates and friends for all the advice and discussion, because of you I am not alone during my PhD years: Riya, Suhas, Neha, Meenal, Yihang, Sun-Ting, Huan-Kuang, Wen-Chen, and Yi-Hwa. I also want to thank my friends in Taiwan and friends in the states for supporting me and always being there for me.

This PhD journey has been very challenging to me. Both my OCD and anxiety disorder have got much worse that I have to take medication and see therapist for four to five years. These mental conditions also affect my productivity significantly since my condition fluctuates frequently and I can only work when I am at a better status. I thought of quitting every semester, but then I just tell myself to hang in there for another semester. I am still not sure if that was a correct thing to do since this degree really takes up both my mental and body health conditions. The price is high but I do see self growing during these years. I get to know myself better. I know where my limit is at and I am learning to not to be that harsh on myself. I would like to thank my parents for their unconditional love and support. They are always there for me no matter what. I came to the states at the age of fifteen and since then I haven't been able to spend much time with my family. For that I want to say I'm sorry and thank you to my parents for understanding me. I also want to thank my husband for always taking care of me and supporting me when I have a mental breakdown. If it weren't you I would not be able to survive through this PhD study.

Table of Contents

Acknowledgements ii

Table of Contents iv

List of Abbreviations vi

List of Tables viii

List of Figures ix

Chapter 1: Introduction 1
- 1.1 Objective of thesis 1
  - 1.1.1 Protein behavior in ionic liquids (ILs) 3
  - 1.1.2 Molecular dynamics 8
- 1.2 Outline of thesis 13

Chapter 2: Evaluate the role of $[EMIM]^+[EtSO_4]^-$ and $[EMIM]^+[Et_2PO_4]^-$ on lysozyme stability and activity 16
- 2.1 Overview 16
- 2.2 Introduction 18
- 2.3 Methods 22
  - 2.3.1 Simulations 22
- 2.4 Results and discussion 27
  - 2.4.1 Evaluate the role of $[EMIM]^+[EtSO_4]^-$ and $[EMIM]^+[Et_2PO_4]^-$ on lysozyme stability 27
  - 2.4.2 Evaluate the role of $[EMIM]^+[EtSO_4]^-$ and $[EMIM]^+[Et_2PO_4]^-$ on lysozyme activity 36
- 2.5 Conclusion 45

Chapter 3: Transferable and polarizable coarse grained model for proteins (proMPT) 48
- 3.1 Overview 48
- 3.2 Introduction 50
- 3.3 Methods 56
  - 3.3.1 Non-Bonded Interactions 59
  - 3.3.2 Bonded Interactions 63
  - 3.3.3 Simulation setup 70

3.3.4 Comparison with replica exchange molecular dynamics simulations with solute tempering (REST) . . . 74
3.3.5 Analysis . . . 74
3.4 Results and discussion . . . 77
3.4.1 Mini-proteins . . . 77
3.4.2 Transmembrane protein Glycophorin A (GpA) . . . 90
3.5 Conclusion . . . 100

Chapter 4: Investigate the role of $[TEA]^+[Ms]^-$ on Abeta aggregation 102
4.1 Overview . . . 102
4.2 Introduction . . . 103
4.3 Methods . . . 105
4.3.1 $[TEA]^+[Ms]^-$ model . . . 105
4.3.2 $A\beta_{16-22}$ peptide model . . . 107
4.3.3 Simulation setup . . . 107
4.3.4 Analyses . . . 109
4.4 Results and discussion . . . 110
4.4.1 Validation for the coarse grained ionic liquid model . . . 110
4.4.2 The morphology is different for $A\beta_{16-22}$ aggregates in water and in IL at different concentrations . . . 112
4.4.3 Mechanism behind the β-sheet disruption . . . 115
4.4.4 Mechanism behind the change of helical fraction . . . 115
4.5 Conclusion . . . 118

Chapter 5: Conclusion and Outlook 120
5.1 Overview . . . 120
5.2 Future work . . . 125

Appendix A: Experimental results for evaluating the role of $[EMIM]^+[EtSO_4]^-$ and $[EMIM]^+[Et_2PO_4]^-$ on lysozyme stability and activity 128
A.1 Methods . . . 128
A.1.1 Materials . . . 128
A.1.2 Sample preparation . . . 128
A.1.3 Calorimetry . . . 129
A.1.4 Fluorescence Spectroscopy . . . 129
A.1.5 Activity assay . . . 130
A.1.6 Circular dichroism (CD) Spectroscopy . . . 131
A.2 Results and discussion . . . 132
A.2.1 Evaluate the role of $[EMIM]^+[EtSO_4]^-$ and $[EMIM]^+[Et_2PO_4]^-$ on lysozyme stability . . . 132
A.2.2 Evaluate the role of $[EMIM]^+[EtSO_4]^-$ and $[EMIM]^+[Et_2PO_4]^-$ on lysozyme activity . . . 136

Bibliography 141

List of Abbreviations

COVID-19	Coronavirus disease
IL	Ionic liquid
MD	Molecular dynamics
CG	Coarse grained
$[EMIM]^+[EtSO_4]^-$	1-ethyl-3-methylimidazolium ethylsulfate
$[EMIM]^+[Et_2PO_4]^-$	1-ethyl-3-methylimidazolium diethylphosphate
ProMPT	Transferrable and polarizable coarse grained model for proteins
$[TEA]^+[Ms]^-$	Triethylammonium mesylate
Aβ	Amyloid beta
Im7	ColE7 immunity protein
$[BMIM]^+$	1-butyl-3-methylimidazolium
$[C_4mpy]^+[Tf_2N]^-$	1-butyl-1-methylpyrrolidinium bis(trifluoromethylsulfonyl)imide
REMD	Replica exchange molecular dynamics
REST	Solute tempering in replica exchange
AA	All-atom
UA	United-atom
AWESEM	The associative memory, water-mediated, structure and energy model
OPEP	Optimized potential for efficient protein structure prediction
WEPCGM	Water-explicit polarizable coarse grained model
PMF	Potential mean force
RDF	Radial distribution function
DSC	Differential scanning calorimetry
UV-VIS-NIR	Ultraviolet–visible–near infrared spectroscopy
CD	Circular dichroism
PDB	Protein data bank
OPLS	Optimized potentials for liquid simulations
SPC	Simple point charge
NVT	Isochoric-isothermal ensemble
NPT	Isobaric-isothermal ensemble
LINCS	Linear constraint solver algorithm

PME	Particle-mesh Ewald method
rxn	Reaction coordinate
SDF	Spatial distribution function
VMD	Visual molecular dynamics
COM	Center of mass
WHAM	Weighted histogram analysis method
R_g	Radius of gyration
RMSD	Root mean squared distance
RMSF	Root mean squared fluctuation
LJ	Lennard-Jones
SR	Short range
vdW	van der Waals
GpA	Glycophorin A
FRET	Förster resonance energy transfer
WT	Wild-type
APP	Amyloid precursor protein
BB	Backbone bead
DPC	Dodecylphosphocholine
MBAR	Multistate Bennett acceptance ratio
DSSP	Hydrogen bond estimation algorithm
NMR	Nuclear magnetic resonance
WEPMEM	Water explicit polarizable membrane model
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
$[HSO_4]^-$	Hydrogen sulfate
$[H_2PO_4]^-$	Dihydrogenphosphate
$[EA]^+$	Ethylammonium
$[DEA]^+$	Diethylammonium

List of Tables

2.1 Number of water molecules, IL ions, and chloride ions for each system along with the dilution factors if applicable. 26
2.2 SR interaction between ILs and lysozyme. 35

3.1 Charges and characteristic bonded potentials for dummy beads. Bond-length is the length of the tether from the primary interaction-center to the charged dummies. k_{angle} is the spring constant preventing deviation of the angle between charged dummies and the primary-interaction-center from 180 degrees. 58
3.2 Bond lengths between primary CG interaction sites. 64
3.3 Bonded interaction potentials between different primary CG interaction-centers. BB: Backbone; S1: First Sidechain; S2: Second Sidechain. 64
3.4 Angular interaction potentials between different primary CG interaction-centers. BB: Backbone; S1: First Sidechain; S2/S3/S4: Second/Third/Fourth Sidechain. . 65
3.5 Dihedral interaction potentials between different primary CG interaction-centers. BB: Backbone; S1: First Sidechain; S2/S3/S4: Second/Third/Fourth Sidechain. . 65
3.6 Dihedral potential parameters for α-helix and 3-10 helix corresponding to equation 1 in the main text. 66
3.7 Simulation setup for each protein. [a] Sequence: KLVFFAE. 71

4.1 Simulation setup for each system. 108
4.2 Helical fraction for systems containing 8 peptides and for system containing a single peptide. 117

A.1 Binding constants obtained from Stern-Volmer and Lineweaver-Burk plots. . . . 136
A.2 Normalized lysozyme activity using the typical activity assay, under different conditions. Prior to rehydration all samples here showed no detectable activity, i.e., activity $= 0\%$. The dilution factor df $= 2.6 \times 10^5$. 138
A.3 Lysozyme melting temperature T_m under different conditions. All samples here are aged in their respective solvents for more than 1 year. The dilution factor df $= 100$. [a] DSC. [b] VP-DSC. 139

List of Figures

2.1 Molecular structures of (a) $[EMIM]^+$ cation, (b) $[EtSO_4]^-$ anion, and (c) $[Et_2PO_4]^-$ anion. White: hydrogen; cyan: carbon; blue: nitrogen; red: oxygen; yellow: sulfur; and tan: phosphorus. The marked atoms are for the identification purpose in Figure 2.7. 23

2.2 Fraction of native contacts, RMSD, and R_g time series for water (black), 25 wt% $[EMIM]^+[EtSO_4]^-$ (red), and 25 wt% $[EMIM]^+[Et_2PO_4]^-$ (blue) systems. The first column is at 300 K, the second column is at 370 K, and the third column is at 400 K. 29

2.3 RMSF of lysozyme Cα atoms in nm, where black represents water, red represents 25 wt% $[EMIM]^+[EtSO_4]^-$, and blue represents 25 wt% $[EMIM]^+[Et_2PO_4]^-$. . . . 29

2.4 PMF plot at 370 K, where the reaction coordinates used for the left column are RMSD and fraction of native contacts and the reaction coordinates used for the right column are RMSD and R_g. (a) and (d) are water system, (b) and (e) are 25 wt% $[EMIM]^+[EtSO_4]^-$ system, and (c) and (f) are 25 wt% $[EMIM]^+[Et_2PO_4]^-$ system. PMF is in RT unit. 30

2.5 The probability distribution of absolute cosine between the normal of $[EMIM]^+$ ring and the normal of Trp62 ring within a cutoff of 0.4 nm. The insets are SDF (200 ns–1 μs) of $[EMIM]^+$ near the average structure of Trp62 (200 ns–1 μs). (a) is in 25 wt% $[EMIM]^+[EtSO_4]^-$ and (b) is in 25 wt% $[EMIM]^+[Et_2PO_4]^-$. Trp62 is shown in purple CPK representation. 32

2.6 Ratio of ILs to H_2O near lysozyme and in bulk in (a) 25 wt% $[EMIM]^+[EtSO_4]^-$ and (b) 25 wt% $[EMIM]^+[Et_2PO_4]^-$ at 300 K. 33

2.7 RDF between selected atoms of ILs and the whole protein in (a) 25 wt% $[EMIM]^+[EtSO_4]^-$ and (b) 25 wt% $[EMIM]^+[Et_2PO_4]^-$. An illustration of the marked atoms is shown in Figure 2.1. 34

2.8 SDF of ILs near lysozyme where red represents $[EMIM]^+$ and cyan represents anions: $[EtSO_4]^-$ in (a), (b) and $[Et_2PO_4]^-$ in (c), (d). Trp62 is shown in purple CPK representation. Panels (b) and (d) are after a rotation of approximate 180 ° of (a) and (c), respectively. 37

2.9 RMSF of water (red), 50 wt% ILs (blue) and rehydrated (green) systems for (a) $[EMIM]^+[EtSO_4]^-$ and (b) $[EMIM]^+[Et_2PO_4]^-$. Blue: Region I; Orange: Region II; Yellow: Region III. (c) is the B factor recorded in 1AKI.pdb. 39

2.10 The alignment of the average structures from water (red), 50 wt% ILs (blue) and rehydrated (green) systems in $[EMIM]^+[EtSO_4]^-$ (left) and $[EMIM]^+[Et_2PO_4]^-$ (right). Pink vdW representations are the active site residues (residue 35 and 52). The high RMSF regions are also indicated. 40

2.11 Overlapping of the lysozyme-ligand complex with the average structure of lysozyme in 50 wt% $[EMIM]^+[EtSO_4]^-$ (left) and $[EMIM]^+[Et_2PO_4]^-$ (right). In both figures, the crystal structures are in gray, the average structures from simulations are in blue, active site residues are in orange vdW, TRP62/63 are in yellow vdW, non-native sugar ligand is in green vdW and $[EMIM^+]$ is in red SDF. The green arrows indicate another region near the active site that is also occupied by high $[EMIM^+]$ density. 41
2.12 Free energy profile ΔG(r) of ligand in water (red) and 50 wt% $[EMIM]^+[EtSO_4]^-$ (blue) for moving the ligand away from the lysozyme active site in the pulling direction. 42
2.13 Umbrella sampling simulations showing the PMF plot for the distance of an $[EMIM^+]$ molecule away from the lysozyme along the pulling direction in a hydrated system. 43
2.14 The time series showing the number of $[EMIM^+]$ ions near the non-native ligand binding site, where (a) is the 50 wt% $[EMIM]^+[EtSO_4]^-$ system, (b) is the rehydrated $[EMIM]^+[EtSO_4]^-$ system, (c) is the 50 wt% $[EMIM]^+[Et_2PO_4]^-$ system, and (d) is the rehydrated $[EMIM]^+[Et_2PO_4]^-$ system. 44
2.15 Illustation showing the close contact between an $[EMIM^+]$ molecule (red) and (a) the active site residues (green) (b) TRP62/63 residues (cyan). 45

3.1 Schematic geometries of CG amino-acids. The bead type is labeled on each bead. 57
3.2 CG geometry of polarizable Martini water. The partial charge on each dummy is labeled. 58
3.3 Non-bonded interactions between polar beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol. . 60
3.4 Non-bonded interactions between polar beads and hydrophobic beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol. 61
3.5 Non-bonded interactions between polar beads and charged/neutral beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol. 62
3.6 Non-bonded interactions between hydrophobic beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol. 63
3.7 Non-bonded interactions between hydrophobic beads and charged/neutral beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol. 64
3.8 Non-bonded interactions between charged/neutral beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol. 65
3.9 BB (previous amino acid)-BB-S1 tabulated angular potentials for each amino acid in kJ/mol. 67
3.10 BB (previous amino acid)-BB-S1 tabulated angular potentials for each amino acid in kJ/mol. 68

3.11 BB (previous amino acid)-BB-S1 tabulated angular potentials for each amino acid in kJ/mol. 69
3.12 Secondary-structure specific dihedral potential used in the CG forcefield. The tabulated potentials are fitted (for α-helix and 3-10 helix) or derieved (β-sheet) to capture maximum value in the dihedral probability distributions 70
3.13 Interaction between Gly backbone of one monomer with Val/Leu backbone of the other monomer before (blue) and after (orange) addition of CαH..O interaction. . 72
3.14 CG geometry of the DPC micelle. Apolar beads are in green, negatively charged bead is in red, and positively charged bead is in purple. The bead names are shown on the right. 73
3.15 The amino acid sequence for WT, T87L, and G79L. The mutation point is colored in red and the GxxxG motif is underlined. The residue index is also marked on the top. 74
3.16 An illustration for the crossing angle between two GpA monomers. 76
3.17 PMF for Trp-cage (a) with RMSD helix BB and native contact as the reaction coordinates. PMF for Trpzip4 (b) with RMSD BB and native contact as the reaction coordinates. Both PDB structure (left) and the representative structure (right) from our model for Trp-cage (c) and Trpzip4 (d) are shown. Blue indicates the specific secondary structure each protein exhibits. 79
3.18 The folded structure of Trp-cage showing the packing of Tyr (red), Trp (green), and Pro (cyan) residues. 79
3.19 Free energy plot for Trp-cage at 290 K from the REST simulations. The folded basin has a free energy of 2.82 kT and the unfolded basin has a free energy of 3.98 kT. The ΔG is estimated to be 1.17 kT (2.81 kJ/mol). 80
3.20 RMSD helix BB time series for Trp-cage at T*=0.82. 80
3.21 Native contact time series for Trp-cage at T*=0.82. 81
3.22 The folded structure of Trpzip4 with the packing of Trp residues shown. 82
3.23 Free energy plot for Trpzip4 at 290 K from the REST simulations. The folded basin has a free energy of 2.12 kT and the unfolded basin has a free energy of 4.00 kT. The ΔG is estimated to be 1.88 kT (4.53 kJ/mol). 82
3.24 RMSD BB time series for Trpzip4 at T*=0.82. 83
3.25 Native contact time series for Trpzip4 at T*=0.82. 83
3.26 PMF for villin with RMSD S1 and RMSD S2 as the reaction coordinates at T*=0.52. The representative structure for each basin is shown as insets. 85
3.27 PMF for villin with RMSD S1 and RMSD S2 as the reaction coordinates at T*=0.82. 85
3.28 RMSD S1 time series for villin at T*=0.82. 86
3.29 RMSD S2 time series for villin at T*=0.82. 86
3.30 RMSD BB time series for (a) WW-domain and (b) β-α-β at T*=0.52 (blue) and T*=0.82 (red). The PDB structure (left) and the representative structure (right) are shown in (c) and (d) for WW-domain and β-α-β, respectively. Blue indicates the specific secondary structure each protein exhibits. 87
3.31 Time series for the fraction of Aβ 16-22 peptide forming β-sheets. An illustration of β-sheets aggregation is shown in the inset with yellow representing Phe residues. 89

3.32 PMF for (a) WT, (b) T87L , and (c) G79L GpA with the average helical content and the number of BB contacts as the reaction coordinates. The criteria to determine the most populated basin is 0.95 to 1 (average helical content) and 9 to 18 (number of BB contacts), 0.95 to 1 (average helical content) and 14 to 22 (number of BB contacts), and 0.85 to 0.92 (average helical content) and 6 to 12 (number of BB contacts) for WT, T87L, and G79L, respectively. The representative conformation for each WT/mutant is shown in the insert figures. Color code: Thr (grey), Gly (green), Val (purple), Leu (orange). 92
3.33 The residue-residue contact maps for (a) WT, (b) T87L, and (c) G79L. 92
3.34 The average number of contacts between WT (blue)/ G79L (orange) and the (a) Choline bead (NC3); (b) hydrophobic bead (C3) of the DPC detergent. 94
3.35 Dihedral angle analysis for G79L, where the x-axis shows the dihedral angle index and the y-axis shows the dihedral angle. The median is shown in orange, the box includes the interquartile range, the bars show the positions of minimum and maximum, and the circles are the outliers. 96
3.36 The crossing angle distribution for WT (blue), T87L (green), and G79L (orange). The x-axis is the crossing angle in degree. The y-axis is the probability of observing a specific crossing angle for the WT or the mutant. 98

4.1 The CG mapping scheme for (a) $[TEA]^+[Ms]^-$ and (b) $A\beta_{16-22}$ peptide. On the left side of (a) shows the atomistic scheme for the IL and water, where nitrogen is in blue, carbon is in cyan, hydrogen is in white, sulfur is in yellow, and oxygen is in red. On the right side of (a) is the CG scheme for the IL where the bead types are marked. 106
4.2 The double well dihedral potential used for $A\beta_{16-22}$ simulations. The minima occur at both α helix and β-sheet. 108
4.3 RDF between atoms/beads for (a) atomistic simulation and (b) CG simulation for 25 wt% $[TEA]^+[Ms]^-$. The CG mapping scheme is shown below, where in atomistic representation: Nitrogen: blue; Carbon: cyan; Hydrogen: white; Sulfur: yellow; Oxygen: red. 111
4.4 Snapshots of 75 wt% $[TEA]^+[Ms]^-$ in (a) atomistic simulations and (b) CG simulations showing the characteristic nano-structure for high concentration IL/water mixture. 111
4.5 Aggregation size for $A\beta_{16-22}$ in (a) water and (b) 10 wt% (c) 50 wt% (d) 90 wt% $[TEA]^+[Ms]^-$. There are 8 peptides in each system. 113
4.6 (a) The fraction of β-sheet for $A\beta_{16-22}$ in water and $[TEA]^+[Ms]^-$ at various concentrations. Representative $A\beta_{16-22}$ cluster in (b) water and in (c) 50 wt% $[TEA]^+[Ms]^-$. BB are in pink and Phe residues are in yellow. 114
4.7 Phe-Phe contact distribution in water and in IL at different concentrations. 114
4.8 Coordinate number of $[TEA]^+$ for each residue for (a) monomers and (b) not monomers. 116
4.9 (a) $[TEA]^+$ coordinate number for each residue for a single $A\beta_{16-22}$ in 90 wt% IL. (b) A snapshot showing the interaction between $[TEA]^+$ and the two ends of the $A\beta_{16-22}$. Peptide backbone: pink; backbone negative dummy: red; backbone positive dummy: blue; Lys: green; Glu: orange; $[TEA]^+$: purple. 118

A.1 Schematic of activity assays. (a) ”Traditional” assay where samples are diluted with phosphate buffer to achieve required final enzyme concentration of 400 U/mL. (b) ”Modified” assay where samples are prepared and mixed at a constant composition, C_{ILB} refers to IL-buffer mixture from 5% to 17% IL. 131
A.2 Effect of IL concentration on lysozyme melting temperature T_m from calorimetry. For comparison, data for the effect of [Gdm][Cl] is also shown[1]. 133
A.3 Effect of IL concentration on the lysozyme (a) transition enthalpy ΔH and (b) effective transition enthalpy ΔH_v from the van’t Hoff equation. 134
A.4 (a) Stern-Volmer plot and (b) Lineweaver-Burk plot from fluorescence quenching of lysozyme with added IL. 135
A.5 Far-UV CD spectra of lysozyme. Freshly prepared sample in phosphate buffer (red), freshly prepared in 50 wt% [EMIM]$^+$[EtSO$_4$]$^-$ and rehydrated with water (green) The dilution factor df $= 1000$. 137
A.6 Normalized lysozyme activity as a function of [EMIM]$^+$[EtSO$_4$]$^-$ content using a modified activity assay to avoid dilution steps. The curve fit is a Boltzmann-type sigmodial function. 139

Chapter 1: Introduction

1.1 Objective of thesis

The maintenance and extension of the thermo-stability and activity of proteins constitute an important goal in biotechnology. Once a protein has lost its native structure, it ceases to be useful. The storage and transportation of protein therapeutics such as vaccines and antibodies require a fixed temperature window referred to as the "cold chain" [2]. Maintaining the cold chain requires significant cost, and failure to maintain the system will result in both economic loss and public health crisis. Especially during the Coronavirus disease (COVID-19) pandemic era, the needs for a more efficient protein preservation technique become higher than ever. Some other protein preservation methods without directly modifying the proteins have also been proposed: freeze-drying, powdering, sterilization by γ-irradiation, immobilization of viral particles onto carbohydrate glass material, addition of sugars, encapsulation in silica, and silk stabilization. [3, 4, 5, 6, 7, 8, 9, 10, 11, 12] In addition, ionic liquids (ILs)/water mixtures have also been shown to be able to extend the thermal protein stability window. These non-traditional solvents carrying many advantages compared to organic solvents are thus emerging to be promising candidates for cold chain substitution and are the focus of this thesis.

ILs are liquids consisting mostly of ions and are generally referred to those that have melting points lower than 100° C. The advantages of ILs include the high chemical and thermal stability, negligible vapor pressure, high ionic conductivity, and lower toxicity than common vaccine formulation additives. The most interesting feature is their designable nature due to the vast number of cation and anion choices that allows us to tune the combination based on availability, cost, toxicity, or other functions. Previous studies have demonstrated both the stabilizing and destabilizing ability of ILs on proteins [13, 14, 15, 16, 17], and it is still not fully understood why some ILs could strengthen protein stability/activity while others do not. The goal of this thesis is to use a combination of experiments and simulations to understand the mechanism behind the change of protein behavior caused by ILs. Since the possible IL-protein pairs are an astronomical number, it is impossible to screen each pair to find the best combination for a specific application. A general rule will be useful for an efficient IL application design. The understanding we obtain from specific IL-pair interaction is thus important to help us build a common rule of thumb for choosing adequate ILs.

While experimental studies could provide a wealth of knowledge on the effects of ILs on proteins, for example, the change of secondary structure, difference in hydrophobic environment, shift in denature temperature, and lost in activity, some microscopic events can not be probed in details with only experiments. Here we apply molecular dynamics (MD) simulations to complement the findings observed by our collaborators at the University of Massachusetts (Bermudez lab). MD is a computational technique that can be used to explore both the structure and stability of proteins. Based on numerical integration of Newton's equations, MD is able to simulate the time evolution of molecules with specific force field that describes potential function from atomic positions. Due to the high viscosity of ILs, an atomistic representation of a protein in a high IL concentration mixture with MD will take extraordinary time to reach convergence, which is not realistic and limits the exploration of different IL combinations and protein sizes. The coarse-grained (CG) MD simulations are therefore applied to ease the burden in sampling. CG MD groups similar atoms into functional sites and by doing this, long length and timescales of interest can be explored with reduced system size and lower computational complexity.

In my thesis, I focus on characterizing which molecular interactions between proteins and ILs would lead to changes in protein behavior (folding kinetics/transitions, activity, and aggregation). Validating with experimental observations, I look at the mechanisms behind these changes. What we learn from each IL-protein pair interaction could help us construct a general picture to design useful ILs for particular applications. Because of the high viscosity of ILs and large temporal-spatial scale of certain systems, I also participate in the development of an in-house CG MD models for proteins and ILs in order to study more complex systems. My thesis will contain three primary objectives:

(1) Investigate the effects of 1-ethyl-3-methylimidazolium ethylsulfate ($[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$) and 1-ethyl-3-methylimidazolium diethylphosphate ($[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$) on the stability and activity of the hen egg white lysozyme

(2) Develop a CG-MD model for proteins (Transferable and polarizable CG model for proteins (*ProMPT*))

(3) Study the effects of triethylammonium mesylate ($[\mathrm{TEA}]^+[\mathrm{Ms}]^-$) on amyloid beta (A$\beta$) aggregation

1.1.1 Protein behavior in ionic liquids (ILs)

Proteins are essential in many biological events such as reaction catalysis, DNA replication, stimuli response and molecular transport. They are also the building block of the body, the main component of muscles, bones, organs and skins. Due to their importance in our daily life, modulation of protein behavior has drawn huge interest from many researchers, and one of the focuses is on the tuning of protein stability and activity. One direct benefit from protein stability and activity modulations is the break down of the expensive ”cold chain”. The cold chain is a common method for protein storage. The storage time for proteins before applications can range from weeks to years. During this extended time, it is important to devoid all possible inactivation processes for proteins [18]. Cold chain is able to keep proteins at an adequate temperature but requires significant cost and human resource to maintain. The use of ILs for the preservation of proteins has been reported in many literature works before [19, 20, 21], and thus makes them a promising candidate to replace cold chain and possibly bring down the risk of protein dysfunction.

ILs are room temperature liquids composed of purely cations and anions. They have drawn huge interest from researchers because of their potential of replacing organic solvents as novel green solvents. ILs can be categorized to four main classes based on their cation type: (1) alkylammonium; (2) dialkylimidazolium; (3) phosphonium; and (4) N-alkylpyridinium [22]. The cations are usually bulky and many of them consist of nitrogen atoms. As for anions, they are usually smaller in size or they can even be just a halide ion. Ethylammonium nitrate, one of the earliest ILs, can be dated back to 1914. Since 2000, ILs have been applied to be solvents for proteins [23]. Their advantages include but are not limited to low flammability, high thermo-stability, high ionic conductivity and good solvent property. Some ILs also show high bio-compatibility or strong ability of enhancing bio-molecule solubility and thus can be potentially used in vaccine formulation [24, 25, 26]. Moreover, the signature property of ILs is their tunability. Enormous combinations can be generated from vast choices of cations and anions. This property gives freedom for researchers to investigate ILs and their applications based on availability, cost and other considerations. In the introduction I will discuss the effects of ILs on protein behaviors, with special focuses on protein stability, activity, and aggregation.

1.1.1.1 Modulation of protein stability by ILs

Protein behaviors in IL/water mixtures have been extensively studied with different globular proteins such as lysozyme [27, 28], ribonuclease A [17, 29], cytochrome c [19, 30], hemoglobin [31], and lipases [32, 33], but no general trend has been established regarding how ILs modulate protein stability. In addition, in many studies the IL concentration could be held quite low. With the goal to enhance ambient and long-term protein stability, the effectiveness of these dilute conditions becomes questionable. Few studies have been carried out with significant IL content. For example, Mann et al. reported that ammonium formate based ILs could promote lysozyme refolding after thermal denaturation [21]. Lysozyme stability is also reported to be increased with increasing choline dihydrogenphosphate concentration [34]. Weingartner showed that with native RNase A, choline based ILs enhance stability, while imidazolium based ILs destabilize the protein [35]. While some studies focus more on the effects of cations, Rodrigues et al. found that with the same cation (choline in this case), different anions could enhance or diminish lysozyme stability [36]. In ColE7 immunity protein (Im7), anion effects are also reported to be dominant, although both cations ($[\text{EMIM}]^+$ and 1-butyl-3-methylimidazolium ($[\text{BMIM}]^-$)) and anions play roles in the destabilization in this case [37]. In general, imidazolium based cations favor destabilization unless the anion effect overplays the cation effect. From studies on more complex proteins that have a tertiary structure, we are able to investigate the effects of ILs on local regions (for example the active site) or critical sites that are important to stability. However, since these proteins are composed of different secondary structures and some special local sites (for example, the active site and native/non-native binding sites), the stability could be affected by multiple factors. Therefore, proteins and peptides with characteristic secondary structures become interesting targets to investigate the effect of ILs because it is easier to characterize the IL effects on smaller systems, to uncouple the effects of ILs on secondary structure from tertiary structure, and then extend the observed rules to more complex proteins. α-helix [AKA_2] was observed to be stabilized by 1-butyl-1-methylpyrrolidinium bis(trifluoromethylsulfonyl)imide ($[C_4mpy]^+[Tf_2N]^-$), while the same IL is observed to destabilize Trpzip4, which is a β-hairpin [38]. All of these studies have paved a road to the systematic design for IL applications in protein stability modulation. The effect of ILs on protein could be protein dependent, IL dependent, or sometimes, even IL concentration dependent. Both the role of cation and anion has to be studied thoroughly in order to understand their specific role at the molecular level.

1.1.1.2 Modulation of protein activity by ILs

An important role of proteins is their activity as an enzyme, which can catalyze critical biological reactions. The protein activity can be affected by the stability of the protein, the status of the active site (whether it is blocked or not), or the protein structural fluctuation. All of these factors could be modulated by ILs. Yu et al. reported that with low concentration of imidazolium based ILs, lysozyme is more stable and there is no change in activity [39]. However, when the IL concentration increases, the protein denatures and loses activity. The same loss of activity is reported for xylanase also in high concentration of imidazolium ILs, but unlike the protein denaturation that causes activity loss in the previous case, in the later case, it is the dampening of dynamic motion and kinetic trapping of cations in the binding pocket that reduce activity [40]. On the other hand, high concentration imidazolium ILs could also cause little loss of activity as for hyperthermophilic enzymes [41] and papain [42]. From the literature it is reported that the same category of ILs could lead to activity gain or loss, depending on the protein type, or sometimes, even depending on the IL concentrations. Therefore, the prediction of how one specific IL could modulate the activity of a protein could be complicated and therefore extensive case studies are necessary to investigate the possible mechanisms affecting protein activity in order to make precise predictions. Having in mind the ultimate goal is to ”break the cold chain”, the ability of protein to maintain or regain its activity after rehydration from long-time preservation in ILs is also critical when investigating the modulation of protein activity by ILs. This direction of study is significantly less explored and will require more experimental and simulation inputs in the future.

1.1.1.3 Modulation of protein aggregation by ILs

Protein aggregation is another important protein behavior to investigate to improve long-time protein storage. In order to efficiently preserve proteins, it is critical to suppress protein aggregation during storage, where protein aggregation would lead to protein dysfuncton. Protein aggregation such as amyloid plaques and neurofibrillary tangles are also signature pathogenesis for pathological diseases such as Alzheimer’s disease, Huntington’s disease, and Creutzfeld-Jacob disease [43]. These diseases will lead to cognitive decline. Amyloid cascade hypothesis, as a primary model for the Alzheimer’s disease, has been widely accepted [44, 45]. Based on this hypothesis, a cascade of events that results in dementia and neuronal dysfunction is initiated by the oligomerization of Aβ peptides. In addition, neurotoxic species is considered to be some soluble oligomers that form during the initial stage of Aβ aggregation. Therefore, it is important to study the origin of the oligomerization process in order to modulate protein aggregation. Another important reason to study protein aggregation is related to biotechnological applications, where protein aggregation hampers the efficiency of the process [46, 47]. It has been reported by several studies that ILs could be used to suppress or enhance aggregation, depending on the specific protein-IL pair. For ribonuclease A, ILs are reported to suppress aggregation, irrespective of their effects on protein stabilization/destabilization [35]. Polar ILs (pILs) can either promote or inhibit amyloid fibrilization for lysozyme [48, 49]. For Aβ peptides, pILs with kosmotopic anions (phosphate or sulphate) can promote fibrilization, while pILs with mesylate anion suppress aggregation [50]. $[\text{BMIM}]^+$[bis(trifluoromethylsulfonyl)imide]$^-$ is reported to accelerate fibrilization for α-synuclein [51].

1.1.2 Molecular dynamics

There are many properties for proteins in ILs that researchers can investigate with wet-lab experiments, for example, protein melting temperature, binding constants between solvents and specific residues, secondary structures, and protein activity, to name a few. However, there are also some disadvantages of wet-lab experiments such as difficulty in isolating individual effects from complex biological systems, inability to look at certain solvent-protein interactions, which are especially crucial for protein-IL studies, and complication in environmental control. MD simulations are uniquely suited to tackle these problems. Based on numerical integration of Newton's equations, MD is able to simulate the time evolution of molecules with specific force field that describes potential function from atomic positions. Through statistical mechanics, microscopic level simulations can be connected to macroscopic observables. This allows MD to contribute significantly toward understanding experimental data with molecular insights (how ILs interact with proteins) and help designing suitable IL-protein pairs for a specific aim.

Due to the size of the protein studied and the high viscosity of ILs, protein in IL could be difficult to simulate with classical MD, where convergence can not be reached within reasonable simulation time. Even for smaller proteins or peptides, limited conformational sampling that stems from slow barrier in rugged energy landscape and much shorter timescale compared to biomolecular processes could also make simulations with just the use of classical MD quite challenging [52]. Herein, enhanced sampling techniques could thus take place to alleviate these convergence problems. These enhanced sampling techniques include but are not limited to replica exchange MD (REMD) (partial REMD, local REMD [53], dual-solution REMD [54], and replica exchange with solute tempering (REST),...etc.), parallel tempering in the well-tempered ensemble [55], classic simulated annealing, umbrella sampling, torsion angle MD and metadynamics. In standard REMD, the number of replicas used increases with the square root of solution's total number of degree of freedom. Since the overall Hamiltonian grows with system size, in order to maintain an adequate acceptance ratio between each neighboring replicas, more replicas will be needed for larger systems. REST is thus developed to target this drawback by modifying the Hamiltonian so that the acceptance probability for replica exchange scales only with the number of degree of freedom for the biomolecule studied, not for the solvent [56]. In the original paper for REST, alanine dipeptide is used as a test system and the simulation is proved to be more efficient than traditional REMD. Even with many advantages that enhanced sampling techniques can deliver, there is still one important drawback, that is, in many cases the dynamics information could be lost. Therefore, it is to the researcher's decision to determine which method to apply. Here, CG-MD thus emerges to be a good alternative to study change of protein behavior in highly viscous ILs.

1.1.2.1 Coarse grained protein model

While enhanced sampling techniques could take place to alleviate the convergence problems in classical MD, there are still some drawbacks including the loss of dynamics and expensive computational costs. Herein, spatial CG can help alleviate the sampling convergence issue in MD simulations. Unlike all atom (AA) and united-atom (UA) MD, CG-MD takes the local averaging of atomic positions to reduce the number of degree of freedom. By doing this, a free-energy landscape can be smoothed and the computation efficiency can be significantly increased. CG is especially suited for protein folding, membrane remodeling and peptide aggregation studies where temporal and spatial scales are large. Some example force fields include PRIME-20 [57], AWESEM [58], OPEP [59], and MARTINI [60]. Some CG force fields are transferable, which means that the parametrization of these force fields aims for a general use that the users do not have to re-parametrize the parameters for every new system. Taking the popular transferable CG model MARTINI as an example, each amino acid is parametrized by matching the free energy of the side chain molecule to the experimental values. These amino acids are then used as building blocks to build protein model based on only primary sequence. In this way, models can be built for various proteins without the need of re-parametrization. Although the advantage is that the force field is independent of the studied systems and can therefore be directly applied to the systems of interest, it is still limited by the biomolecules it can represent and the ability to capture secondary and tertiary structure based solely on the primary sequence. That is, MARTINI is not able to fold proteins from a random coil or an extended structure. In order to study secondary structures with MARTINI, constraints have to be applied to maintain specific secondary structures at all times.

Our lab has previously developed an in-house CG model named Water-Explicit Polarizable Coarse Grained Model (WEPCGM). This model has its root in MARTINI and uses Yesylevskyy's polarizable water model [61]. The novelty of WEPCGM is the implementation of peptide backbone structural polarization. The idea is that the formation of protein backbone hydrogen bonds is critical in generating particular orientation of dipoles in secondary structures, for example, the CO-NH dipoles in an α-helix are parallel to the helix major axis. We use a polar bead to represent the backbone bead (BB) for each amino acid. For the polar bead, there is a central bead with two dummies carrying opposite charges. This set up results in an induced-dipole effect and adds a directionality to peptide backbone-backbone interaction which helps fold secondary structures. This model has been tested against small peptides and proven to have the ability to simulate folding and aggregation without adding external constraints. This is one big improvement from MARTINI force field where the folding process can not be simulated. Here I collaborate with a former Ph.D student in our group, Dr. Abhilash Sahoo, to upgrade WEPCGM and improve the model to be able to capture tertiary structures and complex fold. This model is named "Transferrable and polarizable coarse grained model for proteins (*ProMPT*)". For the side chain beads in *ProMPT*, the bead types can be divided into three major types: charged, hydrophobic, and polar. The side chain beads are assigned with one of these three bead types depending on their hydrophobicity, charge, and the length or size of the side chain. Again, the polar side chain beads are represented by the polar central site with charged dummies.

1.1.2.2 Coarse grained IL model

Compared to CG protein and membrane models, CG IL models have been much less developed, many of them from the Voth group [62, 63, 64, 65, 66]. Because of their high viscosity leading to a large gap between the time scales of atomistic and macroscopic motion, ILs are especially suited for CG simulations to speed up dynamics. Generally the development of CG IL force fields can be categorized to bottom-up philosophy or top-down philosophy. In the bottom-up philosophy, a CG model can be described by many-body potential of mean force (PMF). The PMF is conditioned on the mapping scheme, which usually relies on researcher's chemical intuition or trial-and-error. Ruza et.al. applied machine learning to learn a temperature transferable CG model for $[\text{BMIM}]^+[\text{tetrafluoroborate}]^-$ [67]. They validated their model with radial distribution function (RDF) and transport properties and found a better agreement at transport properties with atomistic simulations using their temperature transferable model. On the other hand, Vazquez-Salazar et.al. has developed a CG model for imidazolium based ILs based on MARTINI3 [68]. These authors verified the validity of their model with macroscopic density against experiments and nanostructures against atomistic simulations. They also demonstrated the compatibility of this IL model with other MARTINI beads. In this thesis we follow the top-down philosophy, where we apply the same bead types from *ProMPT* to represent different part of an IL ion and validate the model against experimental density and atomistic RDF.

1.2 Outline of thesis

In Chapter 2 I will first discuss the effects of two imidazolium based ILs on the stability and activity of hen egg white lysozyme. Here I collaborate with an experimental group at the University of Massachusetts (Bermudez lab) and apply atomistic MD simulations to approach the research questions. We find that lysozyme is destabilized in the presence of both ILs studied due to contributions from both the cations and anions, acting in local and global manners, respectively. The imidazolium cation exerts its effect locally by binding to hydrophobic residues and destabilizing a native Arg-Trp-Arg interaction that has been found previously critical to lysozyme stability. The anions exert their effects globally through electrostatic effects. We then continue to look at the effects of the same ILs on lysozyme activity. From simulations it is observed that the cations occupy the vicinity of the active site. In addition, the ligand-lysozyme complex is less stable in the presence of imidazolium based ILs. These results explain the reduction of lysozyme activity when in high IL content. When undergoing rehydration, the fast leaving of the cations is observed for both IL systems, which explains the reappearance of the active site availability because there is no more cations present near the active site after rehydration. What we found not only provides explanations to the experimental observations, but also help establishing general rule of thumbs to follow when design protein-IL preservation pairs.

In chapter 3 I introduce a CG-MD model I co-develop with Dr. Abhilash Sahoo (a former PhD student in our group). The model is named ”Transferable and polarizable coarse grained model for proteins (*ProMPT*)” to reflect its high transferability and its ability to capture environment-induced polarity. The novelty of *ProMPT* is the implementation of peptide backbone structural polarization. This set up results in an induced-dipole effect and adds a directionality to peptide backbone-backbone interaction which is crucial in folding secondary structures. This is one big improvement from other CG force fields where the folding events and the response to change of environment can not be addressed simultaneously. Using this model we have demonstrated that we can reproduce the free energy landscape for proteins that have a specific secondary structure. In addition, we can also capture the dimerization of transmembrane proteins in a heterogeneous environment. Change of conformational landscape caused by point mutations can also be observed and will be discussed in this chapter.

In chapter 4, I apply *ProMPT* CG model and an in-house CG model for ILs to study the effects of $[\mathrm{TEA}]^+[\mathrm{Ms}]^-$ on Aβ 16-22 aggregation. Aβ 16-22 aggregation has been pointed to be strongly related to the pathogenesis of the Alzheimer's disease by many literature. We first validate our CG IL model with atomistic simulation results and obtain a good agreement in the cation-anion RDF and nano-structuring. After validation of the CG IL model, we perform simulations on multiple Aβ 16-22 peptides in water and in $[\mathrm{TEA}]^+[\mathrm{Ms}]^-$ at different concentrations. The computed β-sheet fraction decreases as the IL concentration increases, and so does the aggregation size. Although aggregates are observed in both water and in the presence of ILs, the formats are not identical. In water, Aβ 16-22 peptides can form a nice hydrophobic core through Phe-Phe interaction, which leads to a stable cross-β structure. On the other hand, in the presence of IL, aggregates can still form but not necessarily with high β-sheet contact, or β-sheet contacts can form within small oligomers, but do not involve all the peptides in the system as seen in water. $[\mathrm{TEA}]^+$-Phe interaction is observed to be an important driving force for the disruption of the clusters in ILs. We also look at the helical fraction in all solvents and find an increase in helical fraction as the IL concentration raises up. This observation could be explained by the higher degree of disruption in inter-peptide contacts at high IL concentrations. In addition, helicity is also enhanced in the presence of ILs in single helix studies. In this case, $[TEA]^+$ attracts the N-terminal of Aβ 16-22 through electrostatic interaction with the backbone charge and thus enhances the dipole alignment within the peptide leading to a higher helicity.

Chapter 2: Evaluate the role of $[EMIM]^+[EtSO_4]^-$ and $[EMIM]^+[Et_2PO_4]^-$ on lysozyme stability and activity

2.1 Overview

This chapter is based on the author's publications: (1) Dual mechanism of ionic liquid-induced protein unfolding. Pei-Yin Lee, Onkar Singh, Silvina Matysiak, and Harry Bermudez. Physical Chemistry Chemical Physics, 2020. [69] (2) Recovery of enzyme structure and activity following rehydration from ionic liquid. Pei-Yin Lee, Onkar Singh, Harry Bermudez, and Silvina Matysiak. Physical Chemistry Chemical Physics, 2022. [70] All experimental work was performed by Onkar Singh and Harry Bermudez, while all simulation work was completed by Pei-Yin Lee and Silvina Matysiak.

Ionic liquids (ILs) are gaining attention as protein stabilizers and refolding additives [13, 71, 72, 73]. However, varying degrees of success with this approach motivates the need to better understand fundamental IL–protein interactions. A combination of experiment and simulation is used to investigate the thermal unfolding of lysozyme in the presence of two imidazolium-based ILs (1-ethyl-3-methylimidazolium ethylsulfate, $[EMIM]^+[EtSO_4]^-$ and 1-ethyl-3-methylimidazolium diethylphosphate, $[EMIM]^+[Et_2PO_4]^-$). Both ILs reduce lysozyme melting temperature (T_m), but more gradually than strong denaturants. $[EMIM]^+[Et_2PO_4]^-$ lowers lysozyme T_m more readily than $[EMIM]^+[EtSO_4]^-$, as well as requiring less energy to unfold the protein, as determined by the calorimetric enthalpy ΔH. Intrinsic fluorescence measurements indicate that both ILs bind to tryptophan residues in a dynamic mode, and furthermore, molecular dynamics (MD) simulations show a high density of $[EMIM]^+$ near lysozyme's Trp62 residue. For both ILs approximately half of the $[EMIM]^+$ cations near Trp62 show perfect alignment of their respective rings. The $[EMIM]^+$ cations, having a "local" effect in binding to tryptophan, likely perturb a critically important Arg-Trp-Arg bridge through favorable π–π and cation–π interactions. Simulations show that the anions, $[EtSO_4]^-$ and $[Et_2PO_4]^-$, interact in a "global" manner with lysozyme, due to the protein's strong net positive charge. The anions also determine the local distribution of ions surrounding the protein. $[Et_2PO_4]^-$ is found to have a closer first coordination shell around the protein and stronger Coulomb interactions with lysozyme than $[EtSO_4]^-$, which could explain why the former anion is more destabilizing. Patching of ILs to the protein surface is also observed, suggesting there is no universal IL solvent for proteins, and highlighting the complexity of the IL–protein environment.

Following the examination of the effects of ILs on protein stability, we use the same set of ILs and protein to investigate the effects of ILs on protein activity. Various spectroscopic experiments and MD simulations are performed to ascertain the structure and activity of lysozyme. Circular dichroism spectroscopy confirms that lysozyme maintains its secondary structure upon rehydration, even after 295 days. Increasing the IL concentration decreases the activity of lysozyme and is ultimately quenched at sufficiently high IL concentrations, but the rehydration of lysozyme from high IL concentrations completely restores its activity. Such rehydration occurs in the most common lysozyme activity assay, but without careful attention, this effect on the IL concentration can be overlooked. From simulations we observe occupation of $[EMIM]^+$ ions near the vicinity of the active site and the ligand-lysozyme complex is less stable in the presence of ILs, which results in the reduction of lysozyme activity. Upon rehydration, fast leaving of $[\mathrm{EMIM}]^+$ is observed and the availability of active site is restored. In addition, suppression of structural fluctuations is also observed when in high IL concentrations, which also explains the decrease of activity. This structure fluctuation suppression is recovered after undergoing rehydration. The return of native protein structure, fluctuation, and activity indicates that after rehydration lysozyme returns to its original state. Our results also suggest a simple route to protein recovery following extended storage.

2.2 Introduction

Due to the importance of proteins in food, catalysis [74], and biomedicine [75], many efforts in recent years have explored ILs to modulate protein behavior [76, 77]. There is also interest in the long-term preservation of proteins without refrigeration [78]. To this end, substantial efforts have been undertaken to explore protein behavior in IL-water mixtures, usually with a focus on model globular proteins such as lysozyme [79, 80], ribonuclease A [81, 82], cytochrome c [83, 84], hemoglobin [85], and lipases [86, 87]. In spite of these efforts no general trend has emerged for how ILs modulate protein stability, suggesting the need for a closer examination of specific protein-IL pairings.

It is important to note that in many reports the amount of IL [88, 89, 90] and/or the protein concentration [81, 83, 91] can be quite low, such that the solutions are largely dilute and aqueous. One benefit of such conditions is that changes in protein stability can be interpreted within the context of Hofmeister series [92] and/or electrostatic screening effects. However, when the goal is to enhance ambient and long-term protein stability, it is questionable if dilute conditions are relevant.

Several recent works examined lysozyme behavior specifically with significant IL content. Most notable is the early work by Byrne et al. [93] who demonstrated multi-year structural stabilization, as confirmed by calorimetry, by using a mixture of 27 wt% sucrose and 31 wt% ethylammonium nitrate. Mann et al. [91] studied the ability of several related ammonium formate ILs to promote lysozyme refolding after thermal denaturation. Ethanolammonium formate performed particularly well at this task, and when at concentrations greater than 50 wt% it also improved lysozyme thermal stability. Weaver et al. [94] studied the effect of choline dihydrogenphosphate and found an increase in lysozyme thermal stability with increasing concentration (up to 40 wt%). The above papers suggest an important role for the hydroxyl group of the cation (ethanolammonium and choline), and yet anions are well known to contribute to protein behavior [76, 81, 95]. Indeed, Rodrigues et al. [96] used ILs of choline paired with various anions to show that lysozyme thermal stability could be strongly affected (either increase or decrease) depending on the anion. It becomes clear that there is a need for closer examination of *both* IL ions on protein behavior, and a focus on either cation or anion alone is likely to be insufficient.

Another limitation to our understanding is that molecular insights are often still lacking, especially in the cases where proteins are destabilized. For example, the details of the unfolding process also remain largely unexplored: do proteins partially unfold? Is the unfolding process cooperative, reversible? Considering specific pairings of proteins and ILs, what molecular interactions are responsible for the observed behavior? These types of questions can be answered more thoroughly through a combination of experiment and simulation, taking advantage of the benefits from each approach. MD simulation is uniquely suited to provide a wealth of detailed information, such as atomic-scale structure and microscopic kinetics. MD also allows us to study the IL organization near the protein surface, as well as short and long-range interactions between the ions and the protein.

Towards the goal of understanding the effect of ILs on protein stability, in this chapter we use a combination of experiment and simulation to investigate the behavior of a model protein, lysozyme, in the presence of two different ILs, $[\text{EMIM}]^{+}[\text{EtSO}_4]^{-}$ and $[\text{EMIM}]^{+}[\text{Et}_2\text{PO}_4]^{-}$. These ILs are chosen due to their miscibility with water and because sulfate and phosphate anions have relevance to protein-protein interactions [97].

While the first half of the chapter focuses on the effects of ILs on proteins' stability, the second half of the chapter turns attention to the effects of the same ILs on protein activity. A fundamental task of proteins is their activity as enzymes, which are needed to catalyze many biological reactions. Protein denaturation will likely result in the loss of protein catalytic function [98, 99]. In addition, blocking of an enzyme's active site by competitive molecules against the natural ligands or changes in protein fluctuations will also affect protein activity. From several literature studies it is reported that the use of ILs is a promising technique in protein refolding and stabilization [13, 100, 101, 102]. The effect of long-time preservation of proteins in ILs on protein activity is critical in protein-IL pairing design.

Much effort has gone into studying the effects of ILs on protein structure and activity [35, 36, 37, 39, 41, 42, 93, 103, 104], but only a few of the literature examples discuss the effects of ILs on proteins after long-time preservation [93, 104]. Previous studies have focused on different proteins and ILs, and many observations are protein dependent, or sometimes IL concentration dependent. Yu *et al.* reported the increase in α-helical content of lysozyme in low concentration of imidazolium based ILs leading to more stable structures and no changes in activity. However, the stabilization is reversed to destabilization as the IL concentration increases [39]. A similar finding is also reported by Mogha *et al* [105]. On the other hand, high IL concentration resulting in little loss of activity after solvated in imidazolium ILs is observed with hyperthermophilic enzymes [41] and papain [42].

It has become increasingly common to study protein behaviors in IL-rich solvents [79, 91, 93, 94, 106]. However, the enzyme concentration is generally held quite low. In case of lysozyme in particular, the most common activity assay is based on turbidity changes due to lysis of *Micrococcus lysodeikticus*. This assay introduces a volume change in the range of *twenty- to thirty-fold*, with a corresponding large effect on the concentrations of the enzyme, and more importantly, the IL. The reported activities of lysozyme in several papers including the works discussed above [39, 80, 91, 94, 105] appear to not account for these changes in IL concentration, possibly leading to erroneous conclusions regarding the effect of ILs on enzyme activity. Therefore, extra attention has to be paid when looking at the protein activity in ILs.

With experimental observations we learn that different protein-IL combinations will lead to different effects on protein activity, but a systematic rule is still lacking to understand whether the change of activity depends on change of protein stability or the availability of active site. In addition, whether the reduction of activity is reversible or irreversible is not thoroughly discussed in the literature. To answer these questions, MD is a powerful tool to provide a wealth of detailed atomic-scale structure and microscopic kinetics when inaccessible to experimental techniques. Properties related to protein activity can be examined through simulations and linked back to the activity measured from experiments. Some properties that can be studied by MD include the changes in protein conformation, the fluctuation degree of the structure, and the interaction between ILs and the protein, especially with the active site residues.

In the second half of the chapter we study the effects of the same ILs on lysozyme structure and activity before and after rehydration, where lysozyme is preserved in ILs for almost a year. We demonstrate that addition of water, or *rehydration*, has a dramatic impact on both lysozyme structure and activity. Experiments and simulations are combined to complement each other to give a more complete picture on how these ILs affect the protein behavior. We first look at the secondary and tertiary structures of lysozyme before and after rehydration and then study the changes of lysozyme activity before and after rehydration and investigate the possible causes via MD simulations.

2.3 Methods

All experiments were conducted by Bermudez group. See the SI for more details on experiment setup.

2.3.1 Simulations

Three systems containing lysozyme (water, 25 wt% $[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$ and 25 wt% $[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$) were simulated with all-atom MD using Gromacs 2018 package [107]. For the starting protein structure, the coordinates of chicken egg white lysozyme of Protein Data Bank (PDB) entry 1AKI was taken with all crystallographical water molecules removed. OPLS-AA force field [108] was used for lysozyme, water and chloride ions. A non-polarized force field developed by Canongia et al [109], which was compatible with the OPLS-AA force field was used for $[\mathrm{EMIM}]^+$ and $[\mathrm{EtSO_4}]^-$ ions. LigParGen server was used to generate force field parameters that were also compatible with OPLS-AA forced field for $[\mathrm{Et_2PO_4}]^-$ [110]. The pH of the systems were set at 7 to match experimental conditions of water. For the water system, lysozyme was solvated in a $7.01 \times 7.01 \times 7.01$ nm^3 cubic box with 10636 SPC water molecules and 8 chloride ions to neutralize the system. For the 25 wt% [EMIM]$^+$[EtSO$_4$]$^-$ system, lysozyme was solvated in a $8.50 \times 8.50 \times 8.50$ nm^3 cubic box with 9041 SPC water molecules, 8 chloride ions, 230 [EMIM]$^+$ ions and 230 [EtSO$_4$]$^-$ ions. For the 25 wt% [EMIM]$^+$[Et$_2$PO$_4$]$^-$ system, lysozyme was solvated in a $8.00 \times 8.00 \times 8.00$ nm^3 with 8456 SPC water molecules, 8 chloride ions, 192 [EMIM]$^+$ ions, and 192 [Et$_2$PO$_4$]$^-$ ions. The molecular structures for all IL ions are shown in Figure 2.1.

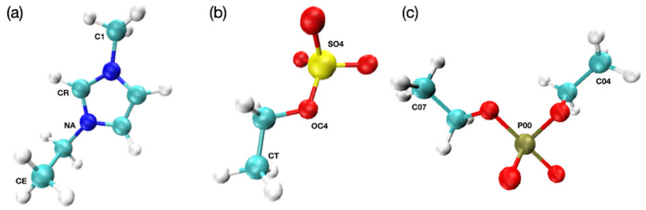


Figure 2.1: Molecular structures of (a) [EMIM]$^+$ cation, (b) [EtSO$_4$]$^-$ anion, and (c) [Et$_2$PO$_4$]$^-$ anion. White: hydrogen; cyan: carbon; blue: nitrogen; red: oxygen; yellow: sulfur; and tan: phosphorus. The marked atoms are for the identification purpose in Figure 2.7.

Prior to production runs, the energy of the system was minimized using the steepest descent method, followed by isochoric-isothermal (NVT) and isobaric-isothermal (NPT) ensemble runs of 100 ps each and at 300 K. Pressure for NPT equilibration was set at 1 bar. The volumes of the simulation boxes after initial equilibration were $6.96 \times 6.96 \times 6.96$ nm^3, $7.12 \times 7.12 \times 7.12$ nm^3, and $6.98 \times 6.98 \times 6.98$ nm^3 for water, 25 wt% [EMIM]$^+$[EtSO$_4$]$^-$, and 25 wt% [EMIM]$^+$[Et$_2$PO$_4$]$^-$ system, respectively. As for the production run, a total of 9 trajectories have been run for 1 μs for the three systems and at 3 temperatures (300 K, 370 K, and 400 K). All production runs were run with NPT ensemble with a time step of 2 fs and pressure at 1 bar. Linear Constraint Solver Algorithm (LINCS) was used for the constraint algorithm and all bonds involving hydrogens were constrained. The cutoff for both short range Coulomb and van der Waals (vdW) interactions was 1.0 nm. These cutoff values were also used to calculate the short range interactions between lysozyme and ILs. Typically a cutoff between 0.8 to 1.2 nm is used. Although longer cutoff values are desirable, 1.0 nm is chosen because of the desire of better sampling. [111] The particle-mesh Ewald (PME) method [112] was used for the long-range electrostatic interactions. Parrinello-Rahman barostat [113] and V-rescale thermostat were used to control pressure and temperature, respectively. The compressibilities of the system were set to 4.50×10^5 bar^{-1}, 4.23×10^5 bar^{-1}, and 4.23×10^5 bar^{-1} for water, 25 wt% [EMIM]$^+$[EtSO$_4$]$^-$, and 25 wt% [EMIM]$^+$[Et$_2$PO$_4$]$^-$ systems, respectively, to reflect different solvent compressibility between water and IL/water mixtures [114].

The last 800 ns data were used for analysis. Native contact pairs were first determined by selecting two atoms that are more than four residues away but less than 0.75 nm in distance from 1AKI.pdb structure. The pair list was then used to determine the fraction of native contacts of a structure with the same cutoff of 0.75 nm. Potential mean force (PMF) was calculated based on eq.1, where $W(rxn)$ is the PMF of a reaction coordinate (rxn) and P is the probability distribution of that rxn.

$$W(rxn) = -\ln P(rxn) \tag{2.1}$$

The probability distribution of the absolute cosine angle between the normal of [EMIM]$^+$ rings near Trp62 and the normal of Trp62 ring is also calculated. An [EMIM]$^+$ is considered to be near Trp62 if its CR atom (Figure 2.1) is within a cutoff of 0.4 nm from either the Cδ2 atom or Cϵ2 atom of Trp62. This cutoff is justified by the RDF between [EMIM]$^+$ CR and Trp62 Cδ2 or Cϵ2, where the maxima occur at around this value. For the spatial distribution function

(SDF) figures, isovalues used for cation and anion in $[\text{EMIM}]^+[\text{EtSO}_4]^-$ are -0.01285 and -0.000586, respectively; isovalues used for cation and anion in $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ are -0.019402 and 0.009118, respectively. All other analyses along with SDF were done with the inbuilt tools in Gromacs or VMD [115]. All structure images were rendered from VMD.

For the study of the effects of ILs on protein activity, although the same set of ILs and protein were used, different set up was applied. Five systems containing lysozyme (water, 50 wt% $[\text{EMIM}]^+[\text{EtSO}_4]^-$, 50 wt% $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$, rehydrated $[\text{EMIM}]^+[\text{EtSO}_4]^-$, and rehydrated $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$) were simulated. The number of water molecules, IL ions, and neutralized ions (Cl^-) for each system are provided in Table 2.1. For the starting protein structure, the coordinates of chicken egg white lysozyme (PDB code ID: 1AKI) was taken with all crystallographical water molecules removed. For the initial conformation of the rehydrated systems, the final conformation of the corresponding 1μs simulation of the 50 wt% system was taken and aligned with the protein crystal structure that is in its ligand-bound form with an oligosaccharide (tri-N-acetylchitotriose) (PDB code ID: 1HEW). All the ILs that are outside of a cutoff (0.6 nm for $[\text{EMIM}^+]$ and 0.6 nm for $[\text{EtSO}_4]^-$) from any atom of the ligand are removed in the rehydrated $[\text{EMIM}]^+[\text{EtSO}_4]^-$ system, leaving only the ILs that are potentially interacting with the active site region. These cutoff values are similar to the shortest distance between the ligand and the active site residues and were thus chosen in order to select a reasonable amount of ILs that aims to mimic the status of the active site when the system is undergoing rehydration. A cutoff of 0.6 nm and 0.52 nm for $[\text{EMIM}^+]$ and for $[\text{Et}_2\text{PO}_4]^-$ were used respectively for the rehydrated $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ system for the same purpose. The reason of different cutoff values for the cation and anion is because by using these cutoff values, the same number of cations and anions can be selected to make the system neutral. Water molecules and neutralized ions were then added to solvate the whole system. The number of water molecules depends on the box size, which has the edge at least 1 nm from the protein and the initially selected ILs. For the rehydrated $[EMIM]^+[EtSO_4]^-$ system, the final concentration is 0.82 wt%, and for the rehydrated $[EMIM]^+[Et_2PO_4]^-$ system, the final concentration is 0.73 wt%.

Table 2.1: Number of water molecules, IL ions, and chloride ions for each system along with the dilution factors if applicable.

	Water	Cation	Anion	Cl^-	df
Water	10636	N/A	N/A	8	N/A
50 wt% $[EMIM]^+[EtSO_4]^-$	6348	485	485	8	N/A
50 wt% $[EMIM]^+[Et_2PO_4]^-$	6292	429	429	8	N/A
Rehydrated $[EMIM]^+[EtSO_4]^-$	12713	8	8	8	60.63
Rehydrated $[EMIM]^+[Et_2PO_4]^-$	19995	10	10	8	42.90

The parameters for energy minimization, NVT equilibration, NPT equilibration, and MD production were the same as the previous set up. For the production run, water and 50 wt% systems were run for 1 μs at 300 K and the rehydrated systems were run for 500 ns at 300K. Three replicas with different velocity random seeds were performed for the rehydrated systems. The last 500 ns data of water and 50 wt% systems and the whole 500 ns trajectories for the rehydrated systems were used for the analysis. For the SDF figures, isovalues of 0.015051 and -0.014899 were used for $[EMIM^+]$ in the 50 wt% $[EMIM]^+[EtSO_4]^-$ system and 50 wt% $[EMIM]^+[Et_2PO_4]^-$ system, respectively. For the $[EMIM^+]$ time series analysis, the cutoff of the $[EMIM^+]$ near active site is the same as that was used to construct the initial rehydrated system.

A set of umbrella sampling simulations were performed in water and in 50 wt% $[EMIM]^+[EtSO_4]^-$. The topology and the OPLS-AA force field parameters for the sugar ligand (tri-N-acetylchitotriose) were generated with the LigParGen server and relaxed through energy minimization and a short NPT equilibration. The equilibrated ligand was then combined with the lysozyme structure

(1AKI.pdb) with the same protein-ligand alignment as that in the lysozyme-ligand complex (1HEW.pdb). The lysozyme-ligand complex was then put into a 6.3 nm $\times$ 6.3 nm $\times$ 13 nm box. The box was then solvated with water and neutralized with 8 chloride ions. The complex system went through a set of energy minimization and NPT equilibration (time step 2 fs for 100 ps). The pulling simulation was then conducted with a pulling rate of 0.01 nm/ps. The center of mass (COM) of the ligand was pulled away from the COM of lysozyme along the direction of the principal axis of the ligand. A total of 30 conformations were saved for the following umbrella sampling simulations. Conformations with different lysozyme-ligand distance were first equilibrated with NPT (time step 2 fs for 100 ps) and then the production run (time step 2 fs for 30 ns). Weighted histogram analysis method (WHAM), a built-in tool in Gromacs, was applied to extract the PMF from various umbrella sampling runs and the free energy landscape along a reaction coordinate (distance between the COM of the ligand and the COM of lysozyme) was then generated.

2.4 Results and discussion

2.4.1 Evaluate the role of $[\text{EMIM}]^+[\text{EtSO}_4]^-$ and $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ on lysozyme stability

2.4.1.1 Compare the stability of lysozyme in water and in ILs/water mixtures

From the calorimetry results (Figure A.2), it is observed that both ILs show gradual destabilizing effect on lysozyme, which is in contrast to the strong denaturing effect exerted by [Gdm][Cl]. Within the two ILs, $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ destabilizes lysozyme more significantly.

From the transition enthalpy ΔH (Figure A.3), $[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$ also shows a lower ΔH, which corresponds to less energy for lysozyme unfolding.

Computationally, we observe a similar trend for lysozyme stability as with experiments. Figure 2.2 shows the fraction of native contacts, C-alpha root-mean-square-deviation (Cα RMSD), and radius of gyration (R_g) time series for water, 25 wt% $[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$, and 25 wt% $[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$ at three temperatures. The starting configurations are folded structures taken right after NPT equilibration for each system. From fraction of native contacts and Cα RMSD at 300 K we see that lysozyme is folded at all times in all systems with RMSD staying at around 0.2 nm. Indeed, lysozyme is very stable thermally considering its rather small size [116, 117, 118, 119, 120]. The reason why lysozyme in water has slightly lower fraction of native contacts can be attributed to the more flexible movements from the loops and turns as seen in the C-alpha root-mean-square-fluctuation (Cα RMSF) analysis (Figure 2.3). The suppression of RMSF of a protein in aqueous IL mixtures has also been found for xylanase in $[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$ [40].

When the temperature increases to 370 K as shown in Figure 2.2 (b) and (e), destabilization of lysozyme is observed starting from around 250 ns for both ILs. This temperature is also the temperature where we start to see different states occur for water system and IL systems kinetically, i.e., lysozyme is partially unfolded in ILs, but still folded in water at around 250-400 ns. A Cα RMSD cutoff of 0.4 nm is used to distinguish folded states and not folded states. This cutoff can be seen more clearly from the PMF plots at 370 K (Figure 2.4), where the folded basins and the intermediate basins [121][122] are separated by this cutoff. As the simulation time becomes longer, lysozyme in water also starts to partially unfold, while lysozyme in $[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$ further unfolds to have a fraction of native contacts of around 0.3. When temperature increases to

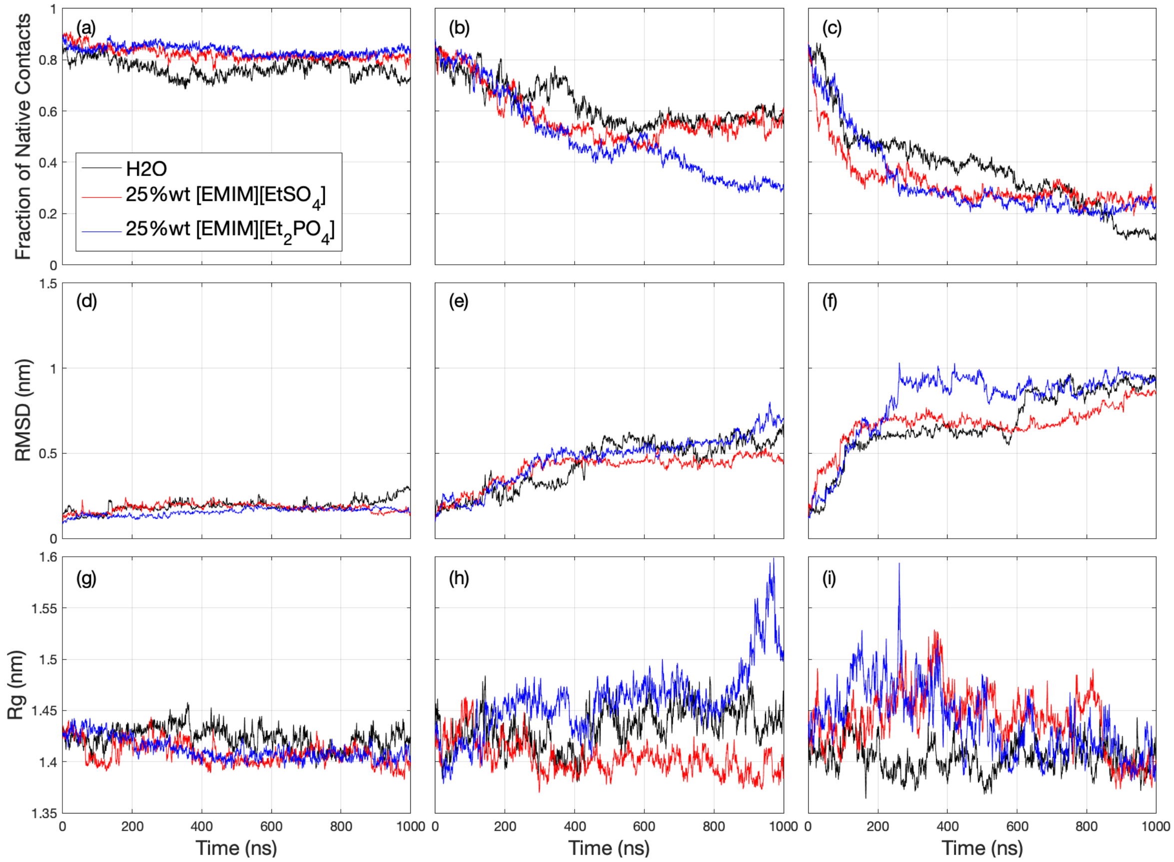


Figure 2.2: Fraction of native contacts, RMSD, and R_g time series for water (black), 25 wt% $[\text{EMIM}]^+[\text{EtSO}_4]^-$ (red), and 25 wt% $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ (blue) systems. The first column is at 300 K, the second column is at 370 K, and the third column is at 400 K.

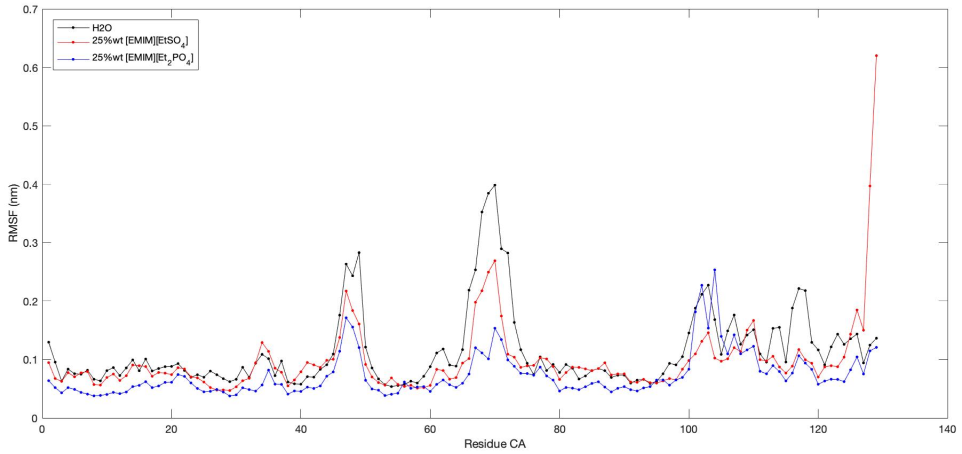


Figure 2.3: RMSF of lysozyme Cα atoms in nm, where black represents water, red represents 25 wt% $[\text{EMIM}]^+[\text{EtSO}_4]^-$, and blue represents 25 wt% $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$.

400 K, ILs still destabilize lysozyme faster than water does, as evident from the fraction of native contacts time series.

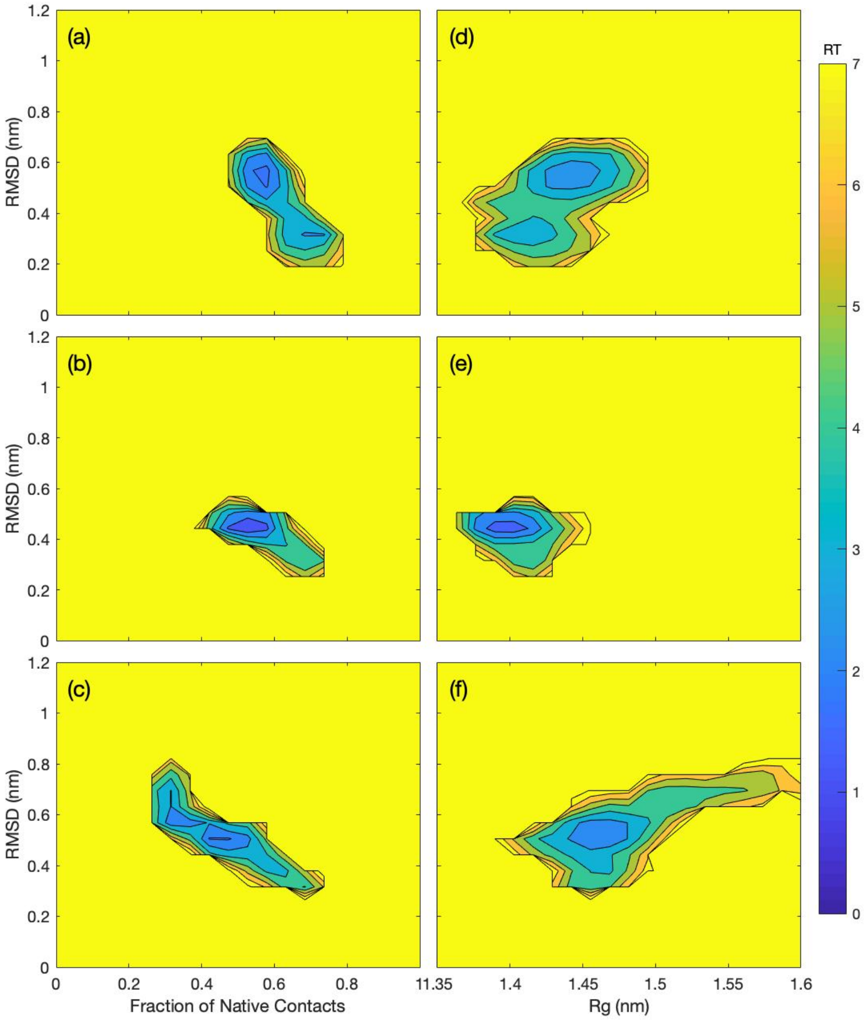


Figure 2.4: PMF plot at 370 K, where the reaction coordinates used for the left column are RMSD and fraction of native contacts and the reaction coordinates used for the right column are RMSD and R_g. (a) and (d) are water system, (b) and (e) are 25 wt% $[\text{EMIM}]^+[\text{EtSO}_4]^-$ system, and (c) and (f) are 25 wt% $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ system. PMF is in RT unit.

For R_g at 300 K, lysozyme in water has a slightly higher value compared to the other two systems, again possibly due to larger fluctuations in the loops and turns. At elevated temperatures, $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ has larger fluctuation in R_g compared to the other two systems. This indicates that lysozyme in $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ samples more swollen conformations in not folded states.

2.4.1.2 Cation effects on lysozyme stability

From the fluorescence experiment data (Figure A.4 and Table A.1), it is observed that the binding interaction between lysozyme (Trp residues) and ILs is predominately dynamic, with $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ having a larger binding constant. Although the effects of the IL cation and anion are complex, as we will discuss below, at least part of their destabilizing effect on lysozyme stability can be attributed to their difference in binding affinity with Trp. Indeed, the importance of Trp62 to lysozyme stability is thought to occur via a ”sandwich-type” interaction between itself and two arginines (Arg73 and Arg113) [123, 124]. A plausible mechanism of interaction between the ILs and Trp62 is through π-π and cation-π stacking between the $[\text{EMIM}]^+$ cation's imidazolium ring and Trp62's indole ring. Such an interaction would compete with the native, and stabilizing, Arg-Trp-Arg sandwich.

Simulations allow us to directly observe the interactions of IL cations and anions at Trp62, and to confirm the competitive binding mechanism. Figure 2.5 shows the probability distribution of the absolute cosine angle between the normal of the $[\text{EMIM}]^+$ ring and the normal of the Trp62 ring within a cutoff of 0.4 nm. The details of the calculation are described in the Experimental section. If the absolute cosine angle value equals to one, that means the rings are perfectly aligned with each other, and if the value equals to zero, that means the rings are perpendicular to each other. If we define the existence of the ring-ring alignment to have an absolute cosine angle value larger than 0.75, then for 25 wt% $[\text{EMIM}]^+[\text{EtSO}_4]^-$ is 89.80% aligned and for 25 wt% $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ is 83.91% aligned, which both indicate strong interaction between rings.

The insets in Figure 2.5 show the SDF of $[EMIM]^+$ near lysozyme for the two ILs. From these figures the high density region of $[EMIM]^+$ near Trp62 can be seen more readily.

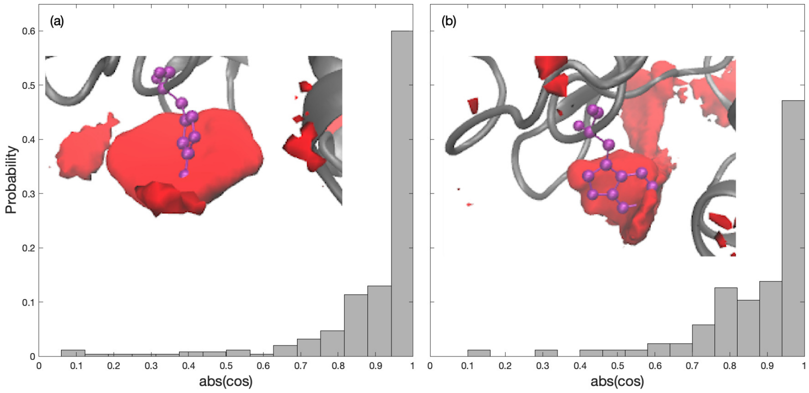


Figure 2.5: The probability distribution of absolute cosine between the normal of $[EMIM]^+$ ring and the normal of Trp62 ring within a cutoff of 0.4 nm. The insets are SDF (200 ns–1 μs) of $[EMIM]^+$ near the average structure of Trp62 (200 ns–1 μs). (a) is in 25 wt% $[EMIM]^+[EtSO_4]^-$ and (b) is in 25 wt% $[EMIM]^+[Et_2PO_4]^-$. Trp62 is shown in purple CPK representation.

The importance of Trp62 in lysozyme stabilization has been previously established both experimentally [123] and also in MD simulations [125][124]. As mentioned earlier, our fluorescence data (Figure A.4) lead us to presume that local π-π and cation-π stacking between $[EMIM]^+$ and Trp62 disturb the native Arg73-Trp62-Arg113 interaction. Our simulation results directly confirm this IL-protein interaction (Figure 2.5). Furthermore, this mechanism explains why lysozyme is destabilized in both ILs studied here: both ILs share the same $[EMIM]^+$ cation. This interpretation is also consistent with the effect of $[EMIM]^+[Cl]^-$ on lysozyme [126], which also reduces the T_m and hence destabilizes the protein (although to a lesser degree than either of the ILs here).

2.4.1.3 Anion effects on lysozyme stability

The significant role of the IL anions is clearly seen in Figure A.2, and simulation provides additional evidence as to why anions are important in a global fashion. Because lysozyme is a highly charged basic protein with an isoelectric point pI $\approx$ 11 [127, 128], it is expected that electrostatic interactions would be prominent in the protein's solubility and stability. Figure 2.6 shows the ratio between ILs and H_2O near protein and in bulk. The ratio between anions and H_2O near protein is greater than that in bulk in both IL systems, which indicates that more anions will partition to the protein surface. In contrast, the ratio between $[EMIM]^+$ and H_2O does not change that much near the protein and in bulk.

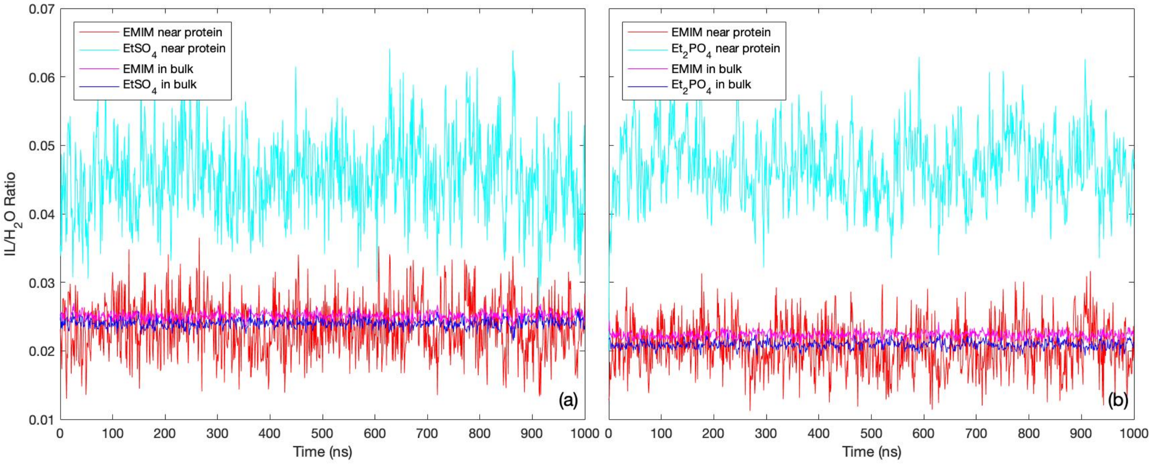


Figure 2.6: Ratio of ILs to H_2O near lysozyme and in bulk in (a) 25 wt% $[EMIM]^+[EtSO_4]^-$ and (b) 25 wt% $[EMIM]^+[Et_2PO_4]^-$ at 300 K.

We next investigate how the specific anion contributes to the difference in lysozyme stability. Figure 2.7 shows the radial distribution function (RDF) between selected atoms of ILs and the whole protein. An illustration of the selected atoms is shown in Figure 2.1. A nearly identical trend is observed for $[EMIM]^+$ in both IL systems, indicating the similar distribution of $[EMIM]^+$ near lysozyme and that the global distribution of $[EMIM]^+$ is not affected by the difference in anions. In addition, the cation peaks are less distinct compared to the anion peaks, indicating the distribution of $[EMIM]^+$ with respect to the protein surface is more delocalized. For the anions, $[Et_2PO_4]^-$ shows a strong orientational effect with phosphorous pointing towards the protein surface (Figure 2.7 (b)), while for $[EtSO_4]^-$ the peaks are less distinct (Figure 2.7 (a)), reflecting this anion's smaller dipole moment. For the same reason, the first coordination shell of $[Et_2PO_4]^-$ is closer to the lysozyme surface than $[EtSO_4]^-$.

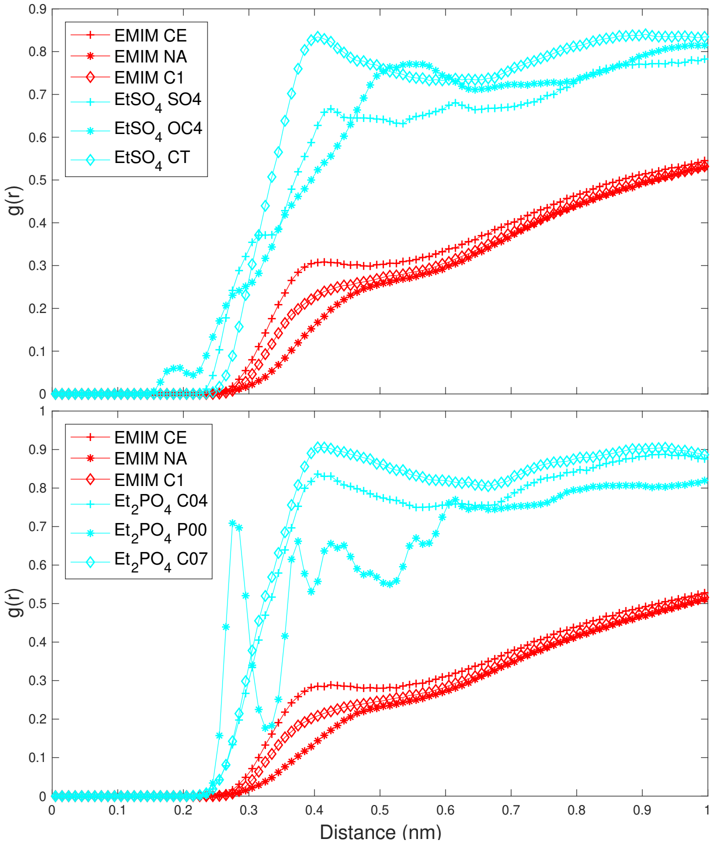


Figure 2.7: RDF between selected atoms of ILs and the whole protein in (a) 25 wt% $[EMIM]^+[EtSO_4]^-$ and (b) 25 wt% $[EMIM]^+[Et_2PO_4]^-$. An illustration of the marked atoms is shown in Figure 2.1.

The short-range (SR) interaction energies between ILs and the protein are further examined in Table 2.2. The cutoff of these two short-range interactions is set at 1 nm. Again, there is no clear difference in Coulomb and Lennard-Jones (LJ) interactions between $[EMIM]^+$ and protein in the two IL systems. However, the Coulomb SR interaction is more than two times stronger between $[Et_2PO_4]^-$ and protein than between $[EtSO_4]^-$ and protein, which agrees with the observation in RDF that $[Et_2PO_4]^-$ has a closer first coordination shell. The two IL anions $[EtSO_4]^-$ and $[Et_2PO_4]^-$ differ in several ways, including that $[Et_2PO_4]^-$ has less charge delocalization and greater hydrophobicity than $[EtSO_4]^-$, providing greater driving forces for interactions with both the protein surface charges and the hydrophobic interior. Since the interaction between the anion and the protein is stronger in $[EMIM]^+[Et_2PO_4]^-$ than in $[EMIM]^+[EtSO_4]^-$, this is another contribution causing the former system to be more destabilizing than the latter one.

Table 2.2: SR interaction between ILs and lysozyme.

Energy Type	Average (kJ/mol)	Standard Deviation (kJ/mol)
	25 wt% $[EMIM]^+[EtSO_4]^-$	
Coulomb (SR) Protein-$EMIM^+$	-40.96	35.87
LJ (SR) Protein-$EMIM^+$	-274.78	43.72
Coulomb (SR) Protein-$EtSO_4^-$	-1389.5	230.15
LJ (SR) Protein-$EtSO_4^-$	-575.76	74.33
	25 wt% $[EMIM]^+[Et_2PO_4]^-$	
Coulomb (SR) Protein-$EMIM^+$	-57.15	40.03
LJ (SR) Protein-$EMIM^+$	-225.21	42.04
Coulomb (SR) Protein-$Et_2PO_4^-$	-3082.48	383.67
LJ (SR) Protein-$Et_2PO_4^-$	-450.96	85.35

Having identified important roles for both the IL cations and the IL anions, we turn our attention to their roles when together. Figure 2.8 shows the SDF of ILs near lysozyme, where lysozyme is shown in gray, red regions correspond to high density of $[\mathrm{EMIM}]^+$ and cyan regions correspond to high density of $[\mathrm{EtSO_4}]^-$ or $[\mathrm{Et_2PO_4}]^-$. We observe that instead of being a homogeneous mixture like the bulk solution, the IL ions distribute in patches across different parts of the lysozyme surface. In both IL systems, positive residues attract anions and negative residues attract cations. In addition, some hydrophobic residues on the protein surface also attract both cations and anions, possibly due to the fact that both the cations and anions have at least one hydrophobic alkyl chain. This non-uniform distribution of ions near the protein surface suggests that there is no ”universal” IL solvent for proteins, but rather, that each protein will need a uniquely tailored IL for optimum solubility and stability.

2.4.2 Evaluate the role of $[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$ and $[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$ on lysozyme activity

2.4.2.1 Lysozyme secondary structure is preserved after rehydration

Far UV CD experiment was first utilized to examine the change of lysozyme secondary structure. Figure A.5 shows that lysozyme is not destabilized after rehydration from high IL content. There is no CD data can be observed in high IL content due to imidazolium cation absorbance. Nevertheless from the CD data we deduce two possibilities, either: (i) lysozyme structure is preserved in high IL content before rehydration, or (ii) lysozyme structure is perturbed in high IL content but is recovered after rehydration. However, *simulations are uniquely able to examine the protein both before and after.*

In order to study computationally the possible changes in structure fluctuation and con-

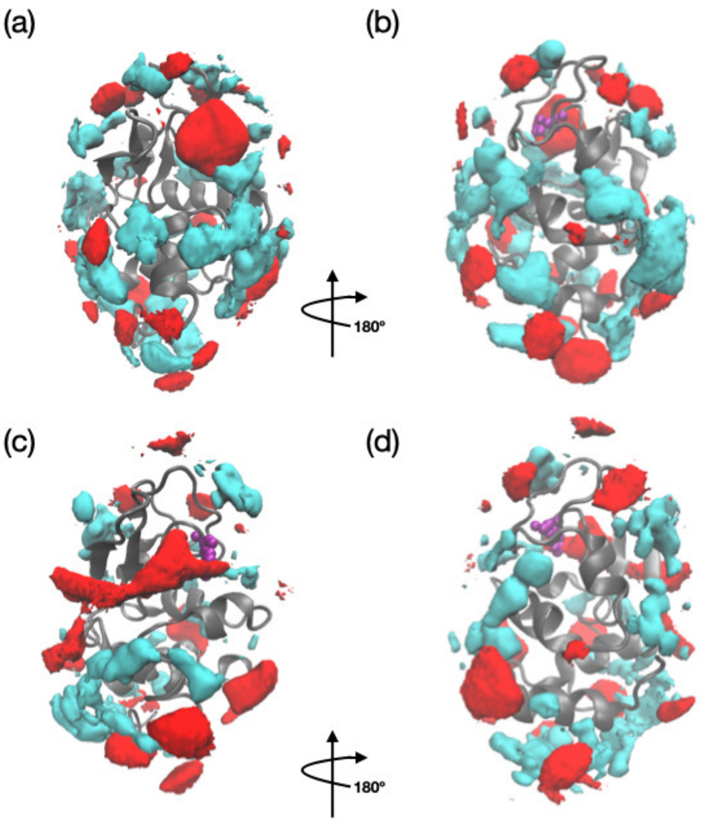


Figure 2.8: SDF of ILs near lysozyme where red represents $[EMIM]^+$ and cyan represents anions: $[EtSO_4]^-$ in (a), (b) and $[Et_2PO_4]^-$ in (c), (d). Trp62 is shown in purple CPK representation. Panels (b) and (d) are after a rotation of approximate 180 ° of (a) and (c), respectively.

formation of lysozyme before and after rehydration, root mean squared fluctuation (RMSF) and the average structure over a 500 ns time course are computed. Figure 2.9 (a) and 2.9 (b) show the RMSF of lysozyme in water, 50 wt% and rehydrated systems for $[EMIM]^+[EtSO_4]^-$ and $[EMIM]^+[Et_2PO_4]^-$, respectively. The RMSF for both 50 wt% IL systems are in general more suppressed compared to that in water, possibly due to high density regions of ILs on the protein surface reducing the conformational freedom as shown in the first half of the results. This dampening of dynamics motion by ILs has also observed for family 11 xylanase in $[EMIM^+]$ based ILs [40]. In addition, the suppression of structure fluctuation is more obvious for $[EMIM]^+[Et_2PO_4]^-$ than for $[EMIM]^+[EtSO_4]^-$, which could be explained by the stronger electrostatic interaction between the $[Et_2PO_4]^-$ and the protein surface as indicated by the first part of this chapter.

After rehydration, the RMSF of both rehydrated systems are recovered to be close to the values in water. There are three regions with particularly high RMSF values (region I: residue 45 - 51; region II: residue 64 - 75; region III: residue 99 - 127). Region I and II correspond to the loops, with region I being near to the active site residues as shown in Figure 2.10. Region III contains some small helical fragments, but is mostly coil and is located at the C-terminus. Other regions that do not show high RMSF values are either buried or well structured (for example, a structured α helix). The reason why the RMSF values of region II and region III for the rehydrated $[EMIM]^+[Et_2PO_4]^-$ system are not fully recovered back to the values in water could be explained by the high B factors of lysozyme in these regions (Figure 2.9 (c)), where a high B factor indicates a higher flexibility of the residue. The high flexibility in residue could possibly lead to a larger deviation in the RMSF values. In addition, one might wonder if the lower RMSF values of region II and III in rehydrated $[EMIM]^+[Et_2PO_4]^-$ could come from ILs having interactions with these loop regions that might cause the slight suppression. In order to answer this question, the trajectory of the rehydrated $[EMIM]^+[Et_2PO_4]^-$ system is examined with a focus on these two regions, and no specific binding of ILs is observed for both loops (results not shown).

In Figure 2.10, the average structures in water, 50 wt% and rehydrated systems are aligned. It is observed that region I in water and in rehydrated system are nicely overlapped in both IL systems, even though the RMSF is not fully recovered. On the other hand, region I of 50 wt% ILs deviates from that in water and in rehydrated system. In 50 wt% $[EMIM]^+[Et_2PO_4]^-$, this deviation in region I is even more obvious. For the 50 wt% systems, although the overall tertiary structures are preserved as well as the secondary structures, the suppression of loop fluctuations and the deviation of region I could possibly affect the activity of lysozyme, considering the close proximity of region I and the active site residues.

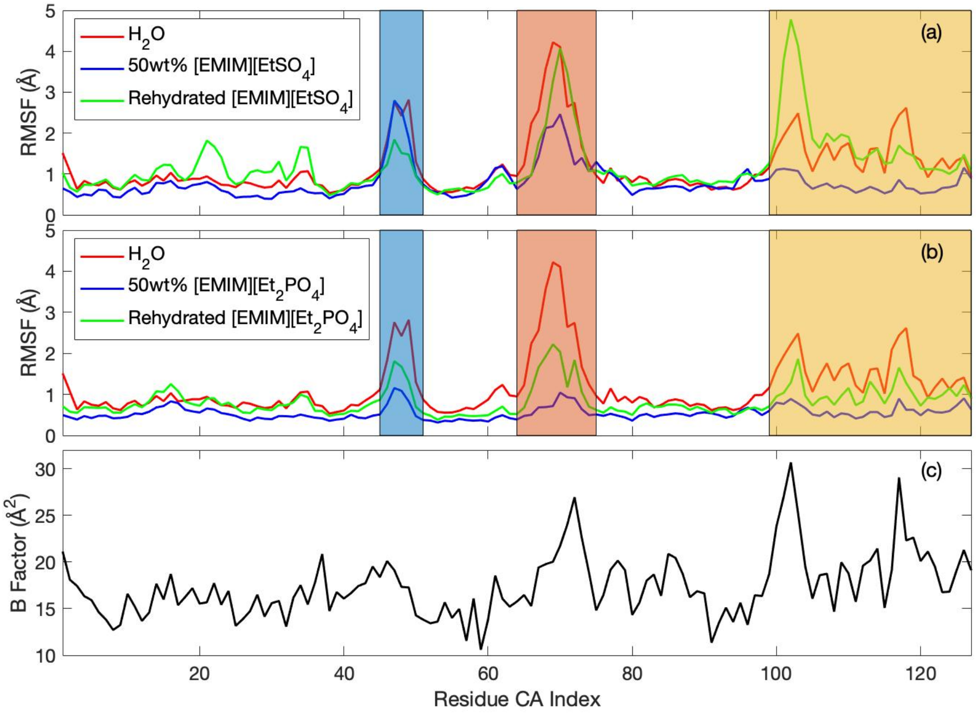


Figure 2.9: RMSF of water (red), 50 wt% ILs (blue) and rehydrated (green) systems for (a) $[\text{EMIM}]^+[\text{EtSO}_4]^-$ and (b) $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$. Blue: Region I; Orange: Region II; Yellow: Region III. (c) is the B factor recorded in 1AKI.pdb.

2.4.2.2 Lysozyme activity is restored after rehydration

Enzymatic activity assay performed by Bermudez group shows that the activity of lysozyme can be restored after rehydration (Table A.2). In addition, as shown in Figure A.6, there is a sharp decrease of activity at moderate IL concentration (around 10 wt%) and no activity in sufficient high IL content. This result is presumably due to blocking of the active site, as interpreted from simulation results that correspond to approximately 1 wt% IL, where the lysozyme activity is fully recovered (see below). The concentration of ILs before rehdyration for simulations is 50 wt%, which is on the right side of the sigmoidal curve, showing no activity.

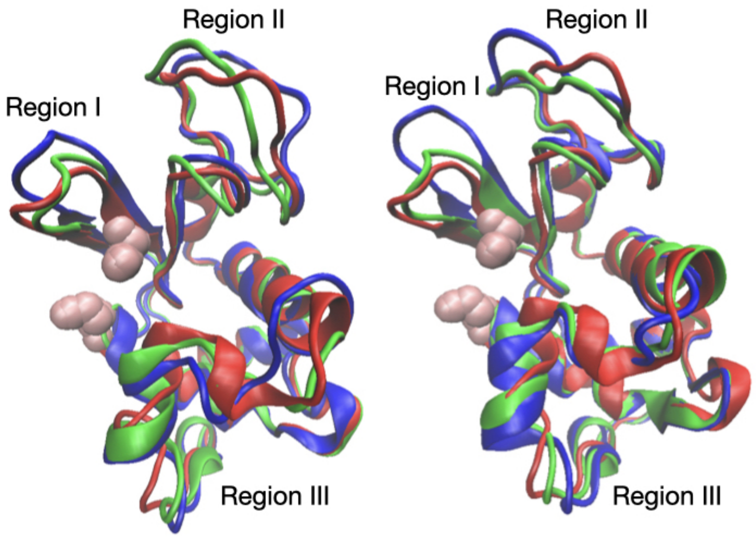


Figure 2.10: The alignment of the average structures from water (red), 50 wt% ILs (blue) and rehydrated (green) systems in $[\mathrm{EMIM}]^+[\mathrm{EtSO}_4]^-$ (left) and $[\mathrm{EMIM}]^+[\mathrm{Et}_2\mathrm{PO}_4]^-$ (right). Pink vdW representations are the active site residues (residue 35 and 52). The high RMSF regions are also indicated.

To gain a molecular understanding on the reduction of lysozyme activity in the presence of 50 wt% ILs, we first compute the average lysozyme structures in 50 wt% ILs and aligned them with the crystal structure of the lysozyme-ligand complex (1HEW.pdb). In Figure 2.11, the orange vdW representations correspond to the lysozyme active site residues (Glu35 and Asp52) while the yellow vdW representations correspond to Trp62 and Trp63. The non-native sugar ligand is shown in green. In addition, the SDF of $[\mathrm{EMIM}^+]$ is expressed in red. Although the IL concentrations are quite high, unlike water as a solvent that is homogeneously distributed around the protein, ILs show specific patching to the lysozyme surface. From the SDF we observe a clear overlapping between the sugar ligand and a high density region of $[\mathrm{EMIM}^+]$ in both systems. The sugar ligand is clamped by Trp62 and Trp63 as illustrated. These two aromatic residues also interact with $[\mathrm{EMIM}^+]$ ions as seen from the high $[\mathrm{EMIM}^+]$ density nearby. Other than the overlapping at the non-native ligand position, $[\mathrm{EMIM}^+]$ also occupies the vicinity of the active site (see [EMIM$^+$] distribution between Trp62/63 and the active site residues and as pointed by the green arrows in Figure 2.11). The occupation of [EMIM$^+$] at the ligand position near the active site could compete with the natural ligands and lead to competitive inhibition.

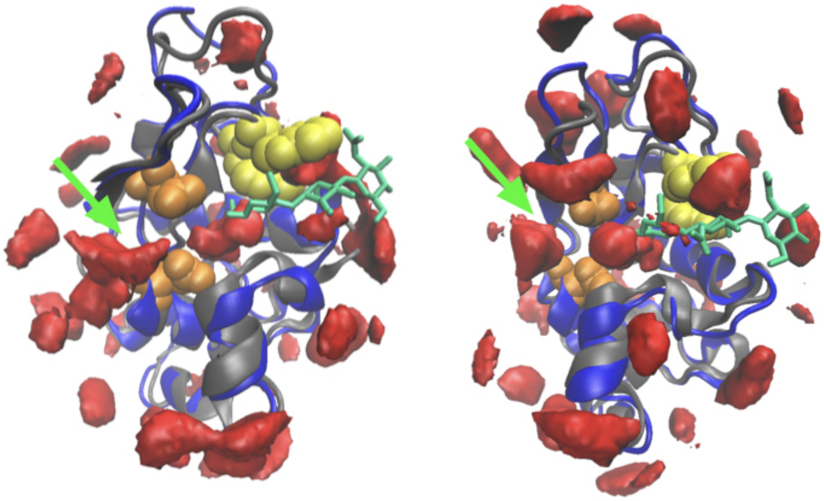


Figure 2.11: Overlapping of the lysozyme-ligand complex with the average structure of lysozyme in 50 wt% $[EMIM]^+[EtSO_4]^-$ (left) and $[EMIM]^+[Et_2PO_4]^-$ (right). In both figures, the crystal structures are in gray, the average structures from simulations are in blue, active site residues are in orange vdW, TRP62/63 are in yellow vdW, non-native sugar ligand is in green vdW and [EMIM$^+$] is in red SDF. The green arrows indicate another region near the active site that is also occupied by high [EMIM$^+$] density.

Figure 2.12 shows the potential of mean force ΔG(r) for the ligand, where r is the distance of the ligand from the lysozyme active site along the pulling direction. The free energy barrier $\Delta G(r)^{TS}$ for the ligand to leave lysozyme is around 47 kJ/mol in water and 18 kJ/mol in 50 wt% $[EMIM]^+[EtSO_4]^-$, which indicates that it takes less energy for the ligand to leave the active site in ILs. Using transition state theory $k \approx \exp(-\Delta G(r)^{TS}/k_B T)$ where k_B is the Boltzmann constant and T is the temperature, a smaller $\Delta G(r)^{TS}$ would translate into faster unbinding kinetics for 50 wt% $[EMIM]^+[EtSO_4]^-$ compared to water. The unbinding free energy in water is around 44 kJ/mol and in 50 wt% $[EMIM]^+[EtSO_4]^-$ is 11 kJ/mol, indicating the protein-ligand complex is less stable in the presence of ILs. We have also calculated the potential energy of the ligand in the active site and in the bulk solvent for both water and 50 wt% $[\text{EMIM}]^+[\text{EtSO}_4]^-$ systems. The difference of ligand potential energy when in the active site and in bulk is similar in the two systems (26.63 kJ/mol in water and 27.93 kJ/mol in IL). Therefore, the lower stability for ligand in IL is driven by a larger entropic gain of the ligand in the bulk IL aqueous mixture. [EMIM$^+$] ions interacting with TRP62/63 and the active site residues could also be a reason why ligand binding is less stable in ILs and explain the dramatic reduction of lysozyme activity observed in experiments when in high IL concentrations.

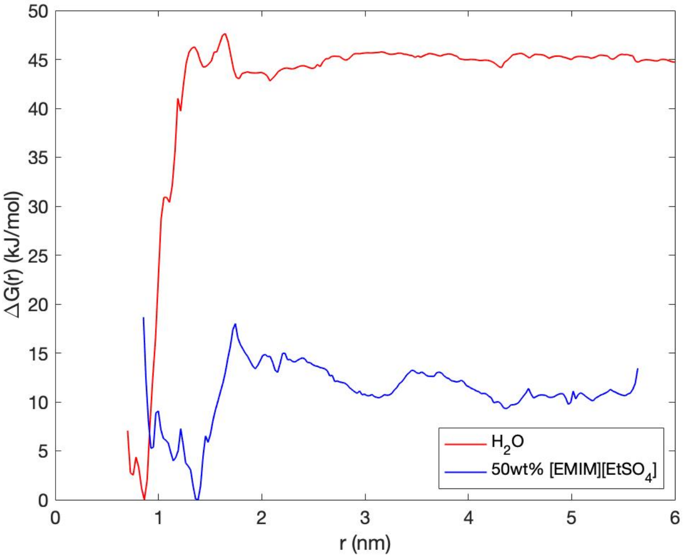


Figure 2.12: Free energy profile ΔG(r) of ligand in water (red) and 50 wt% $[\text{EMIM}]^+[\text{EtSO}_4]^-$ (blue) for moving the ligand away from the lysozyme active site in the pulling direction.

Next we examine the behavior of [EMIM$^+$] ions when undergoing rehydration. Figure 2.14 shows the time series of the [EMIM$^+$] ions near the non-native ligand binding site in both 50 wt% and rehydrated IL systems. In the 50 wt% IL systems, although the number of [EMIM$^+$] fluctuates, in average there are around two [EMIM$^+$] ions near the ligand binding position. However, for the rehydrated systems, within the first 50 ns, the number of [EMIM$^+$] already decreases to one or two ions in both IL systems. After approximately 200 ns the number of [EMIM$^+$] oscillates between 0 and 1. Two other replicas for the rehydrated system for each IL are performed and from all replicas the fast leaving of [EMIM$^+$] is observed. The diffusion of the [EMIM$^+$] ions indicates that the interactions between [EMIM$^+$] and the neighborhood of the ligand position are not strong enough to keep the [EMIM$^+$] ions from leaving the active site in the rehydrated conditions. Figure 2.13 also presents the potential of mean force ΔG(r) for an [EMIM$^+$] ion in the active site, where r is the distance of the [EMIM$^+$] ion from the lysozyme active site along the pulling direction in a hydrated system. [EMIM$^+$] ions have a lower free energy in the bulk solution comparing to that in the active site explaining the leaving of the cations.

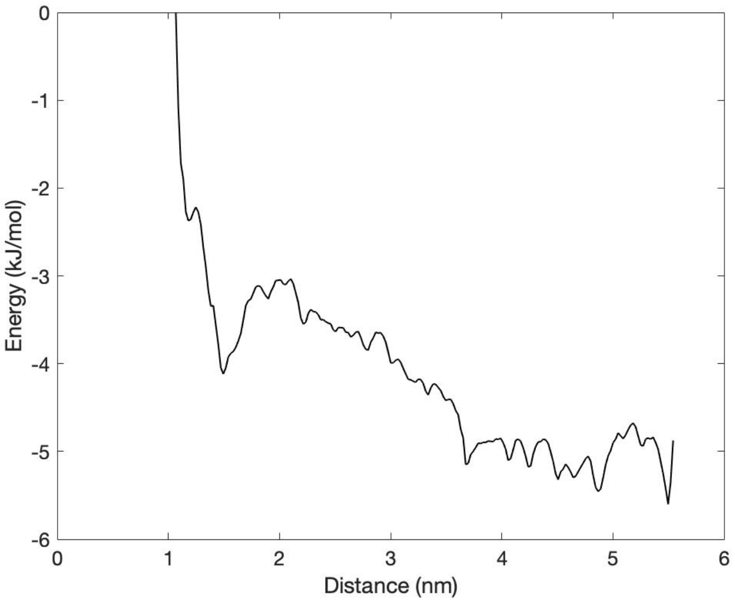


Figure 2.13: Umbrella sampling simulations showing the PMF plot for the distance of an [EMIM$^+$] molecule away from the lysozyme along the pulling direction in a hydrated system.

The re-entrance of the [EMIM$^+$] ion into the ligand binding region after 200 ns is not because of the narrow selection of the cutoff. The rehydrated systems for both ILs are very dilute (less than 1 wt%) and [EMIM$^+$] can diffuse far away from lysozyme. There is no specificity of this one ion, different [EMIM$^+$] ion enters the active site region and leaves. The re-entrance of [EMIM$^+$] ions interact with the active site residues (Figure 2.15(a)) or TRP62/63 (Figure 2.15(b)). The interaction between [EMIM$^+$] and active site residues are mainly electrostatic, while the interaction between [EMIM$^+$] and TRP62/63 could be cation-π and hydrophobic effects. However, the re-entrance of the cations is quite short-lived. The fast leaving of [EMIM$^+$] ions could explain why after rehydration the lysozyme activity can be fully recovered in experiments.

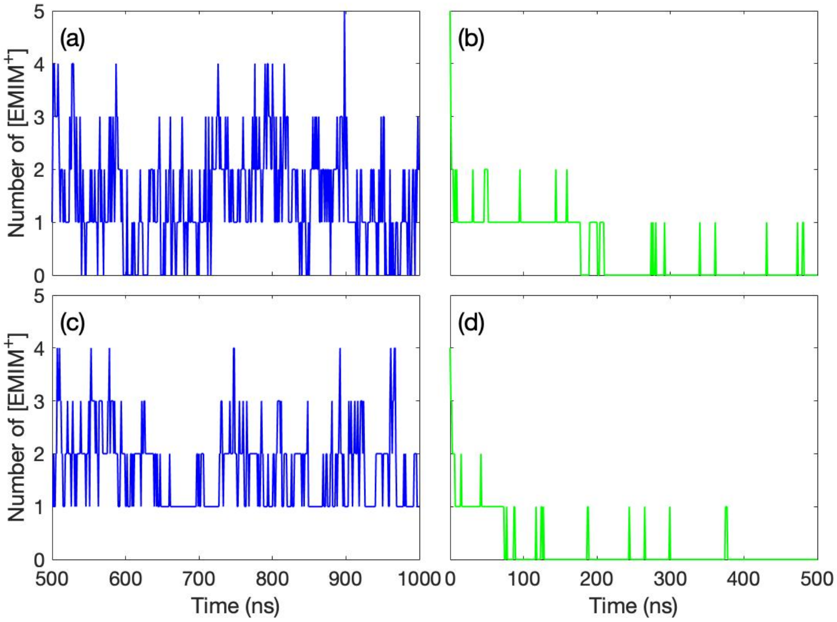


Figure 2.14: The time series showing the number of [EMIM$^+$] ions near the non-native ligand binding site, where (a) is the 50 wt% $[EMIM]^+[EtSO_4]^-$ system, (b) is the rehydrated $[EMIM]^+[EtSO_4]^-$ system, (c) is the 50 wt% $[EMIM]^+[Et_2PO_4]^-$ system, and (d) is the rehydrated $[EMIM]^+[Et_2PO_4]^-$ system.

The results here point to a robust behavior for lysozyme in these two imidazolium ILs. Part of this behavior can be explained by the unusually high stability of lysozyme itself, which has

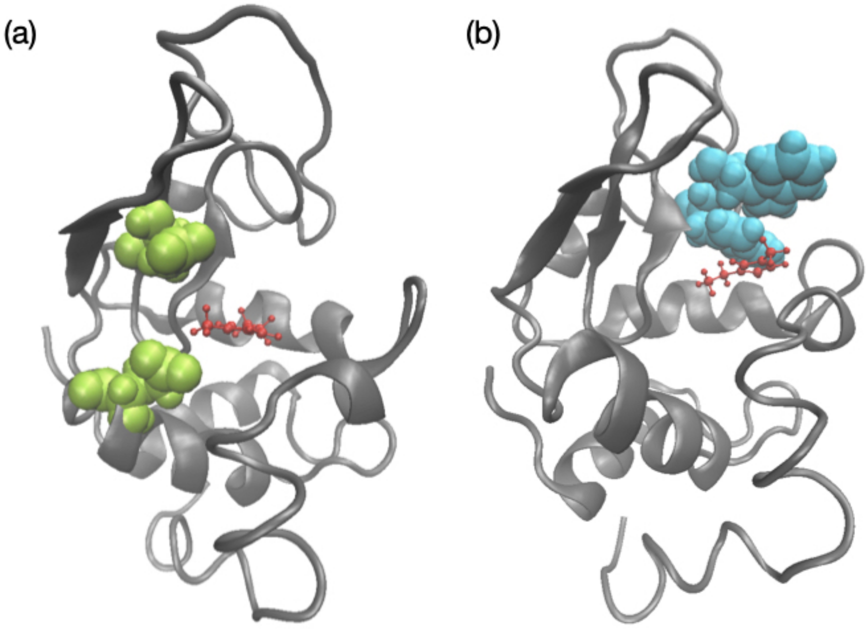


Figure 2.15: Illustation showing the close contact between an [EMIM$^+$] molecule (red) and (a) the active site residues (green) (b) TRP62/63 residues (cyan).

been known for decades [129]. Both ILs have different interaction strengths with lysozyme, with [EMIM]$^+$[EtSO$_4$]$^-$ being the weaker of the two. As a consequence, this IL perturbs lysozyme structure, and activity, to a smaller degree. Nevertheless, with the addition of sufficient water the protein's enzymatic activity is almost completely restored.

2.5 Conclusion

From the first half of the chapter the interactions between lysozyme and two ILs, [EMIM]$^+$[EtSO$_4$]$^-$ and [EMIM]$^+$[Et$_2$PO$_4$]$^-$, are investigated with a combination of experiment and simulation. We find that lysozyme is destabilized in the presence of ILs due to contributions from both the cations and anions, acting in local and global manners, respectively. The [EMIM]$^+$ cation exerts its effect locally by binding to tryptophan and presumably competing with a native Arg-Trp-Arg bridge that is critical to lysozyme stability, via favorable π-π and cation-π interactions. The anions [EtSO$_4$]$^-$

and $[Et_2PO_4]^-$, exert their effects globally through electrostatic effects, with the latter having the stronger short-range interaction. The differences due to the anions are likely due to the lower charge delocalization and greater hydrophobicity of $[Et_2PO_4]^-$. The protein surface heterogeneity is reflected in IL patterning, indicating no universal IL solvent for ILs, but rather, the need for each protein to have tailored ILs.

From the second half of the chapter, circular dichroism data and simulations both show that lysozyme structure is preserved after rehydration from high IL content, even after prolonged storage times (i.e., approaching one year). From the RMSF analysis, the suppressed protein structure fluctuations in 50 wt% ILs are able to recover back to higher fluctuations after rehydration. In addition, the average structure analysis also shows the deviation of the region I position from the aqueous structure (close to the active site) in 50 wt% ILs, and this deviation is able to restore back in the rehydrated systems. This is an important finding given the current interest in IL-based storage media to ”break the cold chain” for protein therapeutics [5].

Activity data and simulations both show that lysozyme’s enzymatic function is suppressed at high IL content but is recovered upon sufficient rehydration with water. From simulations it is observed that $[EMIM^+]$ ions occupying the ligand position and the vicinity of the active site. The ligand-lysozyme complex in 50 wt% $[EMIM]^+[EtSO_4]^-$ has lower binding-unbinding energy than that in water as shown by umbrella sampling simulations, indicating that the ligand-lysozyme complex is less stable in the presence of ILs. These results explain the reduction of lysozyme activity when in high IL content. The presence of $[EMIM^+]$ near the active site leading to loss of protein activity agrees with several other works that concern different protein/ILs systems [40, 130]. When undergoing rehydration, the fast leaving of $[EMIM^+]$ ions is observed for both IL systems, which explains the reappearance of the active site availability because there is no more cations present near the active site after rehydration. This result verifies the concept of recovering functional proteins from IL media.

Chapter 3: Transferable and polarizable coarse grained model for proteins (proMPT)

3.1 Overview

This chapter is based on the author's publications: (1) Transferable and Polarizable Coarse Grained Model for Proteins - ProMPT. Pei-Yin Lee, Abhilash Sahoo, and Silvina Matysiak. Journal of Chemical Theory and Computation, 2022. [131] Pei-Yin Lee performed all the simulation validations and analyses. Abhilash Sahoo performed model parametrization. Pei-Yin Lee and Abhilash Sahoo contributed equally to the writing of manuscript. (2) Modulation of helical conformation of Glycophorin A by point mutations. Pei-Yin Lee, Abhilash Sahoo, and Silvina Matysiak. Submitted. Pei-Yin Lee performed all the simulation validations and analyses. Abhilash Sahoo performed model parametrization. Pei-Yin Lee and Abhilash Sahoo contributed equally to the writing of manuscript.

In this chapter I will first introduce the development of an in-house coarse-grained (CG) model and follow up with several applications in different solvent environments. The application of classical molecular dynamics (MD) simulations at atomic resolution (fine-grained level - FG), to most biomolecular processes, remains limited because of the associated computational complexity of representing all the atoms. This problem is magnified in the presence of protein-based biomolecular systems that have a very large conformational space and MD simulations with fine-grained resolution have slow dynamics to explore this space. Current transferable CG force fields in literature are either limited to only peptides with the environment encoded in an implicit form or cannot capture transitions into secondary/tertiary peptide structures from a primary sequence of amino acids. Here, my collaborator Abhilash Sahoo and I present a transferable CG force field with an explicit representation of the environment for accurate simulations with proteins. The force field consists of a set of pseudo-atoms representing different chemical groups that can be joined/associated together to create different biomolecular systems. This preserves the transferability of the force field to multiple environments and simulation conditions. We have added electronic polarization that can respond to environmental heterogeneity/fluctuations and couple it to protein's structural transitions. The non-bonded interactions are parametrized with physics-based features such as solvation, and partitioning free energies determined by thermodynamic calculations and matched with experiments and/or atomistic simulations. The bonded potentials are inferred from corresponding distributions in non-redundant protein structure databases.

We present validations of the CG model with simulations of well-studied aqueous protein systems with specific protein fold types- TRP-cage, TrpZip4, Villin, WW-domain and β-α-β. We also explore the applications of the force field to study aqueous aggregation of Aβ 16-22 peptides. In addition, we examine the dimerization of Glycophorin A (GpA) and its mutants in the presence of Dodecyl-phosphocholine (DPC) micelles. Dimerization of transmembrane helices is an essential structural feature for several physiological functions. From our simulations, the wild type (WT) GpA is observed to be a right-handed dimer with specific GxxxG contacts, which agrees with experimental findings. Specific point mutations reveal several features responsible for the structural stability of GpA. While T87L mutant forms anti-parallel dimers due to an absence of T87 interhelical hydrogen bond, a slight loss in helicity and a hinge-like feature at the GxxxG region develops for the G79L mutant. We note that the local changes in the hydrophobic environment, effected by the point mutation, contribute to the development of this helical bend. This application of *ProMPT* presents an holistic overview of structural stability of GpA in micellar environment, while taking secondary structural fluctuations into account. Moreover, it presents opportunities for applications of computationally efficient CG models to study conformational alterations of transmembrane proteins that have several physiological relevance.

3.2 Introduction

Physiological functions of protein molecules are closely intertwined with their associated structure and dynamics [132, 133]. This complex macro-organization is shaped through microscopic multi-body interactions that define a protein molecule's conformational landscape. In this direction, computer simulations, primarily MD simulations are being increasingly leveraged to understand protein biochemistry and biophysics [134, 135, 136].

With recent advancements in dedicated high performance computing architectures and graphical processing unit, it is now possible to run long simulations with small proteins for multiple microseconds, at conventional all-atom resolution [137, 138, 139, 140]. Such long molecular simulations could capture a small number of protein folding-unfolding transitions in an unbiased manner. But such processes require significant amount of computing resources and are still not scalable to larger systems. Most biological systems of interest involve large biomolecules that interact over long spatio-temporal scales. To alleviate some of these shortcomings, several enhanced sampling approaches such as metadynamics and replica exchange MD have been proposed to allow faster exploration of conformational space with limited resources [141, 142, 143, 144]. But such approaches require extensive knowledge about the particular biomolecular system and/or can be infeasible for larger systems.

CG-MD involves creating a simplified representation or minimal model of biomolecules that can capture the essential biophysics [145]. This approach allows efficient access to long spatio-temporal scales by directly reducing computational complexity and allowing fast conformational transitions by smoothing the local free energy landscape. Early CG models with simplified phenomenological potentials were instrumental in establishing the foundations of energy landscape theory of protein folding [146, 147, 148]. The CG models require creating interaction sites that are representative of a particular molecule and defining interaction schemes (potentials) that allow these interaction sites to communicate. The interaction potentials, also known as *force fields* can be variations of knowledge-based (developed from analysis of statistical databases) or physics-derived potentials (created for example to fit free energies, partition coefficients, and biomolecular phases.) [149, 150, 151, 152, 153, 154, 155, 156]. CG molecular simulations of proteins have played a significant role in shaping our understanding of physiological processes such as protein folding, protein aggregation and membrane-protein interactions [148, 157, 158, 159, 160, 161, 162, 163].

CG models with varying molecular resolutions and diverse coarse graining strategies have been proposed. The levels of coarse-graining can vary from multiple-residues represented by a single interaction site, to models with representation for all the heavy atoms in a biomolecular system. Several of the CG models employ an implicit description of the solvent environment through interaction potentials [152, 153, 156, 164, 165, 166, 167, 168, 169]. These models have been used to study several protein-based dynamical processes such as folding, adsorption, misfolding and aggregation. While these models provide a significant reduction in system complexity, an implicit representation of the environment cannot be used to study heterogeneous environmental effects that can be essential in crowded physiological systems. Moreover, the role of solvent in governing thermodynamics and kinetics of protein folding is well documented [170, 171, 172]. In some of the implicit-water CG models, the dynamics is biased towards native state, and cannot capture non-native states [147, 173, 174, 175].

MARTINI is a popular CG force field with explicit description for solvents and a modular architecture [150, 176]. The interactions are directed through a set of interaction-levels created by leveraging environment-dependent free energies (solvation, vaporization and partitioning free energies). Due to its reasonable accuracy and ease of use, particularly in heterogeneous and crowded environment, this force field has been widely adopted by the natural science and engineering communities. In the case of proteins, MARTINI has been used to study ligand-binding, aggregation, surface-absorption and membrane translocation [177, 178, 179, 180, 181, 182]. However the model relies on restraining protein's secondary structures through artificial potentials, and the global/tertiary structure also needs to be restrained to prevent spontaneous unfolding. This prevents any study of dynamical changes to protein's conformation over the simulation time using MARTINI. While G$\bar{o}$ model has been employed along with the MARTINI force field to study large-amplitude conformational changes, it can suffer from several shortcomings including loss of amino-acid identity, insensitive to point-mutations and environmental changes [157, 183, 184, 185].

Previous group publication introduced an explicit-solvent polarizable CG model for a selection of amino-acids to study secondary structure transitions starting from primary sequence in presence of different environmental stimulus such as hydrophobic media, interfaces and lipid bilayers [186, 187, 188, 189, 190]. Here we have formalized our approach to model parametrization, introduced new bonded and non-bonded potentials, updated residue geometry; and extended the representation to all natural amino acids to capture accurate protein tertiary structures.

Our Protein Model with Polarizability and Transferability (*ProMPT*) consists of two types of interaction sites (beads) — primary CG beads and dummy beads. The basic biomolecule structure is created with primary interaction sites which feature modular architecture and geometry similar to the MARTINI model that allows for easy transferability; and are parametrized along the MARTINI scales [150, 191]. Through additional off-center dummy charges to the primary sites that represent polarizable entities, we have introduced explicit local dipole moment, that can result in anisotropic Coulombic interactions similar to hydrogen bonds in protein's secondary structure. The angle and dihedral potentials are derived from statistical distributions of these features from the protein data bank (PDB). *ProMPT* can capture secondary and tertiary conformational transitions, along with appropriate intermediate conformations and accurate folding free energy profiles. As such, this CG model can be applied to study spatiotemporally complex biological phenomena and processes involving proteins. We will first present model validations using simulations of small protein structures in aqueous solvent and aqueous protein aggregation, and then explore further the application of this model in the GpA dimer system.

Integral membrane proteins play critical roles in signal transduction and transport across cell membranes. The structure of these membrane proteins primarily feature tertiary structural arrangements such as α-helical and β-strand bundles or their combinations, with membrane spanning helix bundles being the dominant architecture [192]. Therefore, structure and relevant design implications of α-helix bundles is of significant interest to the research community to study their role in several physiological and pathological functions. Experimental and computational studies suggest that the formation of these membrane-spanning transmembrane helix bundles follow a two step pathway [193]. First, the protein folds and inserts into the membrane, that allows it to bury the hydrophobic sidechains into the acyl core of the membrane. Then, the helix association occurs in this membranous environment, driven by a complex interplay of electrostatic and van-der-Waals' effects. The folding of a solvated protein into a helical patch results in a free-energy gain, when the hydrogen-bonded polar backbone gets inserted into the membrane. GpA, found in human erythrocytes, is a well studied benchmark for the studies of transmembrane α-helices because of their functional relevance. The structure of GpA in solution NMR and solid state NMR have been well characterized [194, 195]. GpA is a homodimer packing in a right handed fashion, where the contact surface between the two helices consist of an important motif GxxxG that is found to be critical in helix dimerization [196, 197]. In order to understand the specific contacts allowing the formation of a stable helix dimer, mutational studies are essential as these studies enable us to locate interactions that are critical for dimer formation.

An early mutagenesis experiment, before the solution NMR structure of GpA dimer was solved, had already established that GpA dimerization in detergent micelles is spontaneous and highly specific. Sensitive positions for mutations that would affect dimer stability were identified to occur every 3.9 residues, roughly co-located at the dimerization interface [198]. In addition, L75, I76, G79, G83, V84, and T87 were further identified as the most affected positions for point mutations. An alanine-scanning mutagenesis study further re-iterated that mutating residues at the dimer-interface led to significant disruption of dimer stability, especially at the "GxxxG" motif [199]. In this study, the authors also confirm that the hierarchy of the point mutation stability for GpA is independent of hydrophobic environments. Another FRET experiment has shown that for GpA in detergent micelles, helix formation is uncoupled from helix dimerization for both the WT and mutant G79LG83L [200]. In this study, both the WT and mutant exhibit alpha helical signature in a far-UV circular dichroism spectra despite their dimerization status.

These experimental results have suggested that the specific interactions occurring at dimer surface are critical, and that designed mutants can modulate dimer stability. Fast local structural and environmental fluctuations can be crucial to self-association of GpA helices, particularly in a micellar environment. These dynamic effects can be significantly difficult to characterize through experimental methods because these methods often provide averaged information. On the other hand, MD simulation methods can be leveraged to identify dynamic alterations that shape protein structure and dimerization events.

CG simulations with the MARTINI forcefield of GpA in DPPC bilayer supported dimerization of GpA-helices with a right handed crossing angle, with GxxxG contacts [201]. Disruptive mutants still dimerized but were reported to have a more diffusive contact surface compared to that of WT and non-specific interactions [202]. These results were in-line with the experimental finding that the GxxxG motif is critical for GpA dimerization. These results were instrumental in a molecular explanation of dimer interface, and the mechanisms that contribute to disruption of association.

It is necessary to note that these CG simulations apply external restraints to restrict fluctuations of protein's secondary structure and maintain the α-helix structure for individual GpA monomers. Since, the helicity is fixed in these CG models, it is impossible to study implications of structural undulations on dimer stability. Conformational changes of transmembrane proteins can be important in enzyme activity. For example, it is reported that the structural and dynamical features of the helical transmembrane domain from the amyloid precursor protein (APP) are important in the proteolytic processing by γ-secretase [203]. Mutation on the GxxxGxxxG motif affects the helicity and changes the hydrophobic environment of the protein. In addition, the cleavage point is also affected [203, 204]. These conformational variances can not be studied with CG simulations that employs secondary structure constraints. Therefore, we here study the impact of specific point-mutations on GpA folding and dimerization with our newly developed CG model. *ProMPT* can record accurate local environmental stimulus and institute protein's structural transitions. This forcefield can be instrumental in deducing relationships between GpA/mutant's secondary structure and dimerization. For the GpA dimerization application, we will first examine the performance of *ProMPT* on reproducing the WT results, and then we will continue with detailed characterization for two disruptive mutants, T87L and G79L. We will also discuss the potential reasons on why these mutants are less stable than the WT GpA.

3.3 Methods

The CG interaction sites follow a basic *MARTINI-influenced* nomenclature, with atom-types grouped into polar, neutral, charged and hydrophobic beads [150, 191]. The polar beads have explicit electrostatic dipoles added through charged dummy particles, constrained to the main interaction site (Figure 3.1). For parametrization, we initially run a 50 ns atomistic simulations (using CHARMM36m/TIP3P force field [205, 206] and CHARMM-GUI [207] equilibration protocol) of relevant tripeptides that have the corresponding mapped regions in chemical space. The parameters (charge on the beads - q and distance between the beads - r) are generated to match the dipole moment corresponding to the maximum in dipole moment distribution of the mapped region from the atomistic simulation. The charge and relevant geometry of these charged dummies are provided in Table 3.1. The dummy charges interact with the local environmental through electrostatic interactions (such as change in dielectric constant at membrane-water interface), which then directs protein's structural changes. As these charges are placed off-center to the main interaction site, they can introduce an asymmetry and directionality to interactions, and introduce spatial heterogeneity in local charge distribution. Here, a parallel can be drawn between the direct dipole-dipole and charge-dipole interactions in this CG model and electrostatic alignments in hydrogen bonds. Previous publications with a much simplified variant of this CG model, could generate appropriate sequence-specific secondary structures through alignment of these dipolar charges [186, 187, 188, 189, 190]. Several experiment and simulation-based studies have also underlined the importance of these molecular dipoles in protein folding.

ProMPT has been parametrized to work with the Yesylevsky *et al.*'s polarizable water model, which maps four atomistic water molecules to three CG beads [208]. This water model also features dummy charges that communicate exclusively through electrostatics, and can interact with our protein model (Figure 3.2).

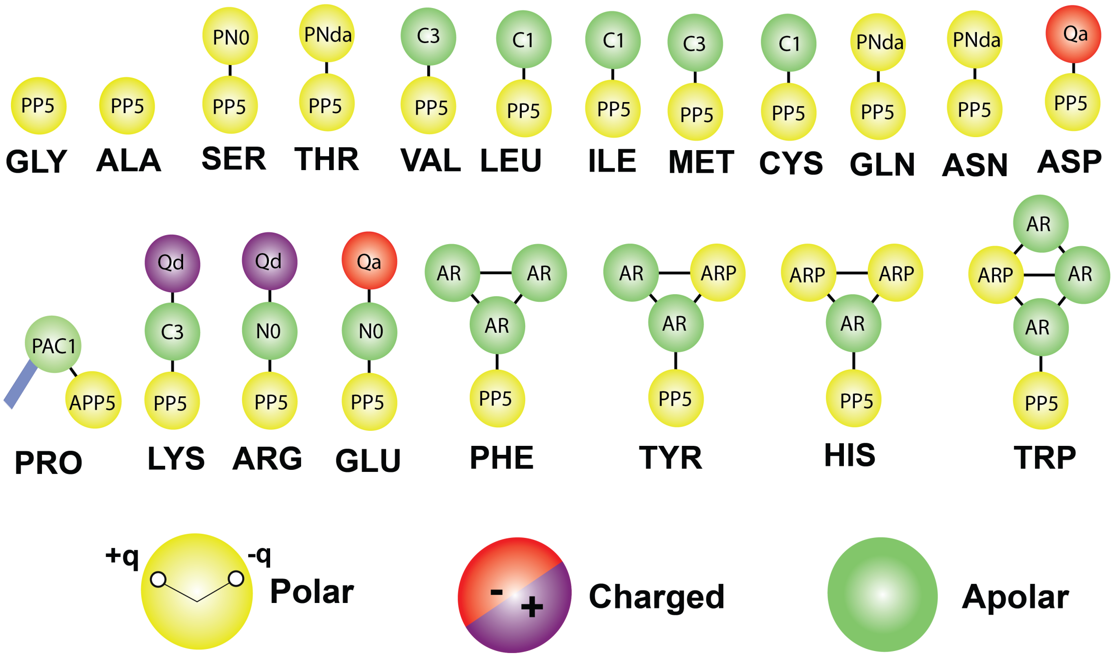


Figure 3.1: Schematic geometries of CG amino-acids. The bead type is labeled on each bead.

The amino-acid geometry defining the structures of the CG amino-acids are shown in Fig-

Table 3.1: Charges and characteristic bonded potentials for dummy beads. Bond-length is the length of the tether from the primary interaction-center to the charged dummies. k_{angle} is the spring constant preventing deviation of the angle between charged dummies and the primary-interaction-center from 180 degrees.

Dummy Types	Charge	Bond-length (nm)	k_{angle} (°)
Backbone	± 0.340	0.14	7.2
SER-Sidechains	± 0.144	0.14	7.2
THR-Sidechains	± 0.153	0.14	7.2
GLN-Sidechains	± 0.256	0.14	7.2
ASN-Sidechains	± 0.256	0.14	7.2
TYR-Sidechains	± 0.138	0.14	7.2
TRP-Sidechains	± 0.136	0.14	7.2

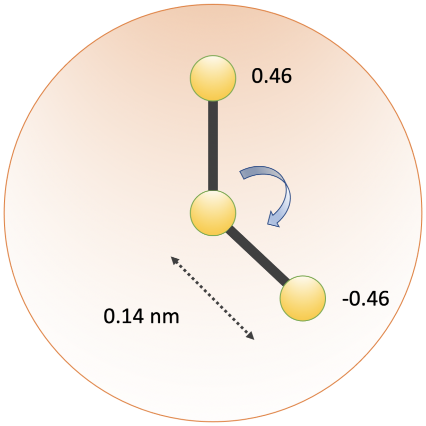


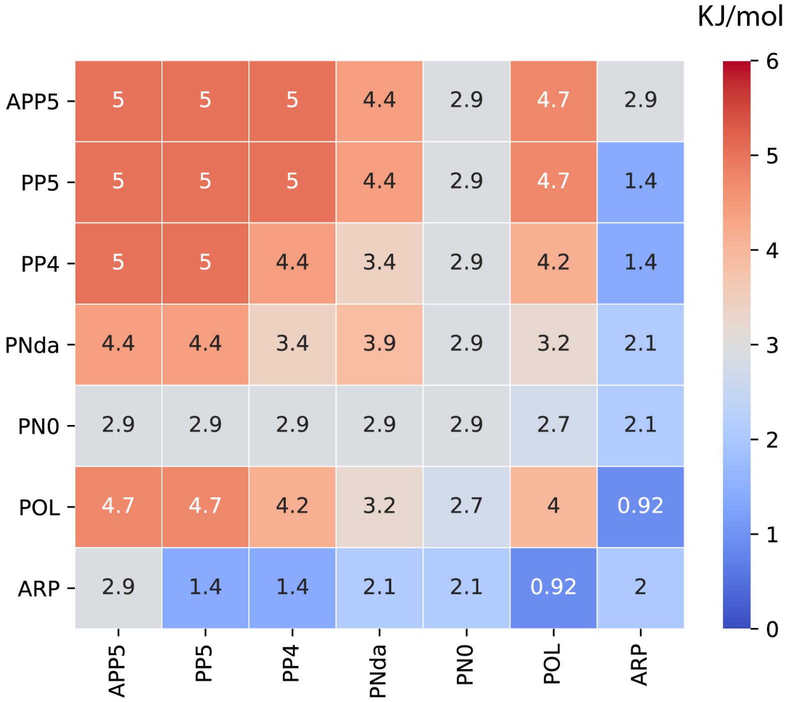
Figure 3.2: CG geometry of polarizable Martini water. The partial charge on each dummy is labeled.

ure 3.1. The interaction sites representing protein backbone and polar sidechains are assigned polarizable atom-types, with off-center dummy charges. The primary CG interaction sites that map to aromatic regions in the chemical space have a radius of 0.43 nm, compared to 0.47 nm for all other interaction sites.

3.3.1 Non-Bonded Interactions

All the non-bonded LJ type interactions between the atom-types are parametrized along the MARTINI interaction levels, which allow for easy transferability, and can therefore be used with all biomolecular environments that can be represented with the MARTINI force field. A complete description of the non-bonded interaction parameters is presented in Figure 3.3-3.8. While, most of the cross-interactions between non-polarizable interaction sites are directly borrowed from the MARTINI force field, the interactions between our polarizable groups had to be re-parametrized to balance out the added electrostatic interactions from the charged dummies. Similar to the MARTINI force field, the parametrization here aimed to fit non-bonded interaction parameters to reproduce accurate free energies of solvation and partitioning. The interaction level between hydrophobic groups and the interaction sites representing water is reduced by upto 50% from the MARTINI level to reproduce appropriate environment-induced structural transitions. These reductions balance out issues of over-polarization by backbone dipoles and allow conformational switching from helices to β-strands.

A set of special interactions were added in an *ad-hoc* manner to capture specific protein-protein interactions. We applied specific attraction (ϵ = 3.0 kJ/mol) between positively charged groups (such as ions and cationic amino acid sidechains) and the aromatic rings to mimic cation-π



	APP5	PP5	PP4	PNda	PN0	POL	ARP
APP5	5	5	5	4.4	2.9	4.7	2.9
PP5	5	5	5	4.4	2.9	4.7	1.4
PP4	5	5	4.4	3.4	2.9	4.2	1.4
PNda	4.4	4.4	3.4	3.9	2.9	3.2	2.1
PN0	2.9	2.9	2.9	2.9	2.9	2.7	2.1
POL	4.7	4.7	4.2	3.2	2.7	4	0.92
ARP	2.9	1.4	1.4	2.1	2.1	0.92	2

Figure 3.3: Non-bonded interactions between polar beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol.

effects [209]. Similarly, interaction between aromatic rings, and proline sidechains with aromatic rings were made attractive (ϵ = 3.0 kJ/mol) to capture π-π stacking and CH-π interactions that have been highlighted in quantum mechanical calculations [210, 211, 212]. These electronic effects are ubiquitous in protein folding. Additional dummy-dummy and dummy-charged group hard-core repulsion ($c_{12} \sim 10^{-7}$ nm) were added to prevent over-interaction between CG groups similar to methods to prevent *polarization catastrophe* in polarizable force fields [213, 214]. This allows for fast switching between conformations through binding/unbinding among charged dummies.

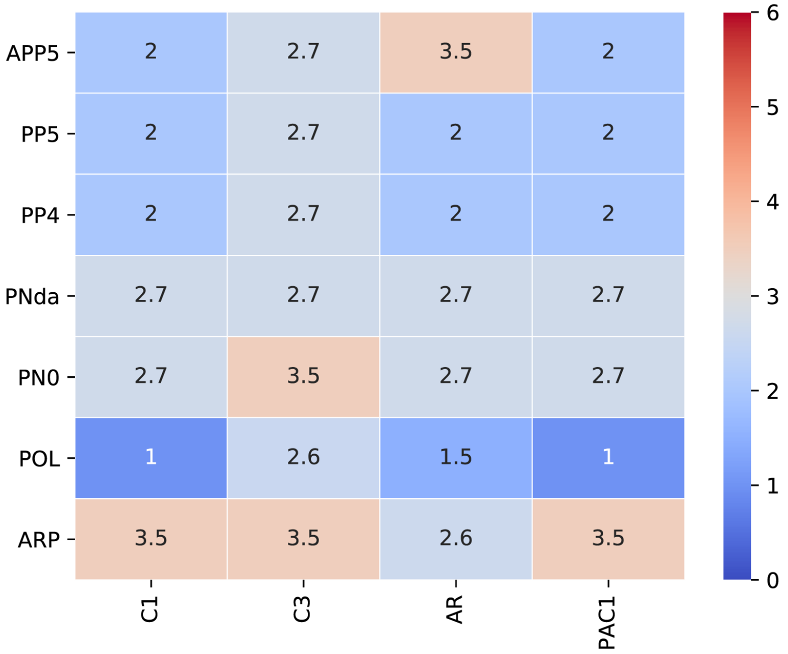
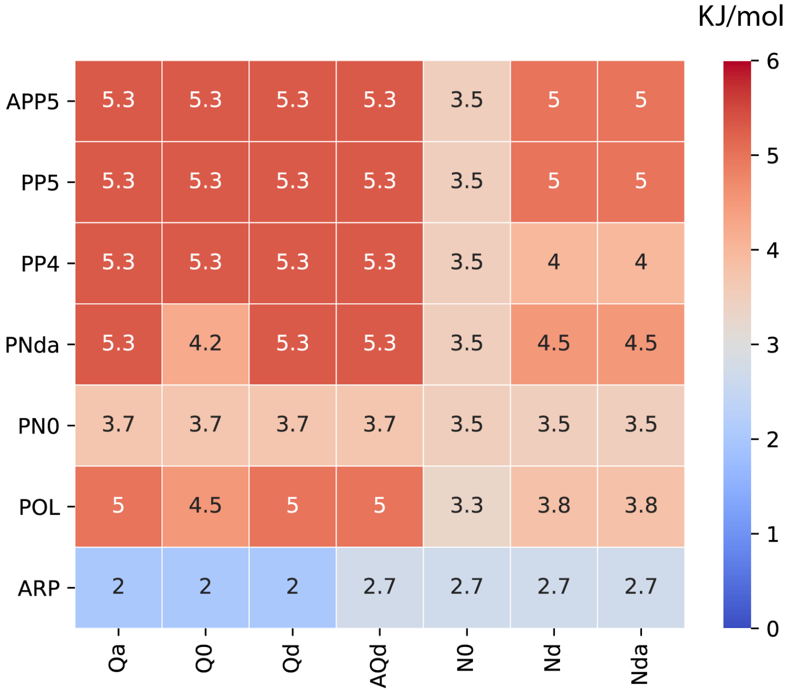


Figure 3.4: Non-bonded interactions between polar beads and hydrophobic beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol.



	Qa	Q0	Qd	AQd	N0	Nd	Nda
APP5	5.3	5.3	5.3	5.3	3.5	5	5
PP5	5.3	5.3	5.3	5.3	3.5	5	5
PP4	5.3	5.3	5.3	5.3	3.5	4	4
PNda	5.3	4.2	5.3	5.3	3.5	4.5	4.5
PN0	3.7	3.7	3.7	3.7	3.5	3.5	3.5
POL	5	4.5	5	5	3.3	3.8	3.8
ARP	2	2	2	2.7	2.7	2.7	2.7

Figure 3.5: Non-bonded interactions between polar beads and charged/neutral beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol.

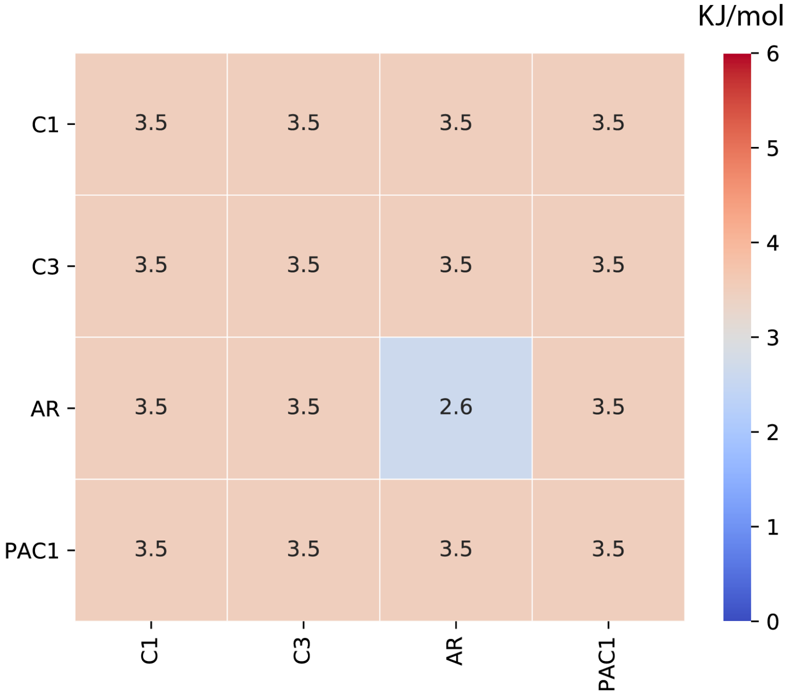
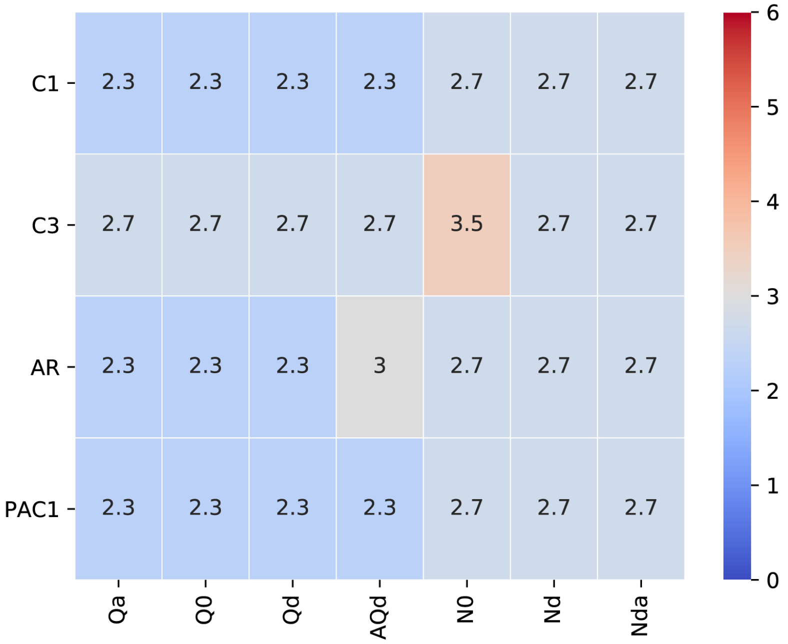


Figure 3.6: Non-bonded interactions between hydrophobic beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol.

3.3.2 Bonded Interactions

Bonded interactions can be broadly categorized as bonds, angles and dihedrals. The features describing bonds in *ProMPT* — bond lengths (Table 3.2) and bond rigidity were borrowed either directly from the MARTINI force field or the previous iteration of this force field [186, 188, 189]. The corresponding parameters are enumerated in Table 3.2-3.6. Some angular (between backbone beads (BB) and first side-chain) and dihedral potentials (backbone only) were informed from a statistical distributions of protein structures.

We used a non-redundancy p-value of 10^{-7} to create a database of about 14000 protein structures from the PDB. The angular and dihedral potentials were based on their corresponding normalized distributions from this database. The angle between the protein backbone sites was



	Qa	Q0	Qd	AQd	N0	Nd	Nda
C1	2.3	2.3	2.3	2.3	2.7	2.7	2.7
C3	2.7	2.7	2.7	2.7	3.5	2.7	2.7
AR	2.3	2.3	2.3	3	2.7	2.7	2.7
PAC1	2.3	2.3	2.3	2.3	2.7	2.7	2.7

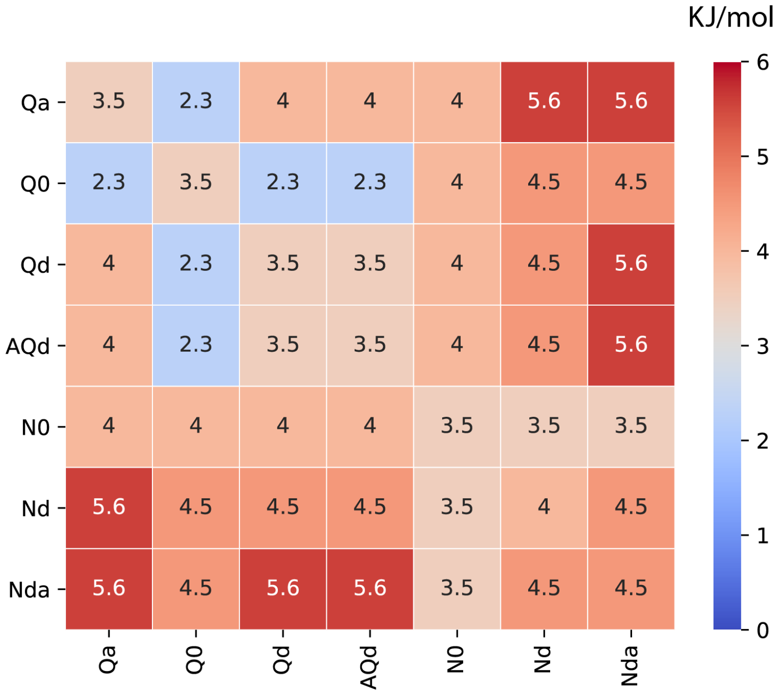
Figure 3.7: Non-bonded interactions between hydrophobic beads and charged/neutral beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol.

Table 3.2: Bond lengths between primary CG interaction sites.

Type	Bond-length (nm)
BB - BB	0.385
BB - Sidechain1 (S1)	0.28
Sidechain1 (S1) - Sidechain2 (S2)	0.25
Between aromatic rings	0.27

Table 3.3: Bonded interaction potentials between different primary CG interaction-centers. BB: Backbone; S1: First Sidechain; S2: Second Sidechain.

Bond-Types	Bond-Length (nm)	Bond-Rigidity/Constraints (kJ mol^{-1}
A. Between backbone interaction-centers	0.385	7500
B. Between BB-S1 of non-aromatic amino-acids	0.250	5000
C. Between S1-S2 of non-aromatic amino-acids	0.280	5000
D. Bonds in planar rings of aromatic residues	0.270	Constraints



	Qa	Q0	Qd	AQd	N0	Nd	Nda
Qa	3.5	2.3	4	4	4	5.6	5.6
Q0	2.3	3.5	2.3	2.3	4	4.5	4.5
Qd	4	2.3	3.5	3.5	4	4.5	5.6
AQd	4	2.3	3.5	3.5	4	4.5	5.6
N0	4	4	4	4	3.5	3.5	3.5
Nd	5.6	4.5	4.5	4.5	3.5	4	4.5
Nda	5.6	4.5	5.6	5.6	3.5	4.5	4.5

Figure 3.8: Non-bonded interactions between charged/neutral beads. Refer to Figure 3.1 for naming nomenclature. The value on the heatmap is the Lennord-Jones ϵ value in kJ/mol.

Table 3.4: Angular interaction potentials between different primary CG interaction-centers. BB: Backbone; S1: First Sidechain; S2/S3/S4: Second/Third/Fourth Sidechain.

Angle-Types	Angle (°)	Angular-Rigidity (kJ mol^{-1} nm^{-2})
A. Between backbone interaction-centers	109	75
B. Between BB-S1-S2 of non-aromatic amino-acids	151	25
C. Between BB-S1-S2/S3 for F Y	150	50
D. Between BB-S1-S2 for W	210	50
E. Between BB-S1-S3 for W	90	50

Table 3.5: Dihedral interaction potentials between different primary CG interaction-centers. BB: Backbone; S1: First Sidechain; S2/S3/S4: Second/Third/Fourth Sidechain.

Angle-Types	Dihedral (°)	Dihedral-Rigidity (kJ mol^{-1} nm^{-2})
A. Between backbone interaction-centers	Tabulated	NA
B. Between BB-S2-S3-S1 for F Y	Improper; 0	50
C. Between BB-S2-S3-S1 for W	Improper; 0	50
D. Between S1-S2-S4-S3 for W	Improper; 0	200

Table 3.6: Dihedral potential parameters for α-helix and 3-10 helix corresponding to equation 1 in the main text.

Secondary-Structure	A	B	C	D	ϕ_0
α-helix	1.799	1.969	-1.199	-1.03	1.227
3-10 helix	1.799	1.969	-1.199	-1.03	1.969

universally set to 109^o, restrained through a harmonic potential. The angular potential between the backbone and the sidechain was created specific to each amino-acid by applying Boltzmann inversion at 300 K to each amino-acid specific distributions. These potential energy functions corresponding to each amino-acid have been reported in Figure 3.9-3.11. Finally, the dihedral potentials between the backbone interaction-centers are secondary-structure specific, with different tabulated potentials for α-helix, 3-10 helix and β-sheets. The functional forms of α-helix and 3-10 helix potentials were taken from previous publications of Cα based CG models [215].

$$\mathrm{V}(\phi; A, B, C, D) = A\left[1 + cos(\phi + \phi_0)\right] + B\left[1 + cos(\phi - \phi_0)\right] + C\left[1 + cos(3(\phi + \phi_0))\right] + D\left[1 + cos(\phi + \phi_0 + \tfrac{\pi}{4})\right]$$

The fitted parameters are provided in Table 3.6. These periodic functions allow for introducing multiple local/global minimums, in contrast to single and deep global minima by Boltzmann inversion. We parametrized these functions to set the global functional minimum to match the maximum value in the dihedral probability distributions of relevant secondary structures, in addition to introducing several local minima. These local extremes introduce additional frustrations to the protein's conformations and has been adopted in previous CG models [215]. The tabulated β sheet potential is generated through a direct Boltzmann inversion (at 300K) of dihedral distributions from regions specific to beta sheets in the protein structure database. Figure 3.12 shows the dihedral potentials employed in *ProMPT*.

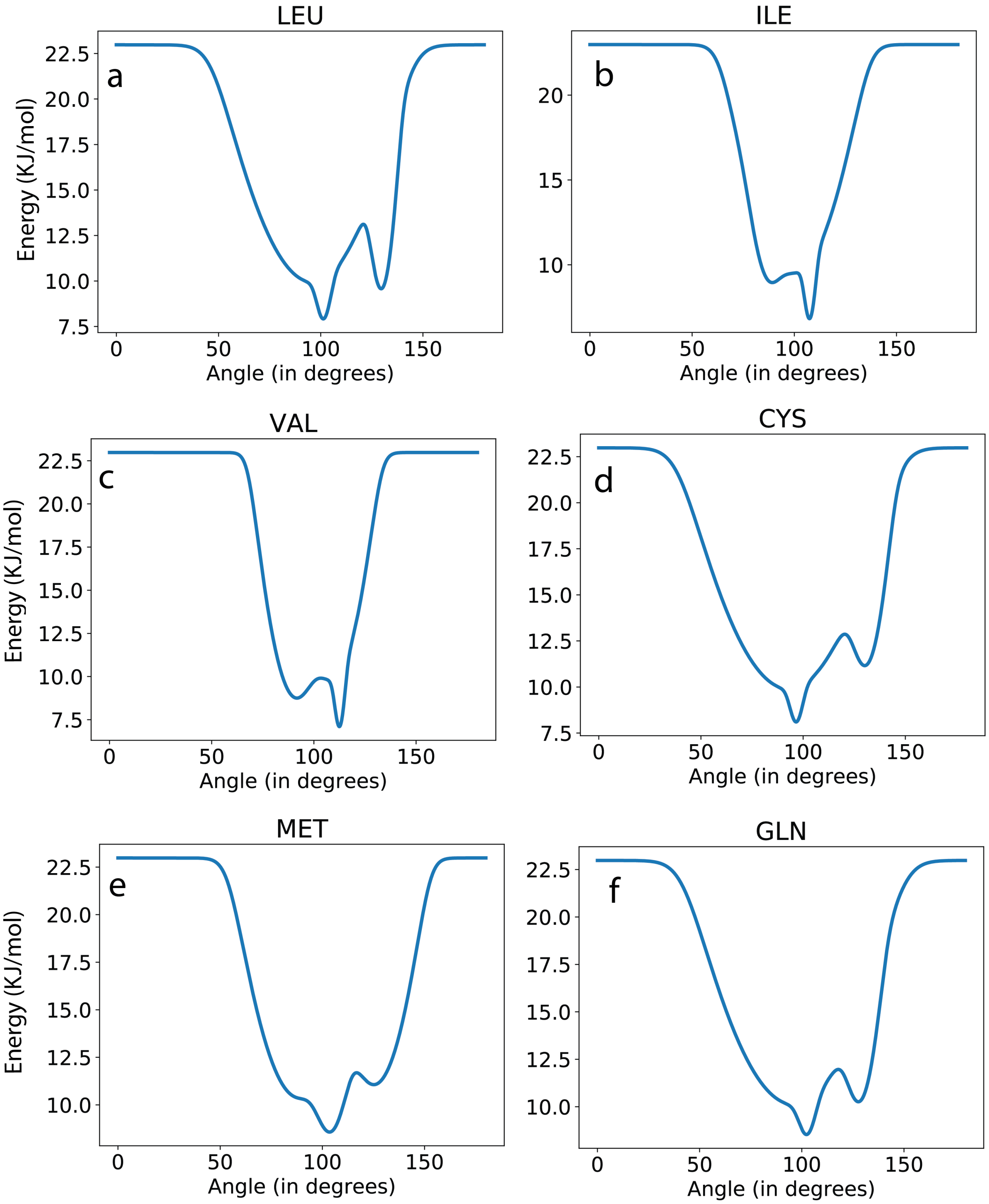


Figure 3.9: BB (previous amino acid)-BB-S1 tabulated angular potentials for each amino acid in kJ/mol.

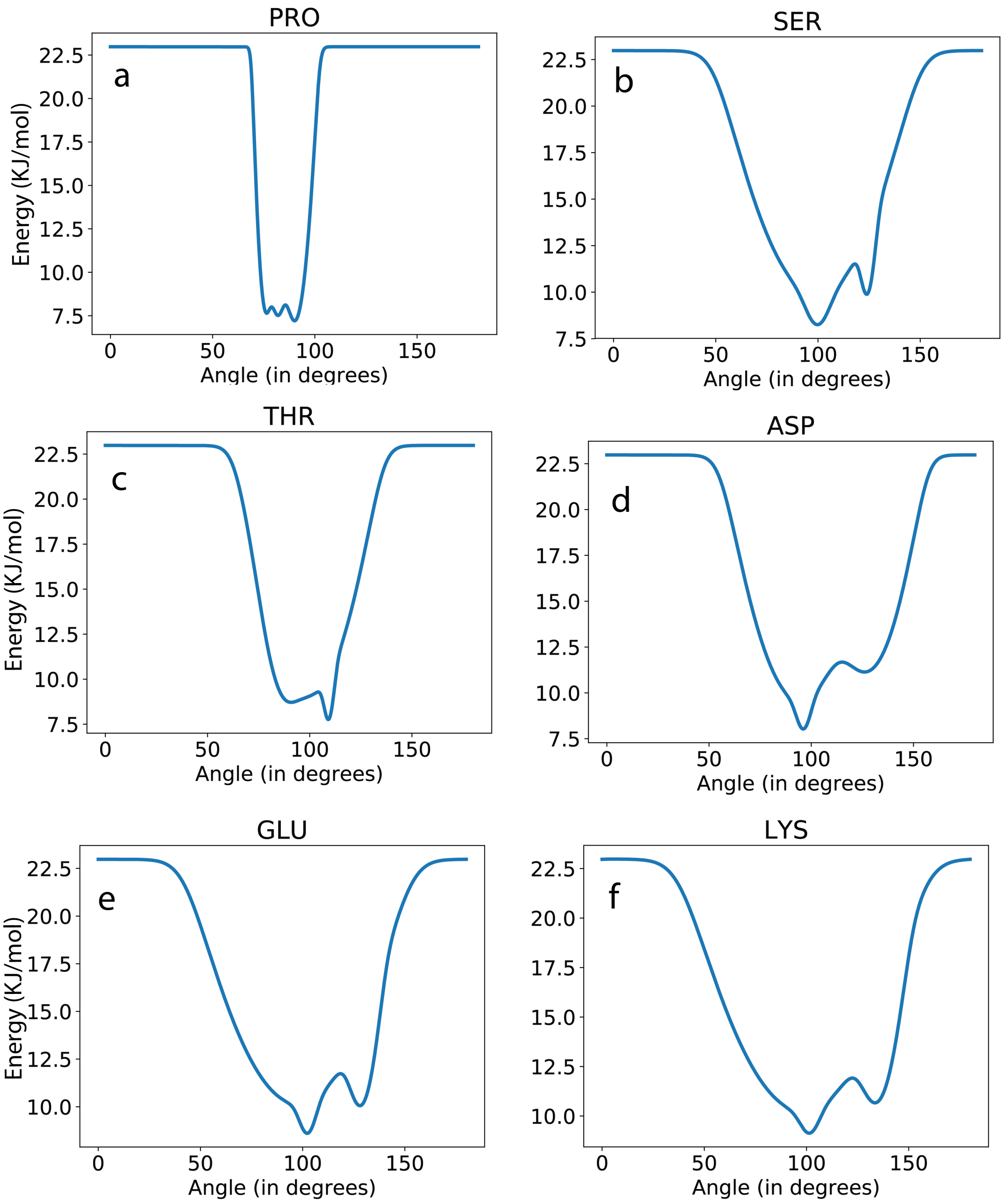


Figure 3.10: BB (previous amino acid)-BB-S1 tabulated angular potentials for each amino acid in kJ/mol.

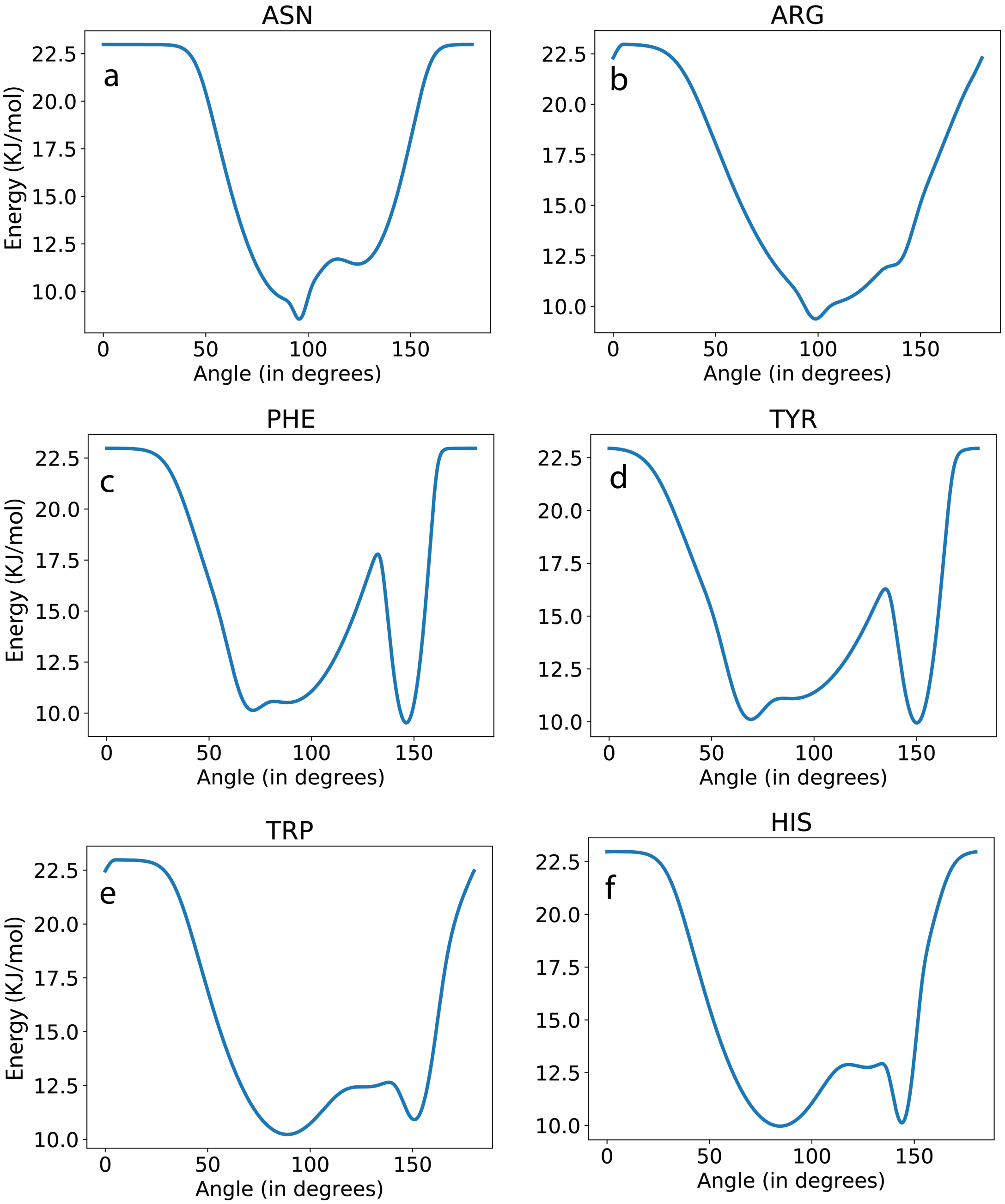


Figure 3.11: BB (previous amino acid)-BB-S1 tabulated angular potentials for each amino acid in kJ/mol.

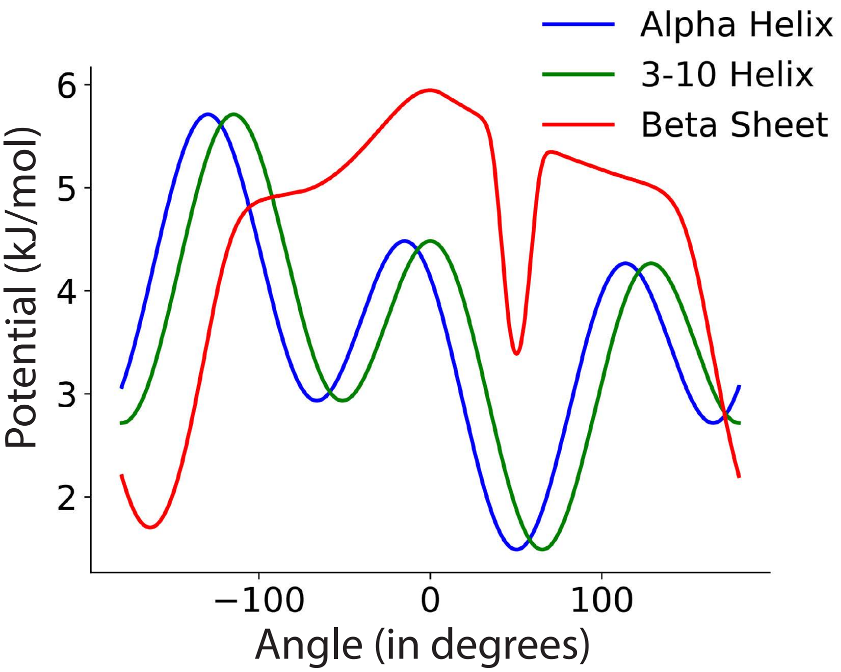


Figure 3.12: Secondary-structure specific dihedral potential used in the CG forcefield. The tabulated potentials are fitted (for α-helix and 3-10 helix) or derieved (β-sheet) to capture maximum value in the dihedral probability distributions

3.3.3 Simulation setup

3.3.3.1 Setup for mini-proteins

All simulations are performed with GROMACS 2019.4 [216]. The starting conformation and the protein topology are generated with in-house codes, taking the amino acid sequence from the reference PDB structures listed in Table 3.7. The initial conformation for all proteins is a random unfolded extended conformation. The setup for each protein system including the number of protein beads, the number of polarizable water molecules, and temperature details are also listed in Table 3.7. Adequate number of ions are added to neutralize the system. We have used reduced temperature, $T^* = K_b T/\epsilon$ as our temperature scale, where ϵ is the highest interaction strength (5.6 kJ/mol) in our CG potential. Energy minimization is first conducted with steepest descent. An NPT equilibration run at T^*=0.52 for 50 ps is followed at 1 bar and with time step 0.001 ps. The canonical production run is then performed with time step 0.01 ps for the simulation time listed in Table 3.7 at different reduced temperatures. Since the box size is constant, the solvent density remains the same across various temperatures. Electrostatic cutoff is 1.6 nm and particle-mesh Ewald (PME) method is applied for long range electrostatic interactions [217]. Nos$\tilde{e}$-Hoover thermostat is used to maintain the system at the desired temperature [218].

Table 3.7: Simulation setup for each protein. ^{a} Sequence: KLVFFAE.

	Trp-cage	Trpzip4	Villin	WW-Domain	β-α-β	Aβ16-22
PDB code	1L2Y	1LE3	1YRF	1E0L	2KI0	-^{a}
Number of proteins	1	1	1	1	1	8
Number of water	1957	1610	26391	30136	24998	2584
Simulation time per temperature (ns)	200	200	600	100	200	300
Temperature range performed	0.52-1.04	0.52-1.04	0.52-1.04	0.52,0.82	0.52,0.82	0.52
Total number of temperature simulated	15	15	19	2	2	1

3.3.3.2 GpA Model

Previous research has emphasized on the importance of Cα-H..O interactions in driving helix dimerization in membranous environments. This allows the polar amino-acids such as glycines to exist on dimerization surfaces. To capture the Cα-H..O interactions in our low-resolution CG simulations, we have added an *ad-hoc* interaction between the main BB of Glycines with Isoleucines and Valines at the dimerization interface (Figure 3.13). The interaction is added as an inverted Gaussian function, with median at 0.3 nm, and an amplitude and standard deviation of 15 kJ/mol and 0.2 nm respectively. The net well depth ($\sim$ 14.5 kJ/mol at 0.49 nm) of this Cα-

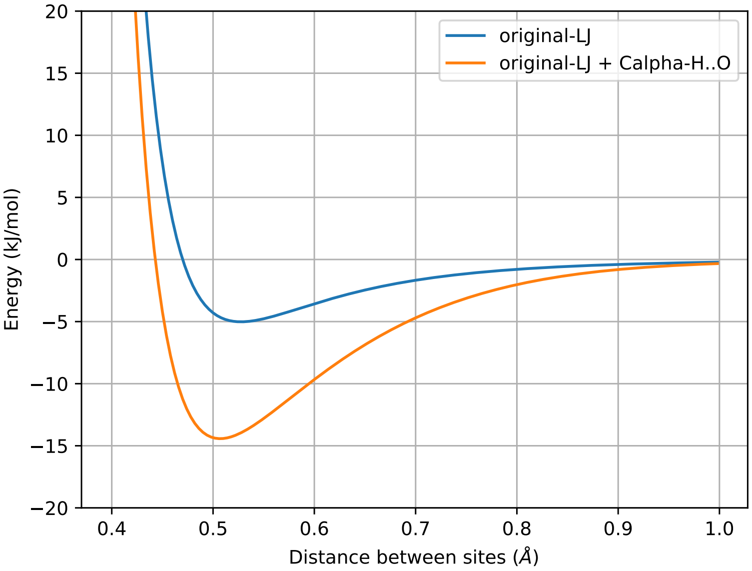


Figure 3.13: Interaction between Gly backbone of one monomer with Val/Leu backbone of the other monomer before (blue) and after (orange) addition of CαH..O interaction.

H..O interaction is close to the estimate from previous reported *ab* initio quantum calculations ($\sim$ 12.5 kJ/mol). Moreover, the Cα-H..O interaction we implemented in the model is an extremely short-ranged interaction, and the potential slowly approaches the initial LJ potential within 1 nm.

3.3.3.3 Dodecylphosphocholine (DPC) Model

Since, *ProMPT* is created with most of the interaction parameters along the MARTINI-defined interaction levels, it allows using the MARTINI forcefield for environmental entities, without any transferability issues. As such, the detergent molecules are created with the MARTINI forcefield. The mapping scheme for the DPC micelle is shown in Figure 3.14.

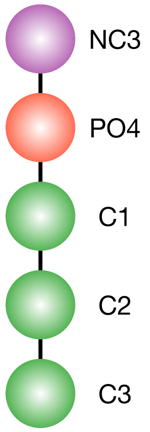


Figure 3.14: CG geometry of the DPC micelle. Apolar beads are in green, negatively charged bead is in red, and positively charged bead is in purple. The bead names are shown on the right.

3.3.3.4 Setup for GpA

A GpA WT monomer with a sequence indicated in Figure 3.15 is constructed as a random coil with VMD. The resulting atomistic topology is then converted to a CG representation using an in-house code adopting the *ProMPT* model. The two resulting CG GpA monomers are then placed in parallel in a 8 nm wide cubic box with a distance of 5.66 nm between the two monomers. The box is then solvated with 80 DPC detergents and 3000 CG water molecules. For mutants T87L and G79L (sequences shown in Figure 3.15), the same work flow is followed with different starting mutated random coil structures. These CG structures are first energy minimized with steepest descent algorithm. An NPT equilibration run at 350K for 5 ns is followed at 1 bar and with a time step of 0.001 ps. The compressibility is 3.50×10^5 bar^{-1}. The production run in an NPT ensemble is then performed with a time step of 0.01 ps for the simulation time of 500 ns at 350 K. The remaining set-up is the same as that for mini-proteins. Six replicas are run for each WT/mutant.

75 80 85 90 95
Wild type: ITLIIFGVMAGVIGTILLISYGI
T87L: ITLIIFGVMAGVIGLILLISYGI
G79L: ITLIIFLVMAGVIGTILLISYGI

Figure 3.15: The amino acid sequence for WT, T87L, and G79L. The mutation point is colored in red and the GxxxG motif is underlined. The residue index is also marked on the top.

3.3.4 Comparison with replica exchange molecular dynamics simulations with solute tempering (REST)

The replica exchange with solute tempering (REST) [219] simulations are performed for Trp-cage and Trpzip4 in water and the Amber99sb [220, 220] atomistic force field is used. For Trp-cage, an unfolded coil is solvated with 2678 water molecules and neutralized with one chloride ion. The system then goes through energy minimization, 200 ps of NVT equilibration and 200 ps of NPT equilibration. After equilibration, the REST protocol is used for the production runs where 10 replicas are simulated spanning a temperature range from 290 K to 540 K. Each replica is run for 700 ns and the resulting exchange rates are between 22% to 36%. For Trpzip4, an unfolded coil is solvated with 3570 water molecules and neutralized with 3 sodium ions. The system undergoes the same equilibration process and the same REST protocol as those for Trp-cage. For Trpzip4, a longer simulation time is needed for convergence, each replica is run for 1400 ns and the resulting exchange rates are between 24% to 32%.

3.3.5 Analysis

We use potential of mean force (PMF) plots of Trp-cage, Trpzip4, and villin to validate the ability of our CG force field to capture a protein's conformational landscape. Backbone native contact and root-mean-squared-deviation (RMSD) of peptide fragments are used as reaction coordinates for PMF calculations. Backbone native contact measures the fraction of backbone contacts with a cutoff of 7 Å that are replicated in our CG trajectories that are native to the PDB-converted-CG structure. RMSD BB is a measure of relative deviation of our CG backbone from the BB in PDB-converted-CG structure. For all proteins, the first 80 ns of trajectory-data is removed and the equilibrated data are collected at all temperatures simulated and re-weighted through Multistate Bennett Acceptance Ratio (MBAR) method [221]. Here, we present the final PMF at $T^*=0.52$. For villin, we use a cluster analysis tool from GROMACS with a RMSD cutoff of 0.3 nm to generate representative conformation corresponding to each basin in the PMF. For the aggregation simulations, two peptides are considered part of β-sheet if more than three dummy-dummy interaction pairs are formed. A fraction of one means that all the peptides (8 of them) are forming β-sheet.

PMF plots are again used to capture the conformational landscape for GpA WT and mutants. For each replica, only the trajectory after at least one intermonomer primary CG site (any CG interaction site that is not a dummy) contacts with a cutoff of 6 Å is used for analysis. These trajectories are then aggregated to construct the PMF plots. The average helical content and the number of BB contacts are used as the reaction coordinates for the PMF calculations. Helical content is defined as the fraction of intramonomer backbone contacts with a cutoff of 6.5 Å, that are native to the PDB-converted-CG structure (PDB code: 1AFO.pdb). For calculating the number of backbone contacts, we also used a cutoff of 6.5 Å. A trajectory from the most populated basin of the PMF plots for each WT and mutant is extracted and used for all the following analyses. The criteria to determine the most populated basin is 0.95 to 1 (average helical content) and 9 to 18 (number of BB contacts), 0.95 to 1 (average helical content) and 14 to 22 (number of BB contacts), and 0.85 to 0.92 (average helical content) and 6 to 12 (number of BB contacts) for WT, T87L, and G79L, respectively. For the crossing angle computation, only the BB is considered. The crossing angle is determined by two vectors, where each vector only considers the position of the tail and the head of the monomer (see Figure 3.16). The contact map is constructed by counting the contacts between intermonomer BB with a cutoff of 6.5 Å and dividing by the total number of frames used for analysis. As for the average number of DPC contact, a contact is counted when the desired bead (Choline: NC3 and acyl-tails: C3) is within 6 Å of any heavy atom from the GpA peptides and then divided by the total number of frames used for analysis. Only the BB is considered when calculating the dihedral angle for G79L. All the insert GpA images are rendered from VMD.

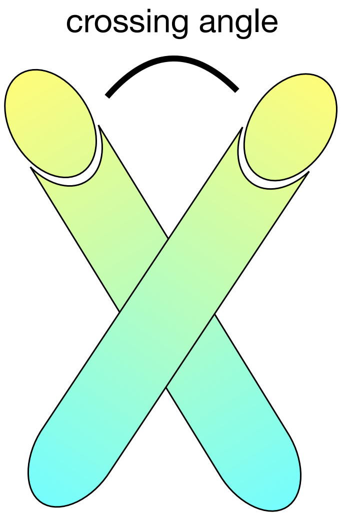


Figure 3.16: An illustration for the crossing angle between two GpA monomers.

3.4 Results and discussion

3.4.1 Mini-proteins

Mini proteins Trp-cage and Trpzip4 are selected to be our first test proteins because they are already well-studied both computationally and experimentally [38, 222, 223, 224, 225]. Trp-cage has an α-helix starting from the N-terminus and a 3-10 helix at the middle part, and an unstructured C-terminal tail, which collapse together forming a hydrophobic core (Figure 3.17 (c)). Trpzip4 has a β-hairpin structure with four Trp residues packing together. Figure 3.17 (a) shows the PMF for Trp-cage using native contact and RMSD helix BB as two reaction coordinates. A two-state folding/unfolding path can be observed from the PMF, which agrees with experiments and REMD simulation results [222, 226]. The folded basin at around 0.1 nm RMSD helix BB indicates a perfect α-helix. The distribution of native contact is relatively broader, from 0.6 to 0.7. The unfolded basin has a a native contact of 0.4 to 0.5 and RMSD helix BB from 0.2 to 0.3 nm, indicating some partial helicity, in agreement with previous simulation reports [227]. The PDB structure and a representative folded structure with our CG model for Trp-cage are shown in Figure 3.17 (c). Both α-helix and 3-10 helix can be well reproduced at the correct positions. The hydrophobic core formed by Tyr3, Trp6, and three proline residues at the C-terminus is also well captured as seen in the experimental structure (Figure 3.18). Our estimation of ΔG for protein folding is 3.2 kJ/mol, close to the experimental estimate of 2.6 kJ/mol [228]. We also perform a set of replica exchange with solute tempering (REST) simulations to estimate the folding-unfolding free energy difference for Trp-cage, from the atomistic simulations, the ΔG is estimated to be around 2.81 kJ/mol at 290 K (Figure 3.19). The time series of the two reaction coordinates we use for PMF at T*=0.82 are shown in Figure 3.20 and 3.21. At this higher temperature (T*=0.82) our CG model could sample frequent transitions between the folded states and the unfolded states. We use the simulations of Trp-cage to evaluate the computational efficiency improved by using our CG model compared to atomistic simulations using an enhanced sampling method. In order to get a converged PMF, we use different protocols for atomistic (REST) and CG (Replicas at different temperatures). Therefore the computational resources needed will also depend on the choices of difference parameters such as the number of replicas and the range of temperatures. Here, we present a simplified comparison of the computational resources used to get the estimate of folding-unfolding free energy. Our CG model improves the computational efficiency by at least ten fold. This has not taken into account the extra simulations needed to fine tune the parameters for REST simulation protocol and the atomistic force field, that will lead to even higher computational cost in practice.

Figure 3.17 (b) shows the PMF for Trpzip4 with native contact and RMSD BB as the reaction coordinates. The PMF of Trpzip4 also shows a two-state folding/unfolding path, which agrees with experiments and atomistic simulations. The folded basin for Trpzip4 is located at 0.35 nm RMSD BB with a native contact of around 0.8. The unfolded basin has a broad RMSD BB distribution from 0.5 to 0.7 nm with a native contact of around 0.4. Figure 3.17 (d) shows both the PDB structure (left) and a representative folded state for Trpzip4 with our model (right). The β-sheet content can be well captured. Four Trp residues are interacting with each other and forming a hydrophobic core as observed in the experimental structure (Figure 3.22). Here, our CG model estimated a ΔG of about 7.3 kJ/mol for protein folding, compared to 12.3 kJ/mol reported in experiments [229]. From the REST simulations of Trpzip4, a ΔG of 4.53 kJ/mol is estimated (Figure 3.23), which is lower than both the experimental data and the estimation

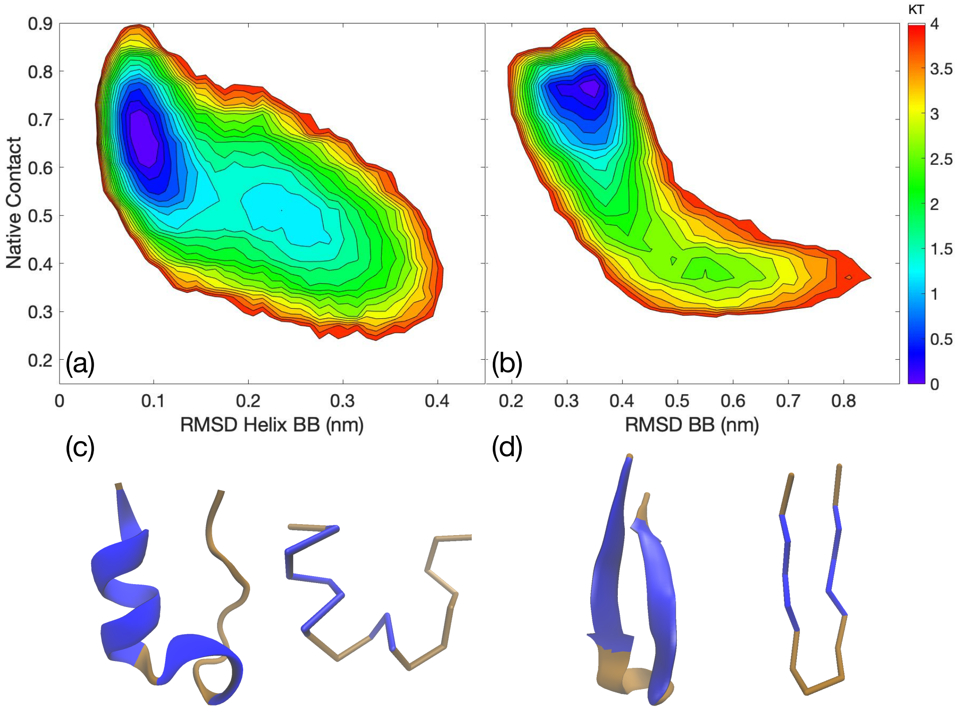


Figure 3.17: PMF for Trp-cage (a) with RMSD helix BB and native contact as the reaction coordinates. PMF for Trpzip4 (b) with RMSD BB and native contact as the reaction coordinates. Both PDB structure (left) and the representative structure (right) from our model for Trp-cage (c) and Trpzip4 (d) are shown. Blue indicates the specific secondary structure each protein exhibits.

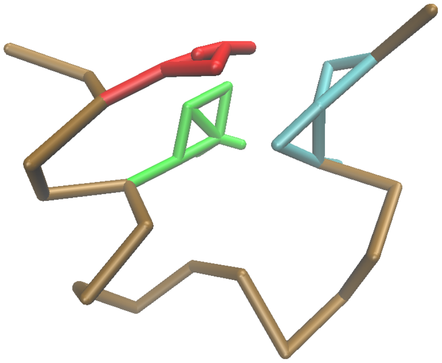


Figure 3.18: The folded structure of Trp-cage showing the packing of Tyr (red), Trp (green), and Pro (cyan) residues.

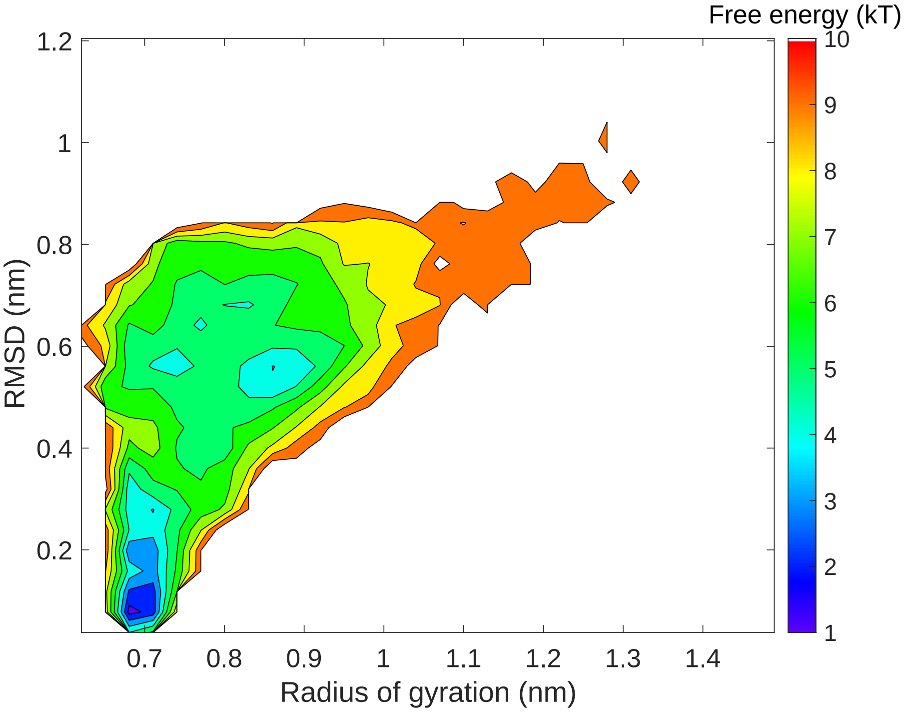


Figure 3.19: Free energy plot for Trp-cage at 290 K from the REST simulations. The folded basin has a free energy of 2.82 kT and the unfolded basin has a free energy of 3.98 kT. The ΔG is estimated to be 1.17 kT (2.81 kJ/mol).

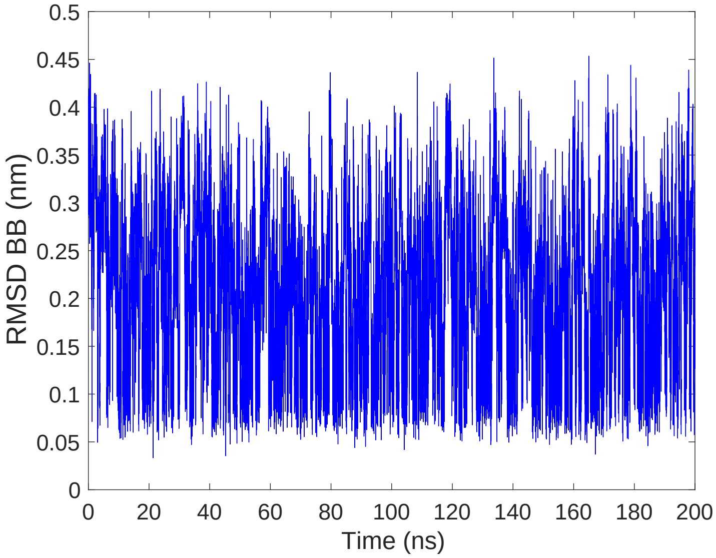


Figure 3.20: RMSD helix BB time series for Trp-cage at T^{*}=0.82.

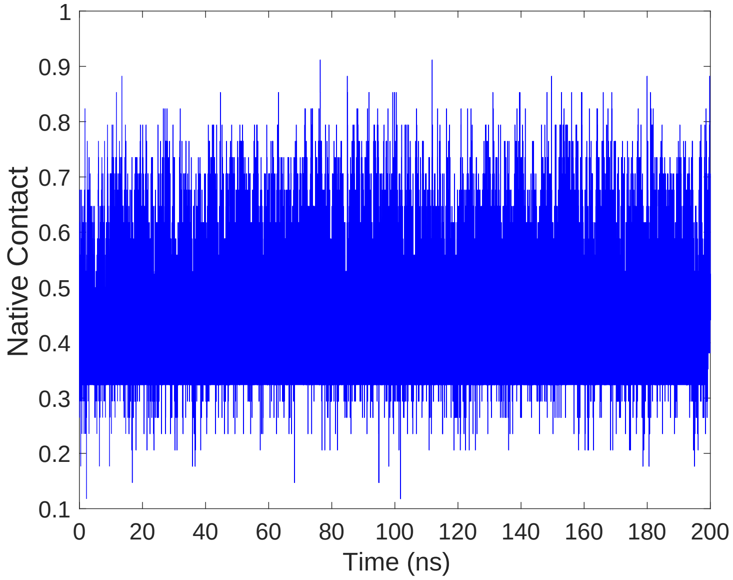


Figure 3.21: Native contact time series for Trp-cage at T*=0.82.

from our CG model. Again, the multiple folding-unfolding transitions at a higher temperature (T*=0.82) for the two reaction coordinates, as time series are shown in Figure 3.24 and 3.25.

Our next target protein - villin has three helices with a hydrophobic core. Successful folding of villin validates our CG protein model for general applications to proteins with tertiary packing of homogeneous secondary structures. Figure 3.26 shows the PMF for villin with RMSD strand 1 (S1) and RMSD strand 2 (S2) as two reaction coordinates, where the locations of the two strands are highlighted in red. Strand 1 contains two helices from the N-terminus (residue 2-18), while strand 2 only contains one long helix (residue 21-32) from the C-terminus. There are one native state and two intermediate states populated at T*=0.52. At higher temperatures, the location of the unfolded basin shifts to more extended conformations (Figure 3.27). All of the states populated at T*=0.52 have similar RMSD S1 values, but different RMSD S2 values, which correspond to different tertiary structures. Indeed, it has been reported in experiments that although there are fluctuations in the N-terminal part of villin, large-scale unfolding will

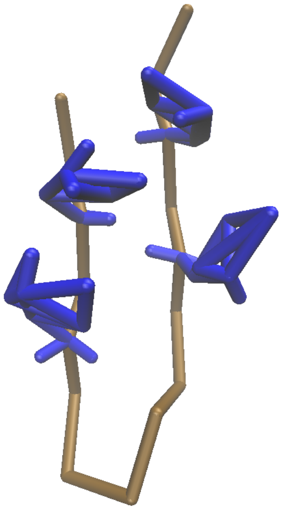


Figure 3.22: The folded structure of Trpzip4 with the packing of Trp residues shown.

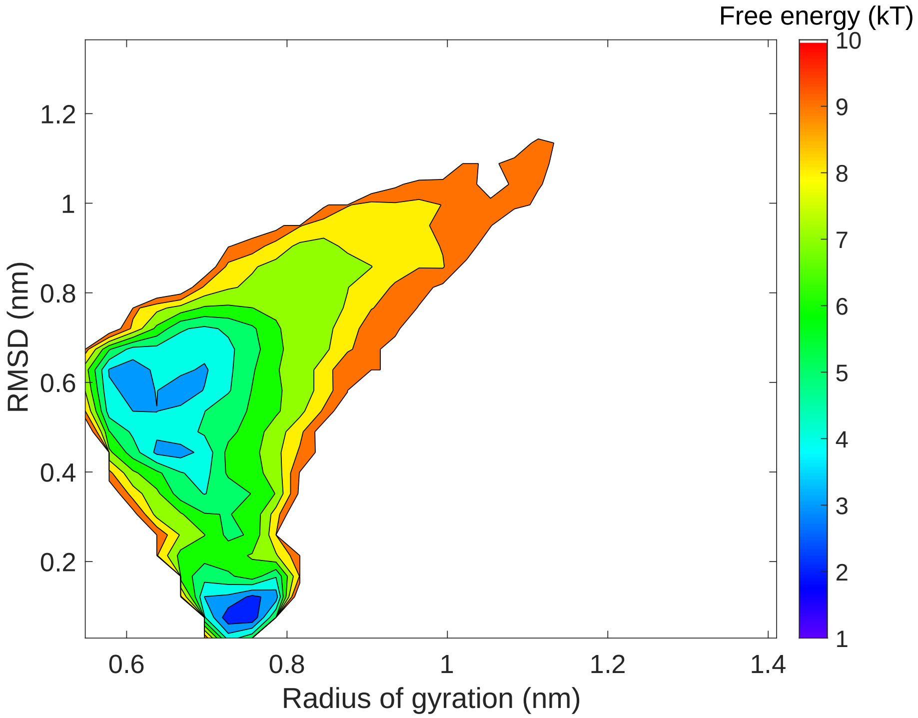


Figure 3.23: Free energy plot for Trpzip4 at 290 K from the REST simulations. The folded basin has a free energy of 2.12 kT and the unfolded basin has a free energy of 4.00 kT. The ΔG is estimated to be 1.88 kT (4.53 kJ/mol).

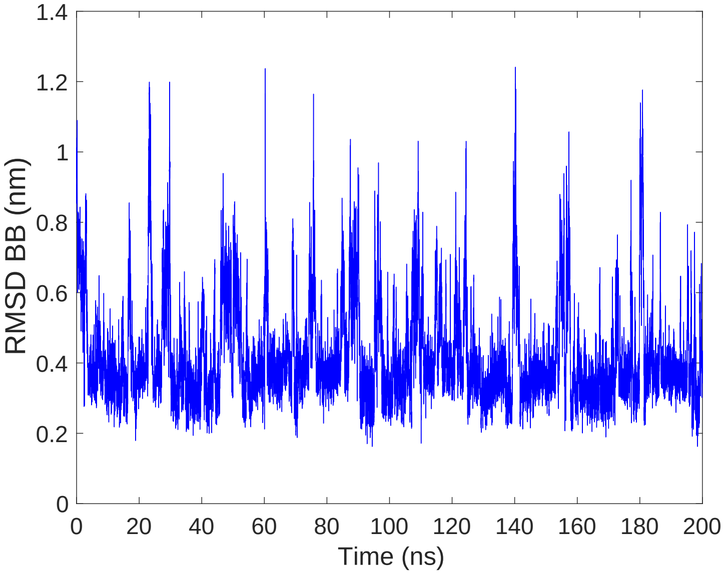


Figure 3.24: RMSD BB time series for Trpzip4 at T*=0.82.

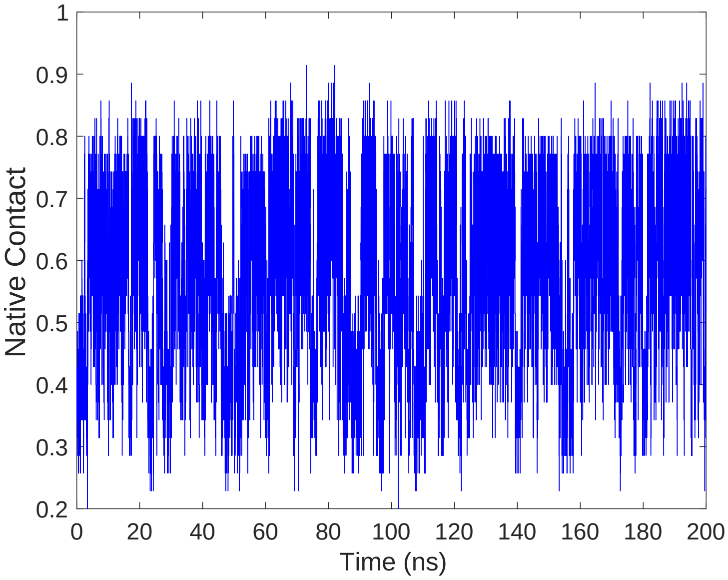


Figure 3.25: Native contact time series for Trpzip4 at T*=0.82.

not occur unless undergoing global unfolding [230]. In our model, global unfolding happens at higher temperatures, where disruption of strand 1 can be found as seen in Figure 3.27. The native state is centered at RMSD S1 0.35 nm and RMSD S2 0.1 nm. The representative structure of the folded basin with our model is very similar to the experimental structure from PDB. For the intermediate states, one is centered at RMSD S1 0.35 nm and RMSD S2 0.2 nm while the other one is centered at RMSD S1 0.35 nm and RMSD S2 0.35 nm. Both of the intermediate basins show a decrease in helicity for strand 2 and losing of hydrophobic core, while the helicity for the basin with larger RMSD S2 is reduced further more. The intermediate states are similar to the states described in the experimental work from Reiner and co-workers, where the intermediate states have an unfolded strand 2 that is undocked from the hydrophobic core [230]. Both the native state and intermediate states are on the folded side of the folding-unfolding landscape. The frequent transitions between folded states and unfold states for RMSD S1 and RMSD S2 are shown at T^{*}=0.82 in Figure 3.28 and 3.29.

For the proteins having a tertiary structure with homogeneous β-sheets, we use WW-domain as our benchmark protein. The PDB structure is shown in Figure 3.30 (c) with three antiparallel β-strands and two long tails at both ends. A representative folded structure with our model is shown in Figure 3.30 (c). In Figure 3.30 (a), the RMSD BB time series without considering the loop regions are shown at two temperatures. From the time series at T^{*}=0.52 we observe that the simulation converges at a value of RMSD BB around 0.3 nm. The simulation can reach the folded state in about 20 ns. We could note frequent transitions between folded and unfolded states at T^{*}=0.82. The last protein we focus on β-α-β (PDB code: 2KI0), which is a designed protein that has a mixture of secondary structures (partially α-helix and β-sheet) with a defined tertiary structure. Using DSSP [231] we assign the secondary structures using

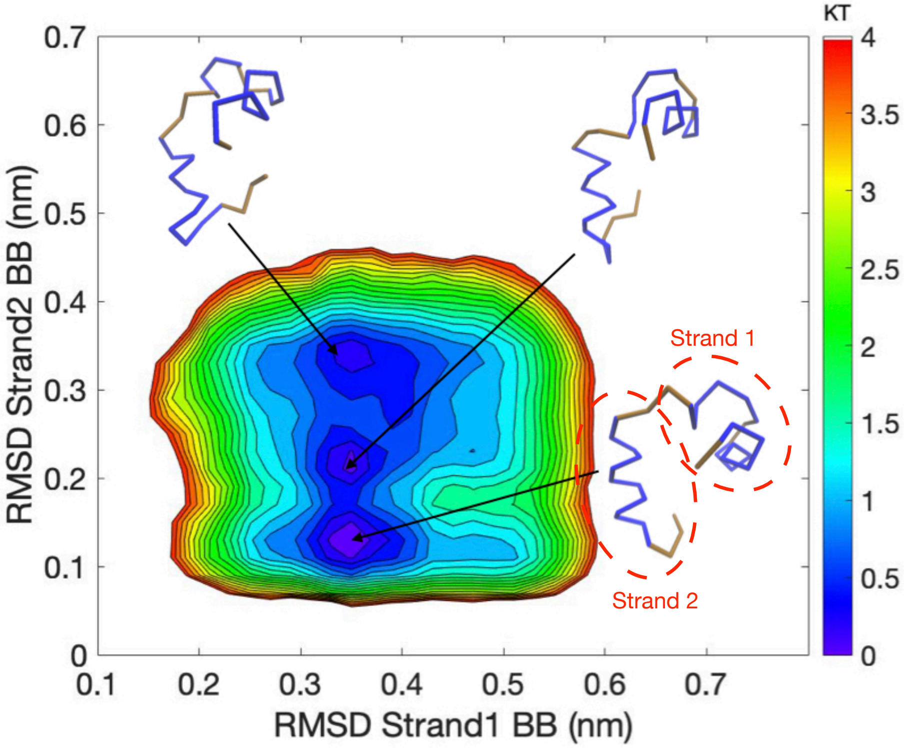


Figure 3.26: PMF for villin with RMSD S1 and RMSD S2 as the reaction coordinates at T*=0.52. The representative structure for each basin is shown as insets.

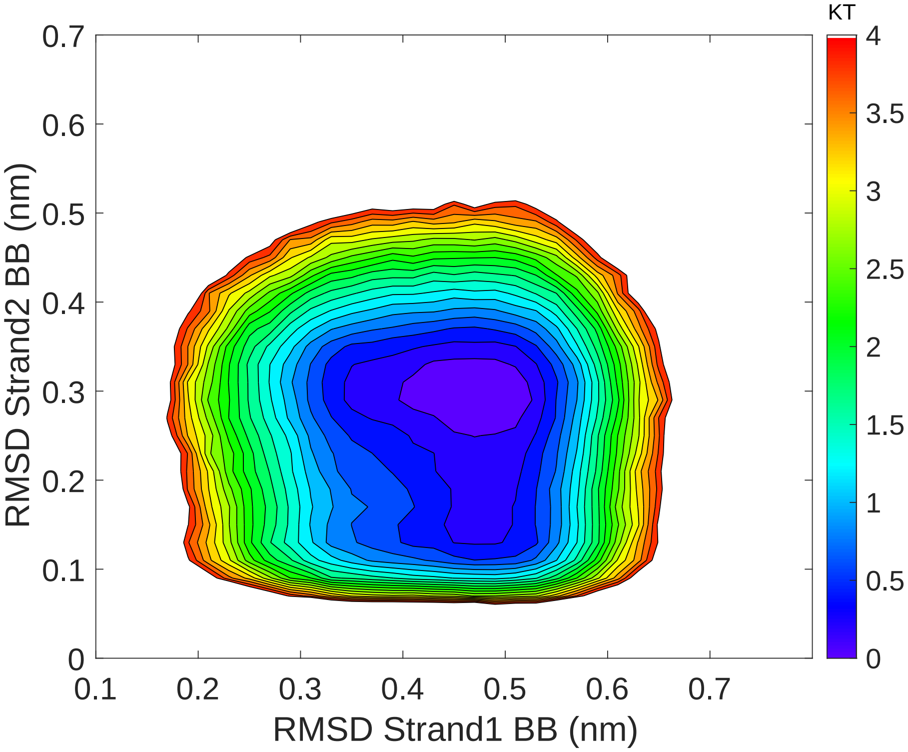


Figure 3.27: PMF for villin with RMSD S1 and RMSD S2 as the reaction coordinates at T*=0.82.

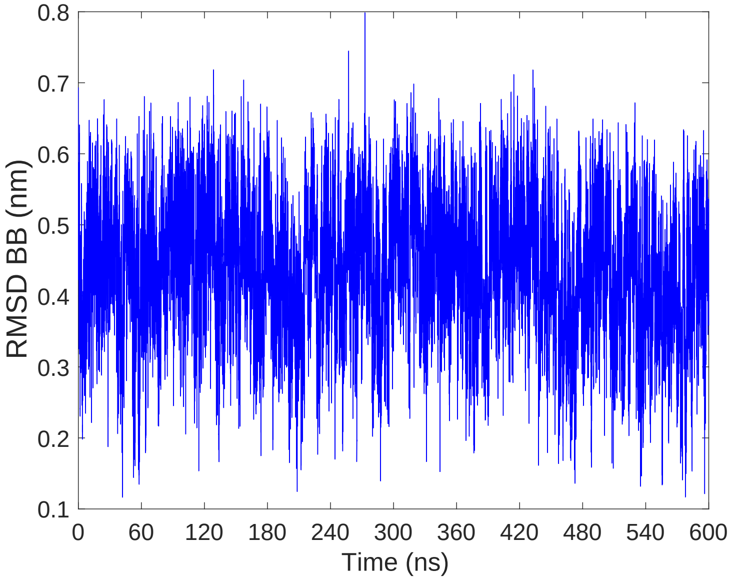


Figure 3.28: RMSD S1 time series for villin at T*=0.82.

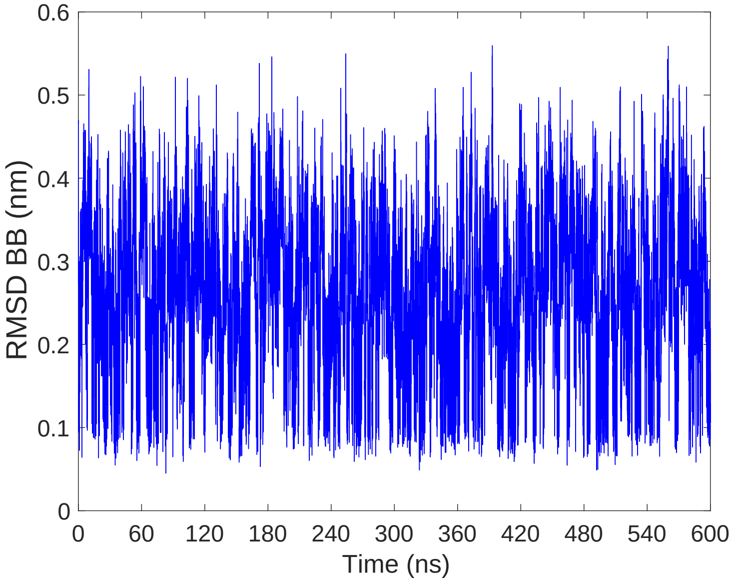


Figure 3.29: RMSD S2 time series for villin at T*=0.82.

one of the deposited NMR structures in the PDB database. The antiparallel β-strands only have four residues on each strand. Figure 3.30 (d) shows a representative folded structure with our model, the α-helix can be well formed, while the β-sheets are formed but shifted. Figure 3.30 (b) shows the RMSD BB time series for β-α-β at two temperatures, without considering the tails. The RMSD BB converges at around 0.5 nm at T*=0.52. The time series at a higher temperature (T*=0.82) is shown in red, which confirms the ability of *ProMPT* to capture frequent transitions between different states.

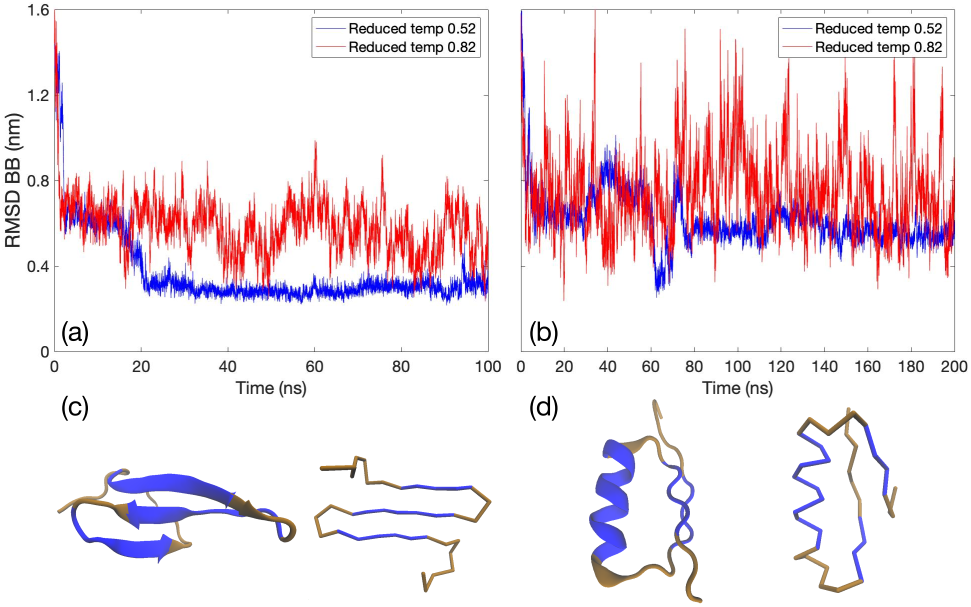


Figure 3.30: RMSD BB time series for (a) WW-domain and (b) β-α-β at T*=0.52 (blue) and T*=0.82 (red). The PDB structure (left) and the representative structure (right) are shown in (c) and (d) for WW-domain and β-α-β, respectively. Blue indicates the specific secondary structure each protein exhibits.

We have demonstrated that with our model we can fold a variety of proteins, which include pure β-sheets, pure helices, and proteins that have a defined tertiary structure composed with homogeneous secondary structures. For proteins that have mixed secondary structures such as β-α-β, the packing could not be perfectly reproduced. We also attempt simulating the folding of another well-studied protein with heterogeneous secondary structures - GB1 (PDB code: 2J52.pdb). GB1 is a slightly larger protein with four β-strands and one α-helix. Although our CG model can generate accurate secondary structure at appropriate regions, the overall tertiary packing is not reproduced (results not shown). We compare the molecular packing of amino-acid side-chains in *ProMPT* for GB1 to GB1 PDB native structure mapped into our CG description. *ProMPT* simulations, with backbone restrained to the native state, resulted in a protein core with packing defects and voids. On the other hand, CG structure mapped directly from PDB native state showed no such anomalies. Therefore, we postulate that appropriate modifications to side-chain size, can achieve adequate protein-core packing required for correct tertiary structures. Proteins with mixed folds will be our next target proteins and could potentially be solved by adjusting the size and shape of each amino acid in the CG model to better mimic the side-chain geometry. Recent parametrization of MARTINI 3 force field has adopted this strategy and included several smaller bead-types to capture the accurate geometrical shapes of biomolecules [176]. This resulted in significant improvement in molecular packing, and could capture correct open and close conformations in transmembrane proteins.

Beyond studying protein folding, *ProMPT* can also be applied to study peptide aggregation. Here we demonstrate the aggregation of Aβ 16-22 peptides in water at T^*=0.52. The protein concentration in the system is 0.0426 M. Figure 3.31 shows the time series for β-sheet fraction. The β-sheet fraction fluctuates between 60% to 80%, indicating significant fibrillation. Occasionally the β-sheet fraction can reach 100%, where all the Aβ peptides are involved in the β-sheet formation. Similar to previous experimental and simulation studies, peptides organize into stable, layered β-sheet rich structures [232, 233, 234, 235, 236]. Phe residues are shown in yellow and a hydrophobic core formed by Phe packing can be observed from the aggregate snapshot (inset of Figure 3.31). We observe anti-parallel β-sheets in our simulations, similar to the structures reported by experiments [237]. Nguyen *et al.* applied replica exchange MD on Aβ 16-22 monomer, dimer, and trimer to study the effects of different atomistic force fields on the formation of β-sheets [238]. While AMBER99 and OPLS generated a diversity of conformations, only GROMOS96 could successfully generate antiparallel β-sheets. With *ProMPT*, we observe aggregation into β-sheet rich, fibril-like structures by 30 ns and can also qualitatively study the kinetics of this aggregation process.

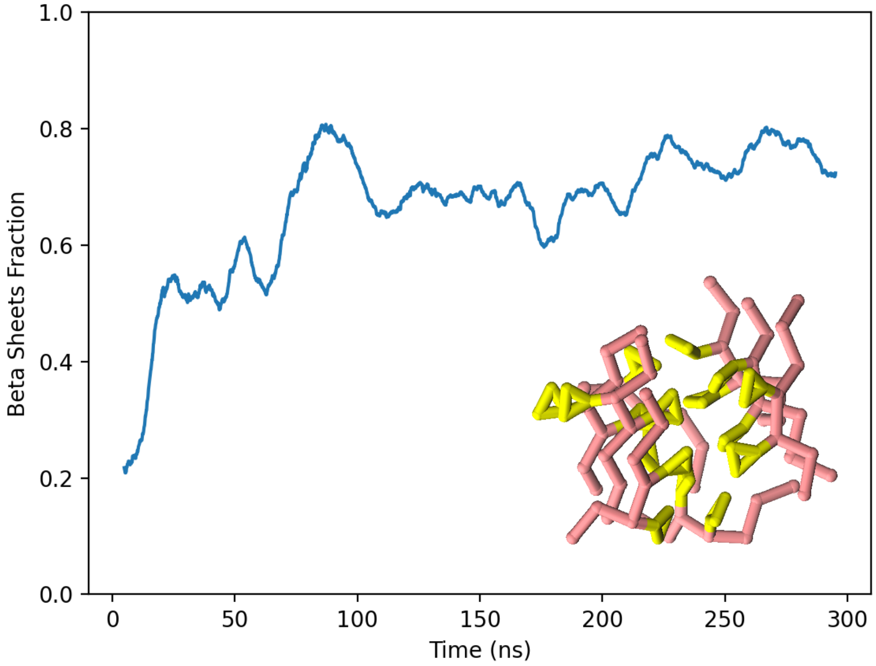


Figure 3.31: Time series for the fraction of Aβ 16-22 peptide forming β-sheets. An illustration of β-sheets aggregation is shown in the inset with yellow representing Phe residues.

Here we must note that although similar energy scales are used, comparisons should not be drawn with the MARTINI family of forcefields as with MARTINI external restraints are instituted to prevent structural transitions and spontaneous unfolding.

The applications and validations of our CG model in this work focus on capturing accurate protein conformational landscape and structural diversity in *aqueous* environment. This explicit treatment of environment can capture local electrostatic changes such as dielectric fluctuations more effectively, which is necessary for several biomolecular processes [208]. Moreover, the importance of these local electrostatic changes was noted in protein folding by previous reports [239]. Also, the use of explicit environment allows for easy transferability of interactions while simulating proteins in different biomolecular environments. We do not have to re-parametrize the protein forcefields as the cross interactions between the protein and the environment would automatically change with changing environments. Therefore, *ProMPT* can have special applications in studies involving chemically heterogeneous biomolecular systems due to its *explicit* description of environment (and solvent). Because of similar parametrization of non-bonded interactions, *ProMPT* can be patched with our *in-house* CG membrane model — Water Explicit Polarizable MEMbrane model (WEPMEM) to study membrane and detergent assisted conformational transitions [188, 189, 190, 240]. Moreover, *ProMPT* can also be patched with our CG model for divalent ions to study metal-assisted protein folding [241]. Finally, the modular architecture of the force field allows for easy parametrization of other biomolecules.

3.4.2 Transmembrane protein Glycophorin A (GpA)

3.4.2.1 Conformational landscape of GpA and mutants

As validation of our simulations and establish comparisons with previous experimental and simulation results, we compute the potential of mean force (PMF). Figure 3.32 (a) shows the PMF of WT GpA, where the most populated basin is observed with a high average helical content and a broad range of number of BB contacts. A representative conformation is shown as the insert figure. In the representative conformation, the WT GpA dimerizes as a parallel dimer with T87-T87 interaction. This T87-T87 interaction can also be seen in the contact map between the two monomers (Figure 3.33). In addition, the $G_{79}\mathrm{xxx}G_{83}$ motif of the two monomers also have contacts with each other (see also Figure 3.33 (a)). These properties agree with the solution NMR structures solved in the DPC micelles where extensive backbone-backbone contacts are found within the G79/G83 region [194]. It has been reported from mutational analysis that there is a high specificity of sequence dependence on GpA dimerization [198]. We first examine the mutant T87L where dimer disruptions have been reported for any non-polar amino acid as a replacement of T87 [198]. The most populated basin in the PMF of mutant T87L (Figure 3.32 (b)) spans a range of average helical content that is similar to that of WT, but with a higher number of BB contacts. A second populated basin is observed with the similar high helical content but with a lower number of BB contacts. The conformations in this second basin can dimerize in both parallel and anti-parallel fashion (Figure 3.33 (b)). The insert of Figure 3.32 (b) shows a representative conformation for a parallel T87L dimer. In this parallel T87L dimer, the critical T87-T87 interaction existing in the WT dimer is disrupted by Leu residues. These Leu residues interact with the surfactants and not between themselves, which is in contrast to the interaction between Thr residues at this position for the WT. The GxxxG motif is observed to have contacts with each other (Figure 3.33 (b)). Last we examine the mutant G79L, which has also been reported from mutational experiments to form monomers [198]. From Figure 3.32 (c) we observe the most populated basin in the PMF to have a lower range of average helicity content compared to that of the WT and T87L. The number of BB contacts also decreases. We note a bent-helix for one of the monomers (insert Figure 3.32 (c)), and the dimer does not have T87-T87 interactions as observed in the contact map of G79L (Figure 3.33 (c)). Moreover, since G79 is mutated to a Leu residue, the original GxxxG motif does not exist and there are no intermonomer contacts between the same region of the monomers.

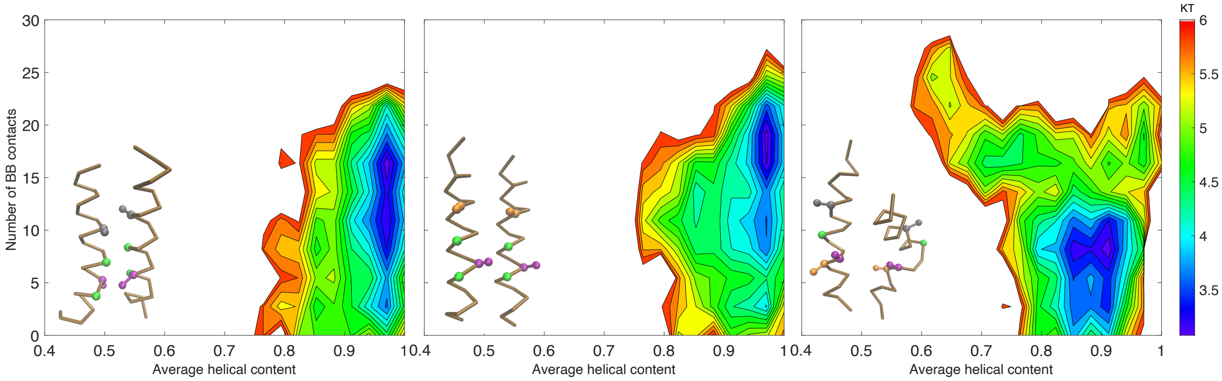


Figure 3.32: PMF for (a) WT, (b) T87L , and (c) G79L GpA with the average helical content and the number of BB contacts as the reaction coordinates. The criteria to determine the most populated basin is 0.95 to 1 (average helical content) and 9 to 18 (number of BB contacts), 0.95 to 1 (average helical content) and 14 to 22 (number of BB contacts), and 0.85 to 0.92 (average helical content) and 6 to 12 (number of BB contacts) for WT, T87L, and G79L, respectively. The representative conformation for each WT/mutant is shown in the insert figures. Color code: Thr (grey), Gly (green), Val (purple), Leu (orange).

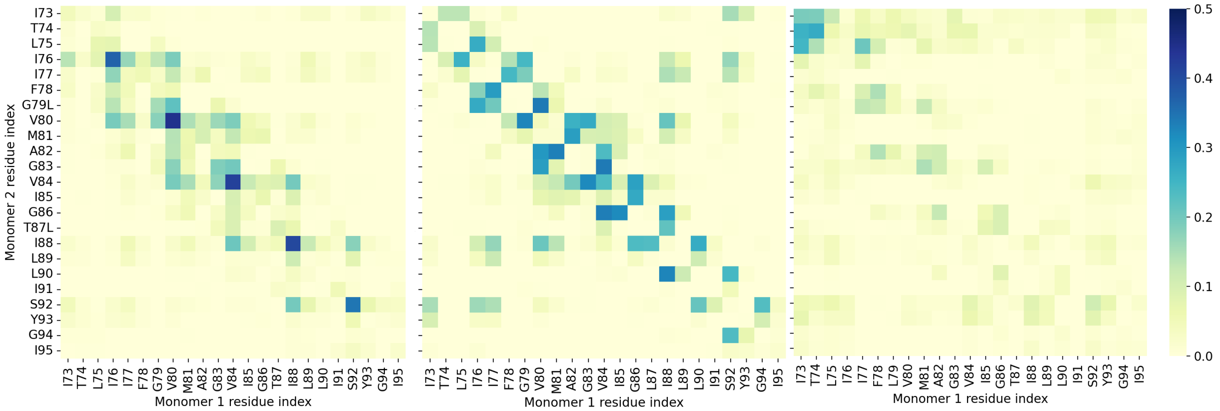


Figure 3.33: The residue-residue contact maps for (a) WT, (b) T87L, and (c) G79L.

3.4.2.2 Changes in peptide-surfactant interactions tune the emergence of helix bend formation

We further inspect the reasons for the monomer-bending in G79L by examining the contacts between G79L and DPC detergent micelles. Figure 3.34 (a) shows the average number of contacts betwteen each WT/G79L residue and DPC headgroup. The major difference between the WT and G79L occurs at the middle part of the monomer, where there are more residue-NC3 contacts for G79L, especially for residue V80 to V84, which belong to the GxxxG motif. Figure 3.34 (b) shows the similar analysis but with the hydrophobic bead (C3) of DPC. In this case, the residues at the central region of the G79L mutant are less exposed to C3, especially for M81 and I85. Moreover, we note that the mutated L79 has more contacts with C3, driven by the hydrophobic effect, and this alters the following six residues (V80 to I85) to be more exposed to the DPC headgroup which is closer to the micelle/water surface. Subtle changes in the hydrophobic environment caused by a point mutation were also observed from previous literature, where a mutation point at the GxxxG motif of Amyloid Precursor Protein (APP) results in different hydration level of the transmembrane protein compared to WT [203]. The change in the hydrophobic environment near the center of the G79L mutant potentially destroys the hydrophobicity balance in the original helix and results in a helical bend.

3.4.2.3 Increased helical flexibility near the GxxxG region

The dihedral angle analysis for G79L (Figure 3.35) provides a quantitative analysis for the bending region of the G79L mutant. From residues L79 to G86, we observe large deviations of the dihedral angle, which marks the hinge of the alpha helix. This result agrees with the previous residue-DPC contact analyses where the same region experiences the most significant environmental variation resulting in a bent conformation. The position of the hinge at the GxxxG region agrees with previous findings for mutated APP where bending and modulation of helicity is observed at the GxxxG region [203, 204]. Lemmin *et al.* noted a loss in helicity and helical bending

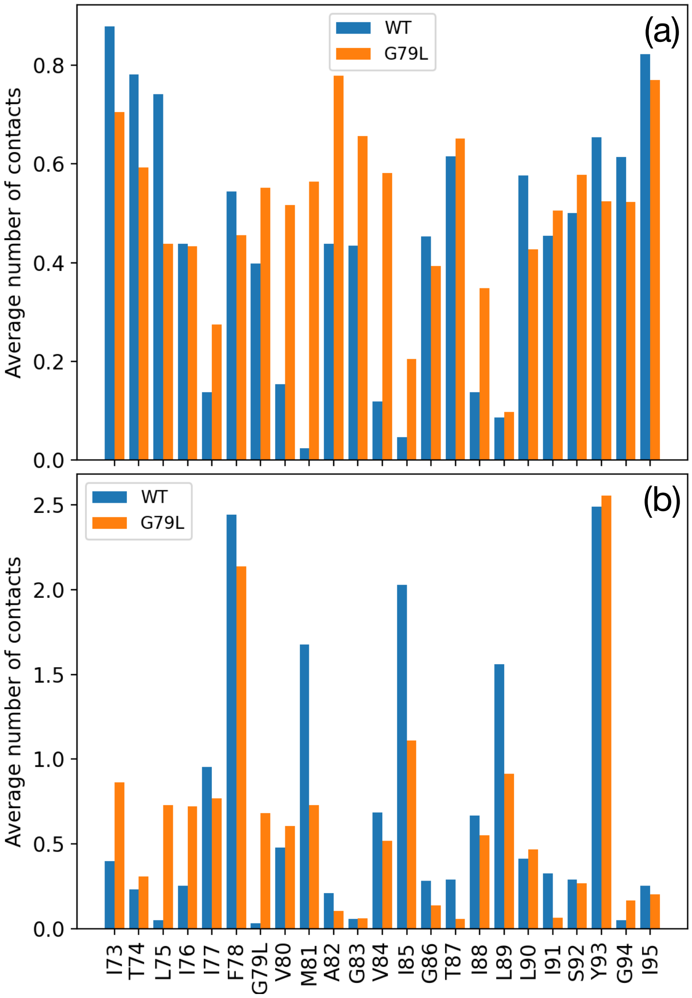


Figure 3.34: The average number of contacts between WT (blue)/ G79L (orange) and the (a) Choline bead (NC3); (b) hydrophobic bead (C3) of the DPC detergent.

due to a mutation of Gly to Leu at the N-terminal side of GxxxGxxxG domain, similar to our observation for GpA. On the other hand, Gotz *et al.* found an increase in helicity if the Gly to Leu mutation is at the C-terminal side, which shows that the change of transmembrane helix helicity is also a function of where the mutation is performed in the GxxxG motif. The characterization of transmembrane protein helix bending is important because the structural and dynamical variance caused by point mutations could result in different enzymatic activity and dimer stability. Since the alpha helix is bent but not destroyed, high helicity can still be expected. A similar, but more disruptive GpA double mutant G79LG83L has been reported to still exhibit alpha helical secondary structure from circular dichroism (CD) data [200]. Therefore, we hypothesize that CD cannot precisely distinguish this slight loss in helicity originating from monomer bending, since the average helicity from simulations is still high (around 90%) (Figure 3.32 (c)) for G79L. In the case of GpA, unlike soluble coils where the formation of helix is coupled to dimerization, the formation of alpha helices and dimerization are independent events [193]. For WT the helices are stable due to the strong hydrogen bonding in a low dielectric environment and the formation of the dimer can be stabilized by vdW interactions between the complementary surfaces from the two helices [200]. However, for G79L, the mutation of Gly to Leu results in a bent alpha helix which does not have a smooth surface at the GxxxG region to aid close backbone interactions. The novelty of our model compared to the popular MARTINI CG model is that there are no specific secondary structure restraints imposed, and therefore folding transitions can be observed. This feature allows us to explore the helix bending for G79L mutant which could not be captured with earlier CG forcefields.

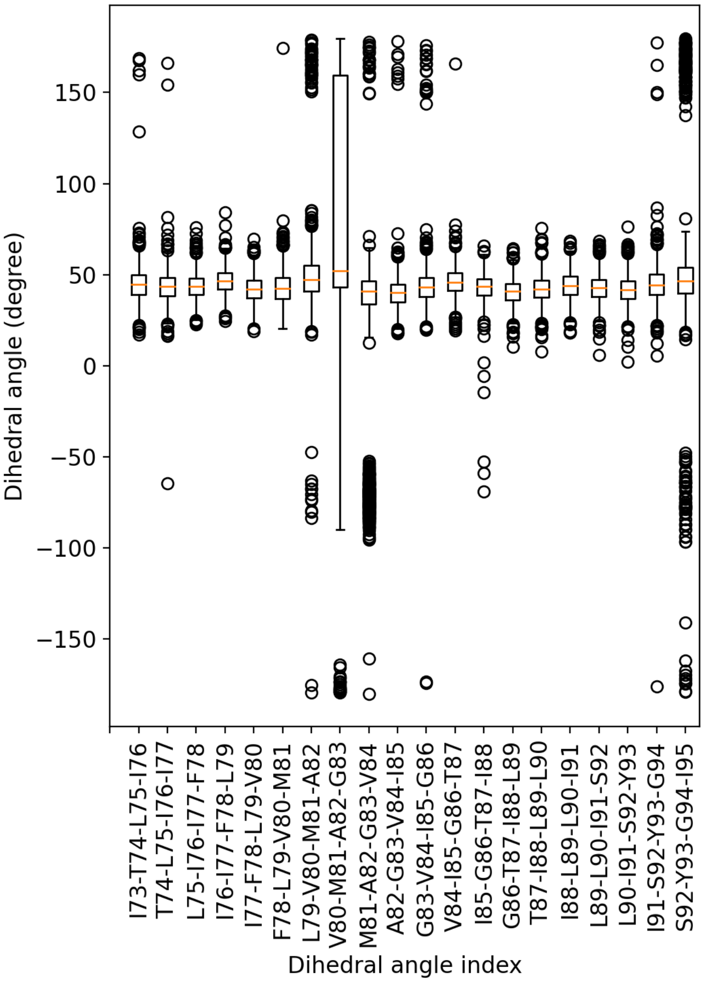


Figure 3.35: Dihedral angle analysis for G79L, where the x-axis shows the dihedral angle index and the y-axis shows the dihedral angle. The median is shown in orange, the box includes the interquartile range, the bars show the positions of minimum and maximum, and the circles are the outliers.

3.4.2.4 Crossing angle analysis shows difference in parallel dimer and anti-parallel dimer population

To characterize the relative orientations of dimer populations, we use the helix crossing angles between the monomers. In Figure 3.36, the majority of the crossing angles for the WT GpA span a range of -50 to 20 degree. The broad range of crossing angles comes from a variety of packing orientations, which could explain the large variation in the number of BB contacts from Figure 3.32 due to different packing fashion. A negative crossing angle indicates a right handed dimer, and thus from our simulations most of the WT GpA dimers are right handed. A skewed bimoidal distribution for the crossing angle is observed for the WT, which agrees with previous CG simulation results performed in a lipid membrane [201]. The collection of black lines centered around -40 degree in Figure 3.36 marks the crossing angles of the solution NMR GpA dimer structures [194] reported in the PDB. The deviation in crossing angles between the NMR structure and our results could come from the ”softer” nature of CG models, which has also been observed in previous CG simulations [201]. While few in number, we also record some anti-parallel dimers for the WT GpA, which have an absolute crossing angle larger than around 150 degree. Anti-parallel dimers have been reported before in atomistic simulations in both implicit membrane and implicit cyclohexane [242]. These dimers could have comparable energy as that of a native dimer. In cyclohexane, there is a higher probability of observing an anti-parallel dimer than in membranes due to the lack of a barrier preventing parallel-anti-parallel dimer flipping. From the residue-residue contacts in the contact map (Figure 3.33 (a)), it can be noted that the dimers are highly symmetric and specific. This specificity of contact surface was also observed in earlier CG simulations in a membrane environment where a mobile helix was anisotropically distributed near another fixed helix [201], with specificity originating from the close contacts at the GxxxG motif.

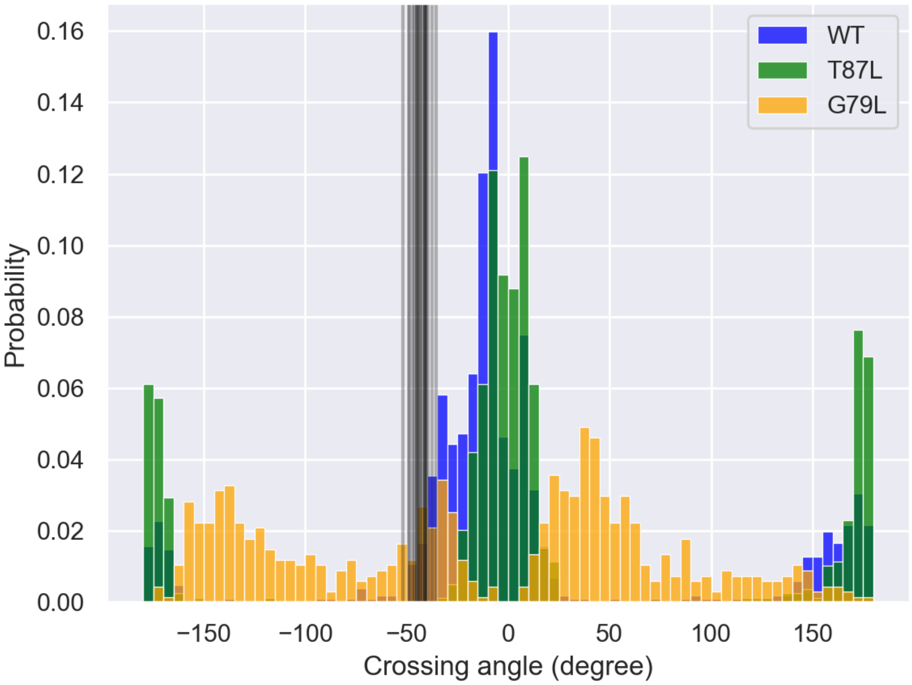


Figure 3.36: The crossing angle distribution for WT (blue), T87L (green), and G79L (orange). The x-axis is the crossing angle in degree. The y-axis is the probability of observing a specific crossing angle for the WT or the mutant.

Figure 3.36 shows the distribution of T87L crossing angles in green. Compared to the crossing angles for WT, the crossing angles for T87L deviate more from the solution NMR structures with a range of -35 to 25 degree. Within this region, parallel dimers with a crossing angle near 0 degree are more populated, which could explain the higher number of BB contacts observed from the PMF of T87L because the monomers are more parallel to each other and create more contacts. However, comparing the total parallel dimer probability (absolute crossing angle less than 50°), it is still more probable to observe parallel dimers in WT. On the other hand, significant amount of anti-parallel dimers are observed for T87L, which surpasses the amount of anti-parallel dimers for WT. The GxxxG contact is also observed for the parallel T87L dimers

(Figure 3.33 (b)), but in a more diffused fashion. While the GxxxG motif contributes largely to the dimerization of the GpA monomers, it is also reported that two third of the stabilizing intermonomer energy comes from outside the GxxxG motif [242]. In this case, the lack of the original T87-T87 interaction could be a reason why the parallel dimer is less sampled in the most populated basin for T87L compared to WT, as T87 has been reported to involve in hydrogen bonding between helices and help dimer stabilization [243]. T87L has been reported to form little to no dimers in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) experiment [198], but in our simulations we observe the formation of some parallel dimers with a lower probability compared to that for the WT. In a previous analytical ultracentrifugation study, mutants that were reported to be disruptive in SDS-PAGE experiment were observed to still form significant dimers in detergent C_8E_5 micelles, but less than WT [244]. In addition, in previous CG GpA simulations, the disruptive mutants reported by SDS-PAGE experiment are also observed with significant amount of dimers in lipid membranes [201, 202]. It could be that with the SDS-PAGE experiment, the amount of dimers was too little to capture. Our results agree qualitatively that parallel dimers for T87L are less stable than the WT dimers.

Finally, a much broader distribution of the crossing angles is observed for G79L due to the bent structure of individual monomers. The contact map for G79L further confirms the fact that there is no specific interactions existing between the two monomers for (Figure 3.33 (c)). Compared to the WT and T87L mutant, the contact map for G79L is significantly more diffusive with only observable contacts between the N-terminus of the two monomers.

3.5 Conclusion

In this chapter, we develop a polarizable CG model with explicit representation for the environment — *ProMPT*. The key objective here is to create a CG MD force field to capture tertiary folding of protein structures with minimal constraints. The CG mapping scheme follows closely the MARTINI CG force field geometry. In *ProMPT*, the polar CG beads have explicit drude-like charges that can couple a protein's environment to its structure and dynamics. We parametrize the non-bonded interactions between CG beads through free energies of their interaction with the environment. The bonded interaction potentials are generated through analysis of corresponding normalized bonded-feature distributions generated from a database of non-redundant protein structures from the protein data bank. This CG force field is validated by reproducing the conformational free energy landscape of several well-studied small protein systems. While *ProMPT* can accurately fold protein structures with homogeneous secondary structures, we observe some side-chain packing defects in folding of proteins composed with heterogeneous secondary structures. Our future research efforts will be focused on fixing sidechain packing through accurate mapping of sidechain geometry. Due to its transferability, *ProMPT* can be patched with our previous *in-house* lipid and divalent ion models to study protein folding assisted by physiological environment.

We further use *ProMPT* to investigate the effects of point mutations on GpA transmembrane folding in micellar conditions. In our initial validations with the wildtype GpA, we observe several structural characteristics as noted by previous literature such as the GxxxG contacts and the helix crossing angle. We explore the structural stability of the helical dimers through specific point mutations aimed at isolating individual interactions at the GxxxG region (G79L) and hydrogen bonding at the C-terminus (T87L). From the PMF plots, T87L can form helical dimers, but G79L looses some helicity, and can not form stable helical dimers. We find that point mutations at the GxxxG motif alter the local hydrophobic environment and contribute to helix bending through a hinge at the GxxxG region. Similar hinge structure at GxxxG has been previously reported for APP. But, even with the hinge we find significant helicity (around 90%) for G79L monomers. A significant fraction of WT exhibits right-handed dimer conformation, while there are more anti-parallel dimers observed for T87L, possibly due to a lack of the T87-T87 lock. For G79L, there are no specific crossing angles observed as there are no stable dimers. In addition, we observe diffused contacts suggesting a decrease in dimerization specificity. The ability to capture this secondary structure change is a novelty of the *ProMPT* model which can not be studied with previous CG models where restraints on secondary structures are imposed. This opens up future possibilities to use foldable and transferable CG models to understand conformational fluctuations and the impact of point mutations in physiological transmembrane proteins, where local environmental fluctuations can play a key role.

Chapter 4: Investigate the role of $[\text{TEA}]^+[\text{Ms}]^-$ on Abeta aggregation

4.1 Overview

We apply *ProMPT*, a CG MD model, to study the aggregation and secondary structure transformation of amyloid beta 16-22 (Aβ_{16-22}) in triethylammonium mesylate ($[\text{TEA}]^+[\text{Ms}]^-$), an alkylammonium based IL. It is found that both the β-sheet fraction and the cluster aggregation size decreases as the IL concentration increases, which agrees with the experimental findings. Moreover, the aggregated clusters exist in different forms in water and in the presence of IL. Aggregates formed in the presence of ILs are less stable and are lack of intact hydrophobic core, while aggregates formed in water have a stable hydrophobic core and form intact β-sheet interactions. $[\text{TEA}]^+$-Phe interaction is one of the driving forces causing the loose of β-sheet fraction by disrupting the Phe-Phe contacts. In addition, the helical content increases with increasing IL concentration, which is caused by the reduction of the inter-peptide contacts existing in β-sheets. From the single peptide studies, it is observed that $[\text{TEA}]^+$ can interact with the end of Aβ_{16-22}, creating a charge-dipole induce effect that enhances the helix alignment.

4.2 Introduction

Protein aggregation is of significant interest for both researchers and bio-chemical companies because of its role in neuro-degenerative diseases and industrial processes. The origins of several neuro-degenerative disorders such as Alzheimer's, Huntington's, Parkinson's, and prions diseases are suggested to be strongly related to the aggregation of corresponding proteins (amyloid-beta, polyglutamine, α-synuclein, and prions, respectively) [245]. On the other hand, protein aggregation can also be a major challenge faced in industrial pharmaceutical processes. For example, protein aggregation can interfere with therapeutic peptides, reduce protein activity, and lead to potential immunogenicity [246]. The existence of insoluable aggregation is a huge hindrance for the release of bio-products [247]. There are many factors that could induce protein aggregation, which include but are not limited to temperature, pH, protein concentration, ionic strength, mechanical pressure, and presence of additives [248]. These environmental changes are frequently present during bio-processes and hence make protein aggregation a critical event to control. Amyloid-beta (Aβ), which links to the Alzheimer's disease, is cleaved from amyloid precursor protein (APP) by β- and γ- secretase. Aβ_{42} fragment is considered to be the dominant species to form fibrils in human brain, and is more toxic than a slightly shorter fragment (Aβ_{40}) that also receives much investigation [249]. The most widely accepted model for the etiology of the Alzheimer's disease is the formation of extracellular amyloid plaques, although some literature have discussed that it is the oligomers causing the neuro-decay, not amyloid fibrils [250]. In either case, a full understanding of the mechanism behind oligomer/fibrils formation is highly desired. Moreover, the ability to modulate the level of Aβ aggregation will be strongly beneficial for future medication design.

Some ILs are reported to be able to modulate Aβ aggregation, either enhance or reduce fibrilization. Debeljuh et al. have done a set of experiments for Aβ_{16-22} in the presence of protic ILs [50, 251]. Aβ_{16-22} fragment is considered to be the smallest fragment of Aβ that can form hydrophobic fibrils. The authors studied both [TEA$^+$] based ILs ([TEA]$^+$[HSO$_4$]$^-$, [TEA]$^+$[H$_2$PO$_4$]$^-$, [TEA]$^+$[Ms]$^-$) and [Ms$^-$] based ILs ([EA]$^+$[Ms]$^-$, [DEA]$^+$[Ms]$^-$, and [TEA]$^+$[Ms]$^-$) and found out that in both cases, [TEA]$^+$[Ms]$^-$ suppressed the formation of fibrils the most at high concentration, while at low concentration, this suppression was less significant. In addition, the helicity of Aβ_{16-22} peptide was also found to increase in the presence of high concentration [TEA]$^+$[Ms]$^-$. They hypothesized that both [TEA]$^+$ and [Ms]$^-$ could play a role impeding the formation of aggregation, where cation showed less proton transfer and anion exhibited reduced salting-out effect. As for the computational studies, Dasari et al. performed umbrella sampling on Aβ_{16-22} in 10% and 70% [TEA]$^+$[Ms]$^-$ [252]. At low IL concentration, β-sheets are the most probable structures. However, if the concentration increased to 70%, the most probable conformation became α-helices, which validated the experimental results. They explained this shift of the structural minimum by the more ”membrane-like” environment high concentration [TEA]$^+$ created, which could induce α-helix formation. Since umbrella sampling is not feasible to study the secondary structure transformation and only two peptides are studied in the system, which does not allow aggregation study, further simulations can be of use to investigate the dynamics of conformational change of the Aβ_{16-22} peptide and fibrilation mechanism. In addition, detailed microscopic interactions between [TEA]$^+$[Ms]$^-$ and Aβ_{16-22} also need to be examined, which experiments are not capable to.

Due to the high viscosity of ILs, an atomistic representation of a protein in a high IL concentration mixture with classical MD simulations will take extraordinary time to reach convergence, which is not realistic and limits the exploration of different IL combinations and protein sizes. Even with enhanced sampling techniques, the convergence can still be out of reach and kinetics information could lost. CG MD simulations are therefore applied to ease the burden in sampling. CG MD groups similar atoms into functional sites and thus larger system size and longer timescales of interest can be explored by reducing the computational complexity. Here we apply our previously developed CG model (Protein Model with Polarizability and Transferability (*ProMPT*)) to study the effects of different $[\mathrm{TEA}]^{+}[\mathrm{Ms}]^{-}$ concentrations on $\mathrm{A}\beta_{16-22}$ fibrilization. We present unique structural features of $\mathrm{A}\beta_{16-22}$ with $[\mathrm{TEA}]^{+}[\mathrm{Ms}]^{-}$ and provide a mechanistic insight into how this IL modulates $\mathrm{A}\beta_{16-22}$ aggregation.

4.3 Methods

4.3.1 $[\mathrm{TEA}]^{+}[\mathrm{Ms}]^{-}$ model

The CG mapping scheme for $[\mathrm{TEA}]^{+}[\mathrm{Ms}]^{-}$ and water is shown in Figure 4.1 (a). $[\mathrm{TEA}]^{+}$ is represented by a triangle with a charged virtual site at the center. The beads on each side of the triangle are hydrophobic SC1 beads from *proMPT*. Each pair of beads is connected through a bond of 0.45 nm and with a force constant of 5000 kJ mol^{-1} nm^{-2}. The virtual site carries no mass and is charged positively. The virtual site is given a new bead type named D3 that is similar to D2 in *ProMPT*, but has an *ad hoc* cation-π interaction with the side chain beads of the aromatic groups (LJ based potential with σ = 0.47 nm and ϵ = 2 kJ/mol). $[\mathrm{Ms}^{-}]$ is represented by a single anion bead. All other *ProMPT* parameters can be found in our previous publication [131]. We use the water model developed by Yesylevskyy that has an 1:4 mapping (one CG molecule corresponds to four atomistic molecules) to represent water [208].

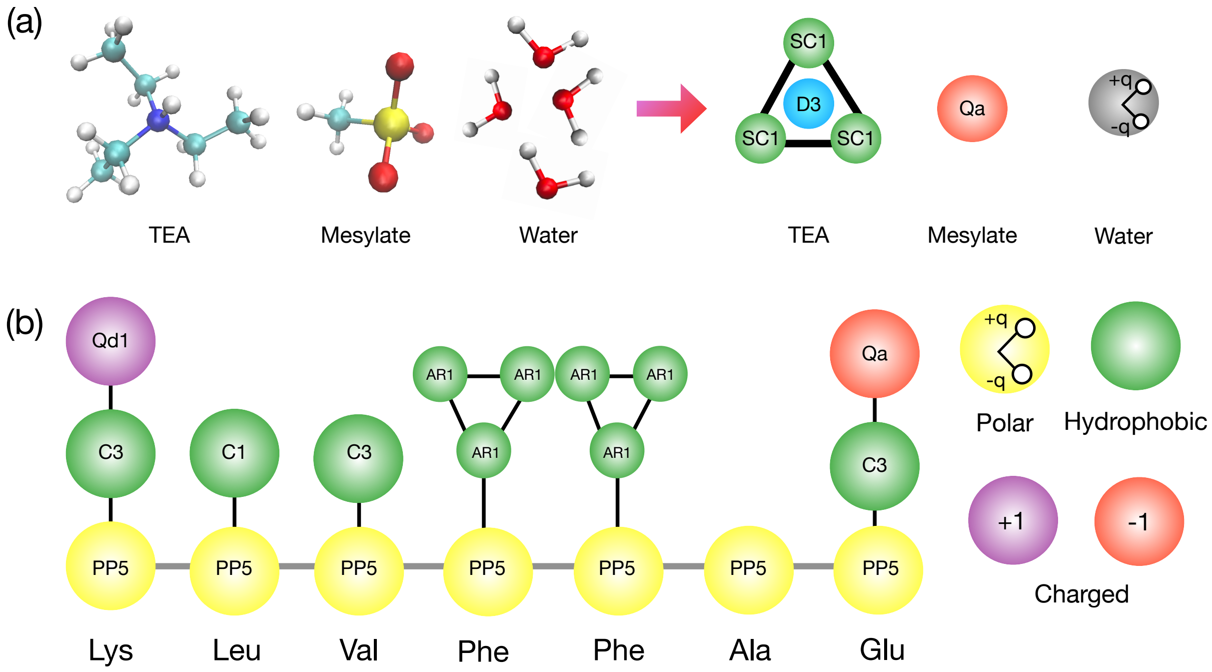


Figure 4.1: The CG mapping scheme for (a) $[\mathrm{TEA}]^{+}[\mathrm{Ms}]^{-}$ and (b) $\mathrm{A}\beta_{16-22}$ peptide. On the left side of (a) shows the atomistic scheme for the IL and water, where nitrogen is in blue, carbon is in cyan, hydrogen is in white, sulfur is in yellow, and oxygen is in red. On the right side of (a) is the CG scheme for the IL where the bead types are marked.

4.3.2 Aβ_{16-22} peptide model

Aβ_{16-22} has a sequence of KLVFFAE. We construct the Aβ_{16-22} peptide in an extended conformation. The CG mapping scheme is shown in Figure 4.1 (b), where all BB are represented by polar beads. The polar bead carries two dummies having opposite partial charges (BBp and BBm) that can move freely with different angles and result in change of dipole moments. The dihedral potential is applied to backbone CG beads. The potential used for Aβ_{16-22} is a double well potential as shown in Figure 4.2, where the minima occur at both β-sheets and α-helix without a bias. An *ad hoc* cation-π interaction is added between the side chain beads of Phe residues and the charged bead of Lys, in both inter- and intra- molecular fashion.

4.3.3 Simulation setup

Simulations are performed with Gromacs 2019.4 [216]. There are four systems for Aβ_{16-22} aggregation study with different IL concentrations (0 wt%, 10 wt%, 50 wt% and 90 wt% $[\text{TEA}]^+[\text{Ms}]^-$). All systems start with 8 randomly placed extended Aβ_{16-22} coils in a cubic box. The number of peptide, water, $[\text{TEA}]^+$, and $[\text{Ms}]^-$ are listed in Table 3.7. Two systems containing a single Aβ_{16-22} peptide in water and in 90 wt% $[\text{TEA}]^+[\text{Ms}]^-$ are also set up to investigate the effects of solvents on a single peptide. Energy minimization is first run and followed by a short NPT equilibration. A production run of 800 ns is performed for each system containing 8 peptides, while a production run of 500 ns is run for the single Aβ_{16-22} peptide systems. In all simulations, PME is used for long range Coulomb interaction with a cutoff of 1.6 nm [217]. vdW type cut-off is used with a cutoff of 1.2 nm. All systems are performed in an canonical ensemble using a Nos$\tilde{e}$-Hoover thermostat for temperature coupling [218].

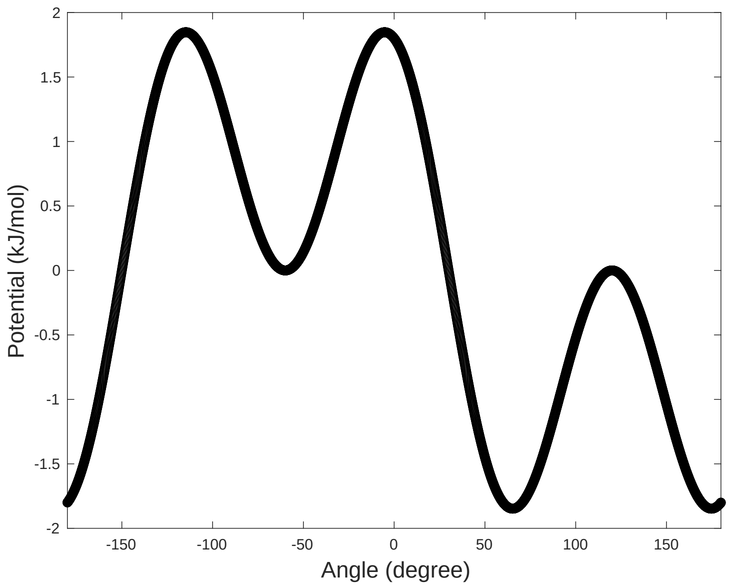


Figure 4.2: The double well dihedral potential used for Aβ_{16-22} simulations. The minima occur at both α helix and β-sheet.

Table 4.1: Simulation setup for each system.

	Water	10 wt%	50 wt%	90 wt%	Single peptide in water	Single peptide in 90 wt%
Number of peptide	8	8	8	8	1	1
Number of water	2584	2500	1095	500	1688	366
Number of $[\text{TEA}]^{+}$	0	101	400	1640	0	1200
Number of $[\text{Ms}]^{-}$	0	101	400	1640	0	1200

4.3.4 Analyses

To compute the β-sheet fraction, only the dummies (BBp and BBm) are taken into account. When a pair of BBp and BBm from different peptides comes within a cutoff of 3 Å, it is counted as an interacting pair. A contact between two Aβ peptides is defined when there are more than 3 interacting pairs between them. At any time, the system needs to have at least one aggregate containing more than three peptides in order to compute the β-sheet fraction, if not, a zero β-sheet fraction is assigned. To compute the aggregation size, only the BB are considered. A cutoff of 7 Å is used to determine if peptides are forming an aggregate, but not necessarily β-sheets. For the computation of Phe-Phe contacts, only the side chain ring beads at the outer most of the triangle are considered with a cutoff of 6 Å between those beads. A contact between $[\text{TEA}]^+$ and Aβ is defined when any CG bead of $[\text{TEA}]^+$ is within a cutoff of 6 Å from a specific residue. The average coordinate number of $[\text{TEA}]^+$ for each residue is then calculated for two scenarios. One scenario only considers peptides that are monomers, and the other one only considers peptides that are not monomers. In order to compute the helical fraction, the dihedral angle is calculated for each dihedral turn of a peptide. A dihedral angle within the range of 35 ° to 58 ° corresponds to an α helix turn and a dihedral angle within the range of 58 ° to 72 ° corresponds to a 3-10 helix. The helical fraction is calculated by dividing the number of specific helical turns by the number of all dihedral turns. Visualizing figures and RDF analysis are rendered and performed by VMD.

4.4 Results and discussion

4.4.1 Validation for the coarse grained ionic liquid model

As shown in Figure 4.3, $[TEA]^+$ is modeled as a triangle with a charged virtual site at the center and $[Ms]^-$ is modeled as a simple one bead anion. We first validate our CG IL model with the atomistic RDF. Both atomistic and CG simulations are performed for 25 wt% $[TEA]^+[Ms]^-$ for 10 ns and with temperature at 300 K and pressure at 1 bar. Amber99sb force field is used for the atomistic simulation. The center of the cation maps to the nitrogen in atomistic $[TEA]^+$, while the anion bead maps to the sulfur in atomistic $[Ms]^-$. As seen from the RDF of Figure 4.3, the locations and heights of the two peaks of the N-S interaction obtained from CG simulations agree well with atomistic simulations. In addition, the location of the first peak of the N-N interaction agrees well between atomistic and CG simulations, with the CG model giving a higher integration. The S-S interaction is harder for the CG model to reproduce the g(r), given the over simplified CG model for the anion, but the locations of the peaks are matched. The nano-structuring nature for high concentration ILs that have been reported in many literature before is also captured by our CG model. In both atomistic and CG simulations to observe nanostructuring, we perform simulations at 75 wt% and for 100 ns since the systems need time to develop nano structures. From Figure 4.4, the local aggregation of $[TEA]^+[Ms]^-$ can be seen clearly in both atomistic and CG simulations. Our model can not only reproduce the correct N-S interaction, but also the local network forming with respect to water.

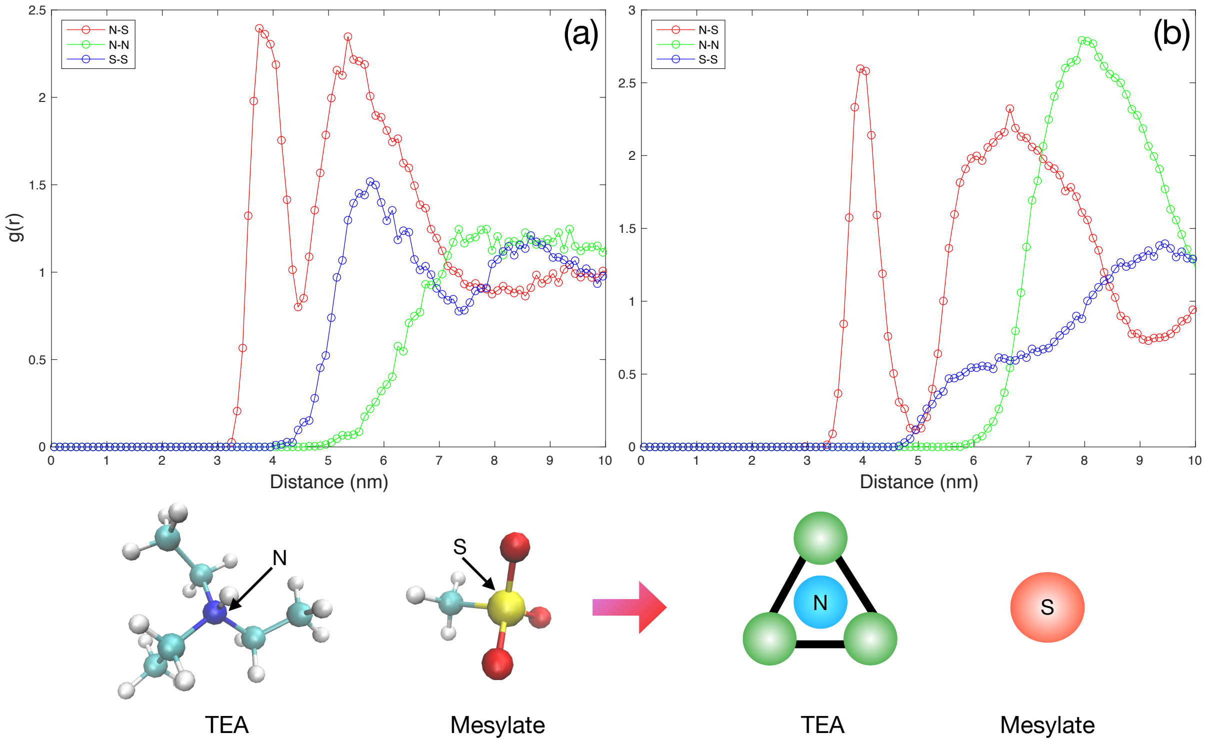


Figure 4.3: RDF between atoms/beads for (a) atomistic simulation and (b) CG simulation for 25 wt% $[\mathrm{TEA}]^{+}[\mathrm{Ms}]^{-}$. The CG mapping scheme is shown below, where in atomistic representation: Nitrogen: blue; Carbon: cyan; Hydrogen: white; Sulfur: yellow; Oxygen: red.

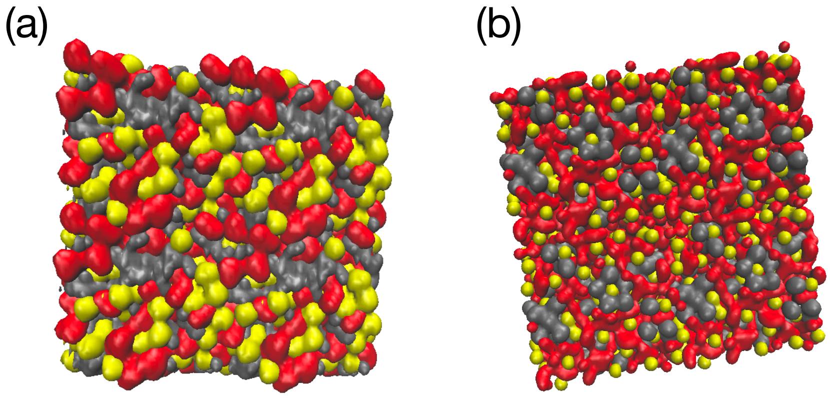


Figure 4.4: Snapshots of 75 wt% $[\mathrm{TEA}]^{+}[\mathrm{Ms}]^{-}$ in (a) atomistic simulations and (b) CG simulations showing the characteristic nano-structure for high concentration IL/water mixture.

4.4.2 The morphology is different for Aβ_{16-22} aggregates in water and in IL at different concentrations

Figure 4.5 shows the aggregation size time series at different IL concentrations. The aggregation size in water mostly stays at 8, which corresponds to a bulk aggregate that consists all peptides in the system. Figure 4.6 (a) shows a similar analysis but for β-sheet fraction. In water, an ordered single aggregate is formed within 100 ns, as shown in Figure 4.5 (a) and Figure 4.6 (a). On the representative Aβ_{16-22} aggregate (Figure 4.6 (b)), peptides form a cross-β sandwich structure as previously observed computationally/experimentally [253, 254]. In this cross-β sandwich structure, the Phe-Phe interactions are seen forming a hydrophobic core. The high number of Phe-Phe interactions in water can also be seen in Figure 4.7. After 400 ns, the peptides mostly form a single aggregate but local rearrangement causes fluctuations in β-sheet fraction from 0.6 to 1. A retardation of ordered aggregation is observed for 10 wt% $[\mathrm{TEA}]^+[\mathrm{Ms}]^-$. Even though the β-sheet fraction fluctuates between 0.5 to 0.8 after 350 ns, many small aggregates are observed in contrast to the simulation in water where a single aggregate is mostly observed after 100 ns. In this case, there are more than one aggregate and each aggregate forms some degree of β-sheet contacts, which is the reason for the high β-sheet fraction. Umbrella sampling results from Dasari et al also confirms β-sheet structures to be the most probable structure at this IL concentration [252]. The Phe-Phe contacts are less than those in water as seen in Figure 4.7. At 50 wt% $[\mathrm{TEA}]^+[\mathrm{Ms}]^-$, the β-sheet fraction generally stays at values less than 0.2, but at around 420 ns the β-sheet fraction reaches a value close to 0.6. This peak corresponds to a transient formation of β-sheets which then dissociates to oligomers. The aggregates formed at this IL concentration are not stable, as seen in both Figure 4.6 (a) and Figure 4.5 (c). In addition, the number of Phe-Phe decreases significantly, as shown in Figure 4.7. At 90 wt% $[\mathrm{TEA}]^{+}[\mathrm{Ms}]^{-}$ aggregate formation is impeded, although some small disordered aggregates (less than 6 peptides) are formed (Figure 4.5). The suppression of amyloid fibrilization at high concentration $[\mathrm{TEA}]^{+}[\mathrm{Ms}]^{-}$ is also shown in previous experimental studies [50, 251]. There is almost no Phe-Phe contacts as seen in Figure 4.7 and the peptides mostly stay as monomers or dimers.

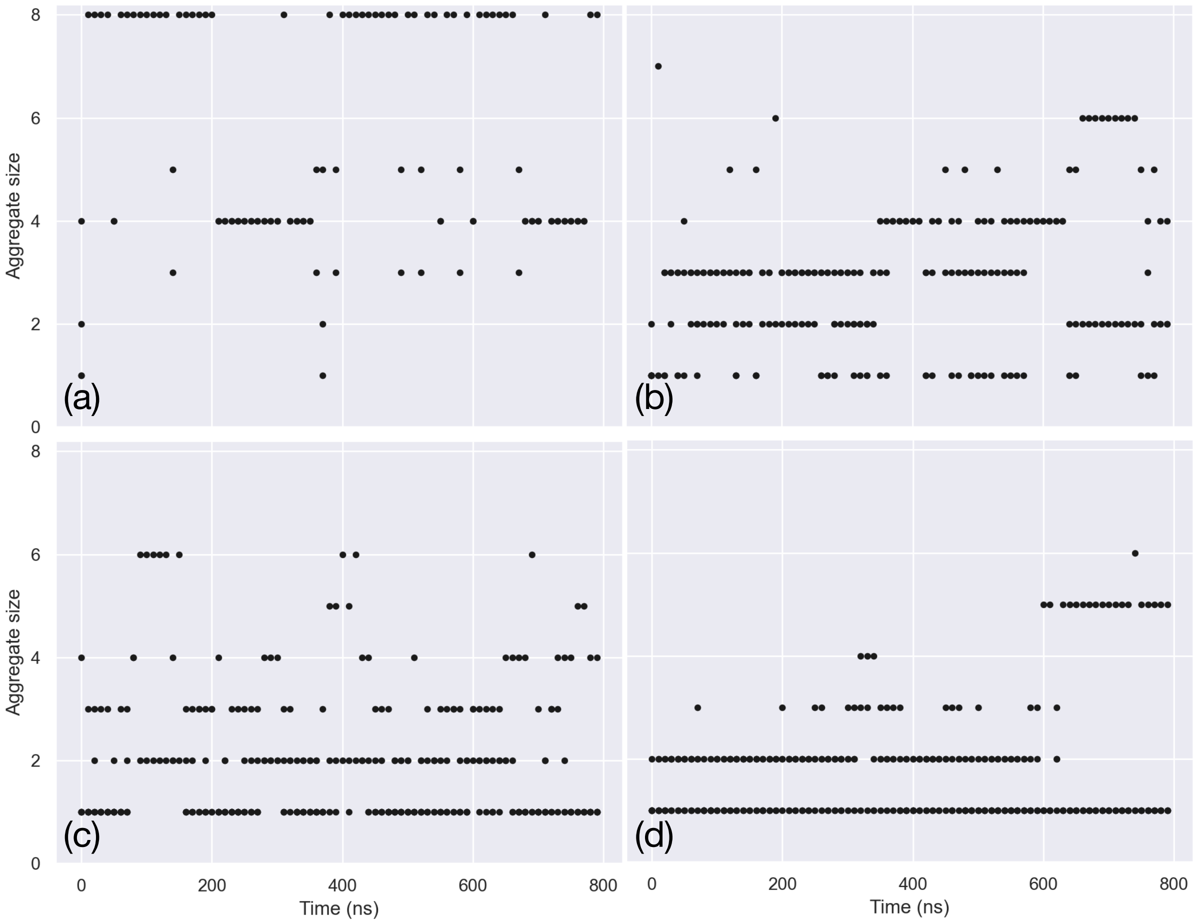


Figure 4.5: Aggregation size for $A\beta_{16-22}$ in (a) water and (b) 10 wt% (c) 50 wt% (d) 90 wt% $[\mathrm{TEA}]^{+}[\mathrm{Ms}]^{-}$. There are 8 peptides in each system.

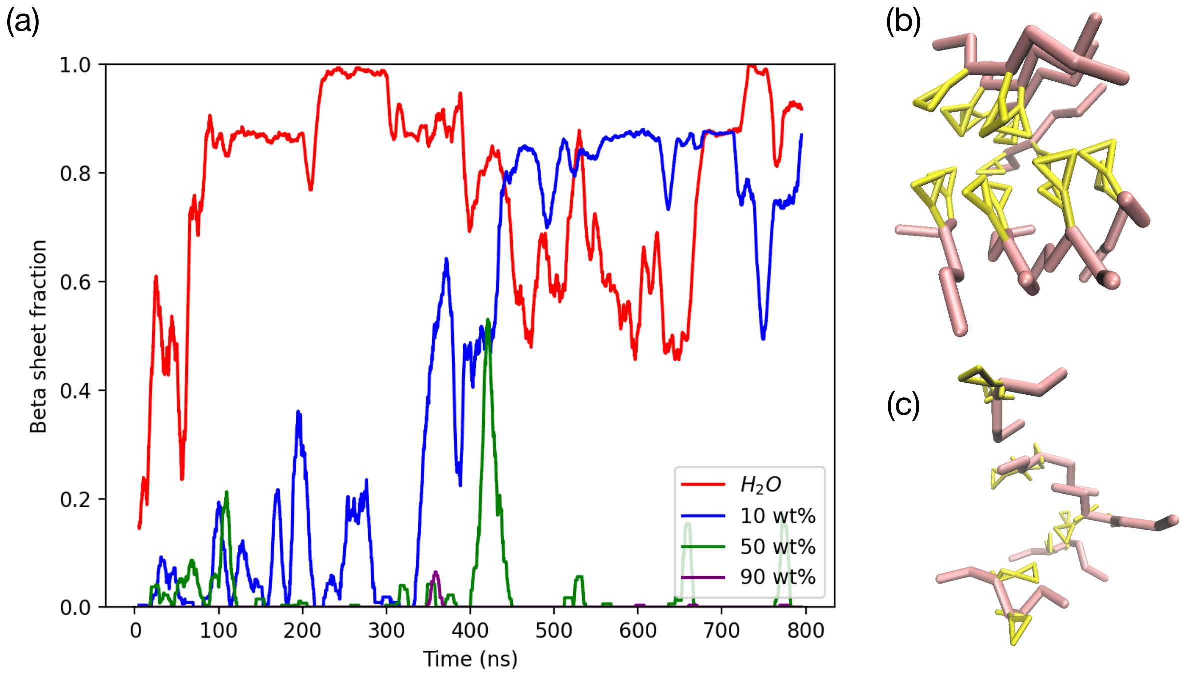


Figure 4.6: (a) The fraction of β-sheet for Aβ_{16-22} in water and [TEA]$^+$[Ms]$^-$ at various concentrations. Representative Aβ_{16-22} cluster in (b) water and in (c) 50 wt% [TEA]$^+$[Ms]$^-$. BB are in pink and Phe residues are in yellow.

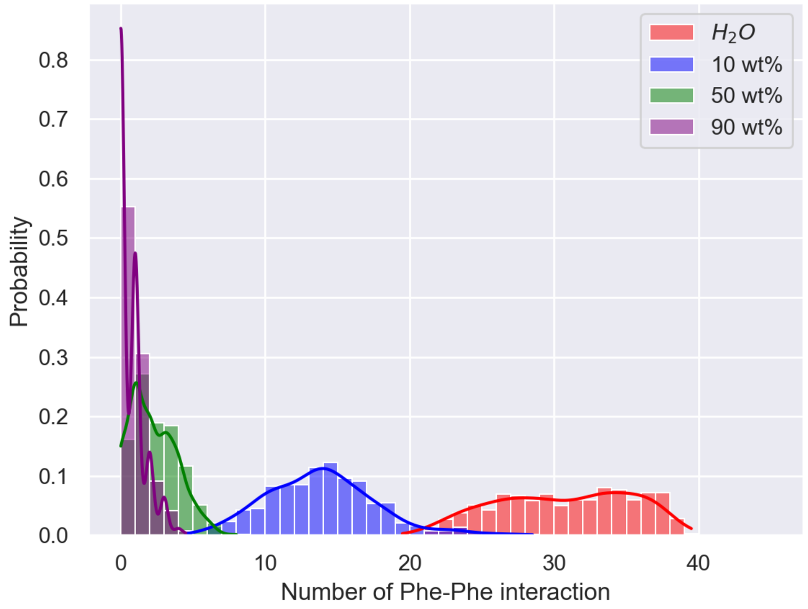


Figure 4.7: Phe-Phe contact distribution in water and in IL at different concentrations.

4.4.3 Mechanism behind the β-sheet disruption

The coordinate number of $[\mathrm{TEA}]^+$ for each $\mathrm{A}\beta_{16-22}$ residue when the peptide is a monomer is shown in Figure 4.8 (a). At all IL concentrations, there are significant higher numbers of contacts between $[\mathrm{TEA}]^+$ and Phe residues, which is a form of cation-π interaction. The $[\mathrm{TEA}]^+$-Phe contact is at its peak when at 90 wt% IL concentration, which could explain the strongly suppressed aggregation since the necessary hydrophobic core that leads to a stable aggregate is absent. There is also a high coordination of $[\mathrm{TEA}]^+$ around Glu residues, which could be explained by electrostatic attraction. The same analysis but for any $\mathrm{A}\beta_{16-22}$ peptide that is not a monomer is shown in Figure 4.8 (b). The coordination number of $[\mathrm{TEA}]^+$ for each residue decreases for every IL concentration, with Phe residues still make the most contacts with the cations. It is observed that in general, despite the difference in IL concentrations, the $[\mathrm{TEA}]^+$-Phe contact reduces when peptides approach each other due to Phe-Phe contact formation. At high IL concentrations, there are still significant large number of $[\mathrm{TEA}]^+$-Phe contacts (around 6 pairs), which could cause the instability of the aggregated structures.

4.4.4 Mechanism behind the change of helical fraction

Table 4.2 shows the α-helical, 3-10 helical, and the total helical fraction of $\mathrm{A}\beta_{16-22}$ in water and in different IL concentrations. The total helical fractions are similar in water and in 10 wt% IL, possibly due to the highly aggregated structures in 10 wt% IL. As the IL concentration increases, the total helical fraction also increases, which is also reported from previous experiments and simulations [251, 252]. From previous umbrella sampling studies, α-helices become the most probable structure at high $[\mathrm{TEA}]^+[\mathrm{Ms}]^-$ content, with the second probable structure

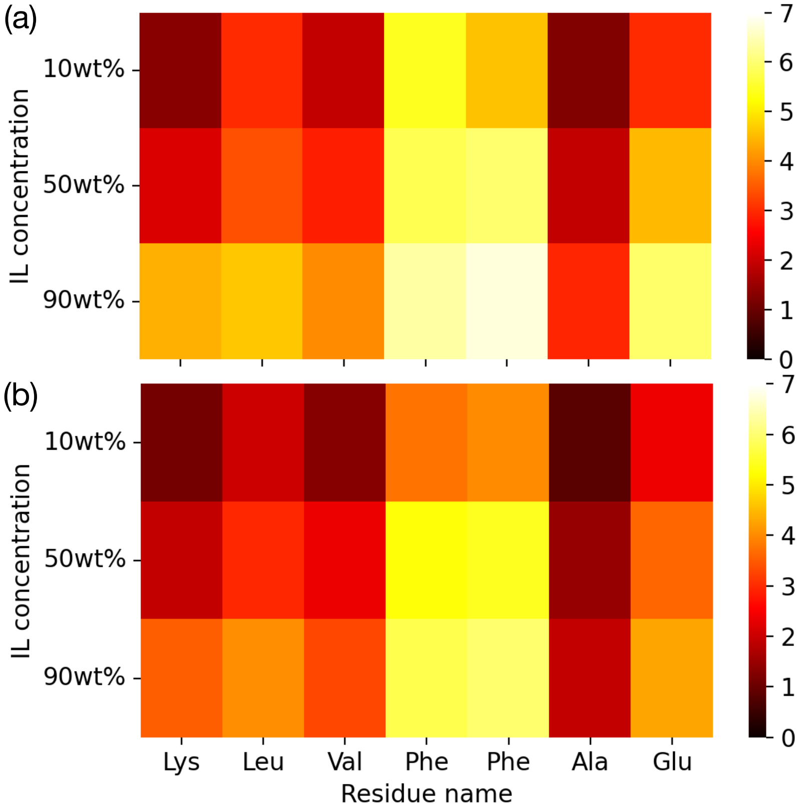


Figure 4.8: Coordinate number of $[TEA]^+$ for each residue for (a) monomers and (b) not monomers.

being coils [252]. The formation of β-sheets restricts the peptide from forming a helix. As the β-sheet fraction decreases, the helical fraction increases. Note that with *ProMPT*, neither the helical structure or the β-sheet structure is biased. The transition of secondary structure, particularly at higher IL concentrations, is not feasible to be captured by atomistic simulations because of the high solvent viscosity and long temporal scale for the conformation change. Single peptide studies are also performed for water and 90 wt% $[\text{TEA}]^+[\text{Ms}]^-$ and a higher helical fraction is found for the peptide in IL. In addition, the average helicity is higher than the system containing 8 peptides at the same IL concentration, which is due to the fact that some small oligomers can still form in the system containing multiple peptides that will lead to a decrease in helicity. Figure 4.9 (a) shows the $[\text{TEA}]^+$ coordinate number for each residue in the single Aβ system in 90 wt% IL. From the bar chart it can be observed that $[\text{TEA}]^+$ mostly interacts with Phe and Glu residues, which agrees with Figure 4.8. A snapshot of a helix interacting with $[\text{TEA}]^+$ in 90 wt% $[\text{TEA}]^+[\text{Ms}]^-$ is shown in Figure 4.9 (b). From this figure we can conclude that both ends can interact with the cations, but in a different fashion. The head residue (Lys) interacts with $[\text{TEA}]^+$ through the charge-dipole interaction, while the tail residue (Glu) interacts with $[\text{TEA}]^+$ through its negatively charged bead. In this case, the induced dipole alignment can promote the formation of a helix and lead to higher helicity. The charge-dipole induced helicity has also been reported in previous literature [255, 256].

Table 4.2: Helical fraction for systems containing 8 peptides and for system containing a single peptide.

	Water	10 wt%	50 wt%	90 wt%	Single Aβ in water	Single Aβ in 90 wt%
α-helical fraction (%)	10.95	11.14	12.29	14.08	15.54	18.36
3-10 helical fraction (%)	19.91	18.41	21.59	31.74	36.02	38.86
Total helical fraction (%)	30.86	29.55	38.25	45.82	51.56	57.22

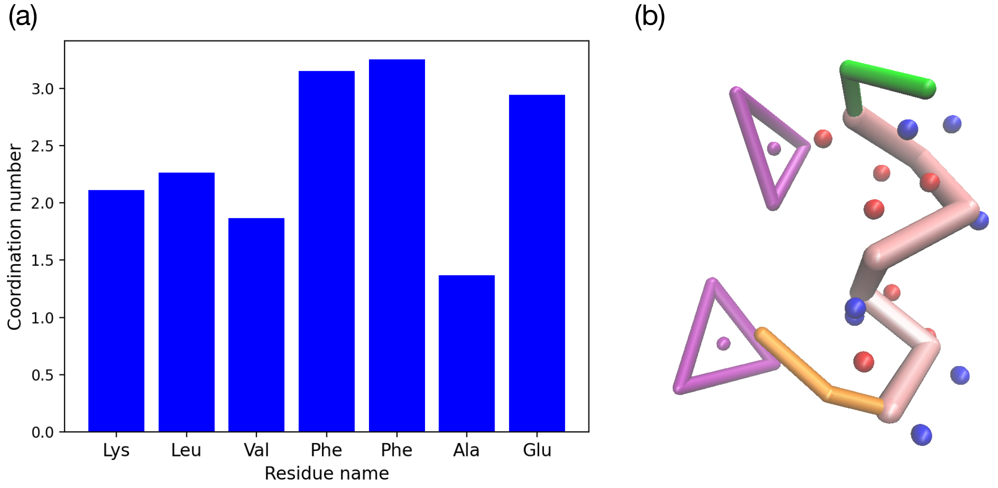


Figure 4.9: (a) $[\mathrm{TEA}]^+$ coordinate number for each residue for a single $\mathrm{A}\beta_{16-22}$ in 90 wt% IL. (b) A snapshot showing the interaction between $[\mathrm{TEA}]^+$ and the two ends of the $\mathrm{A}\beta_{16-22}$. Peptide backbone: pink; backbone negative dummy: red; backbone positive dummy: blue; Lys: green; Glu: orange; $[\mathrm{TEA}]^+$: purple.

4.5 Conclusion

In this work, we study the effects of $[\mathrm{TEA}]^+[\mathrm{Ms}]^-$ on $\mathrm{A}\beta_{16-22}$ peptide aggregation. In agreement with previous experimental findings, we found that the β-sheet content is disrupted as the IL concentration increases. The aggregated structures can still form in the presence of ILs, but are mostly disordered. The hydrophobic core of Phe residues present in the cross-β structure in water looses its stability in the presence of IL. The cause of β-sheet disruption origins from the interaction between $[\mathrm{TEA}]^+$ and the Phe residues that reduces the possible Phe-Phe contacts. At high IL concentrations, $[\mathrm{TEA}]^+$ can still interact with Phe of the oligomers and lead to instability of the whole aggregate. The helical fraction is observed to increase as IL concentration raises up, in agreement with the work of Debeljuh et al. and Dasari et al. [251, 252]. Formation of β-sheet reduces the occurrence of helix because of the forming of inter-peptide connections. The helical fraction for a single $\mathrm{A}\beta$ is also found to be increased in the presence of IL compared to that in water. In this case, the dominant interaction occurs between the $[\text{TEA}]^+$ and Phe/Glu residues. $[\text{TEA}]^+$ can interact with both ends of Aβ peptide through electrostatic interaction, but in a different fashion. $[\text{TEA}]^+$ interacts with Lys (N-terminal) through charge-dummy interaction and with Gly (C-terminal) through the side chain charged bead. Dipole alignment could be induced by these external $[\text{TEA}]^+$ contacts leading to an enhancement in helical fraction. The charge induced dipole alignment leading to an enhancement of helicity has also been reported before [255, 256]. This work also serves as the first work using simulations to study secondary structure transformation of peptides in ILs, which is not feasible for classical atomistic simulations and CG models that apply secondary structure constraints.

Chapter 5: Conclusion and Outlook

5.1 Overview

The main objective of this thesis is to investigate the effects of ILs on the modulation of protein properties. By studying the mechanisms behind the change of protein behavior caused by ILs, we aim to elucidate general rules on how particular IL functional groups interact with local/global protein structures. These general rules can serve as a universal guide to design protein-IL pairs that are useful in many bio-chemical applications such as vaccines preservation, medication, and reaction catalyst. Here we focus on the modulation of three protein properties: stability, activity, and aggregation. The proteins we select are lysozyme and Aβ_{16-22}. The former protein is a common protein found in human saliva and tears, while the second protein is linked to the pathogenesis of Alzheimer's disease. The ILs we study here include imidazolium based ILs and alkylammonium based ILs.

We approach the research questions mainly through molecular dynamics (MD) simulations. We also collaborate with an experimental group from the University of Massachusetts (Bermudez lab) to validate our simulation findings. Classical MD utilizes Newton's equations to calculate the time evolution of atoms based on a specific force field applied. A force field describes the potential based on the atom's coordinates. Different force fields have their own advantages and disadvantages and could be good at capturing properties for certain systems but not for others.

Here we use all-atom explicit MD simulations so both biomolecules and solvents (water or ILs) are described explicitly. MD is excellent in complement experimental findings as some limitations are posed by experiments including structural transience, contradicting explanations on experimental findings, and complexity in the de-convolution of the environmental factors.

In this thesis we have explored the effects of ILs on protein stability, activity, and aggregation. We found out that not only the type of cations and anions would affect protein behavior, but also the IL concentration. ILs tend to interact with specific regions of the protein and exert different effects. From our studies we have shown that $[\text{EMIM}]^+$ could disturb the sandwich long range interaction that stabilizes the tertiary structure of lysozyme and lower the denaturation temperature. In addition, both $[\text{EtSO}_4]^-$ and $[\text{Et}_2\text{PO}_4]^-$ interact with the lysozyme surface through electrostatic interactions and destabilize the whole protein. $[\text{EMIM}]^+$ could also interact with the lysozyme active site residues to produce a semi-blocking effect that decreases the stability of the ligand-protein complex. $[\text{TEA}]^+$, on the other hand, can modulate the aggregation degree of $\text{A}\beta_{16-22}$ peptides through its interaction with the Phe residues. These findings all point to important roles of cation-π, π-π, and electrostatic interactions between ILs and proteins. Especially with the imidazolium based ILs that have a ring for their cations, the interactions involve π electrons become more dominant. The mechanisms we have deduced in this thesis will help further constructing a guide for protein-IL pair design in many bio-chemical applications, despite no universal rule can be established for the complex IL-biomolecular interactions.

In chapter 2 we explore the effects of two imidazolium based ILs ($[\text{EMIM}]^+[\text{EtSO}_4]^-$ and $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$) on the stability and activity of lysozyme. These two ILs are found to not only destabilize lysozyme structure, but also reduce lysozyme activity. Both cation and anions are involved in the destabilization mechanisms, but in a different fashion. $[\text{EMIM}]^+$ reduces lysozyme stability through π-π and cation-π interactions with a local hydrophobic residue involving a long range interaction that stabilizes lysozyme. On the other hand, both anions destabilize lysozyme through short range electrostatic interactions with the protein surface. Strong interactions between $[\mathrm{EMIM}]^+$ and the active site residues are also observed to be a factor affecting lysozyme activity. The cations occupy the vicinity of the active site and possibly compete with the natural sugar ligand. In addition, umbrella sampling simulations point to a lower stability for the protein-ligand complex in the presence of ILs compared to water. Rehydration simulations show a fast leaving of $[\mathrm{EMIM}]^+$ from the active site, which explains the reappearance of the active site for recovered activity.

Simulations achieving large spatio-temporal scale are still not feasible with classical all-atom MD nowadays. For systems that have high viscosity such as IL solvents and systems that require long simulation time to reach convergence as many bio-molecular processes do, the all-atom classical MD becomes not realistic. Here, CG MD can thus come into play. CG MD groups similar atoms into functional sites to reduce computational complexity. For CG MD carrying transferability, these building blocks can be joined or associated freely to construct various systems in interest. In chapter 3, we update our in-house CG model and construct a new CG model named protein model with polarizability and transferability (*ProMPT*). The novelty of this model is the ability to capture secondary structure transition and protein folding-unfolding. There are three bead types in *ProMPT*: polar, hydrophobic, and charged beads. The polar bead has two dummy charges attach to the center bead. These two dummy charges carry opposite partial charges and can move freely with different angles to produce different dipole moments that respond to the change of environment dielectric constant. The bonded interactions in *ProMPT* are parametrized by performing Boltzmann inversion on the distribution of non-redundant protein structures taken from the PDB to give bonded potentials. Several *ad hoc* interactions such as π-π, cation-π, and CH-π interactions are also included to better describe some important intra-protein interactions. We validate *ProMPT* with several benchmark proteins including Trp-cage, Trpzip4, villin, ww-domain, and β-α-β. For Trp-cage, Trpzip4, and villin, both the free energy landscape and the representative folded structure can be well captured. The model can also reproduce the correct aggregate of Aβ_{16-22} in water. With *ProMPT*, we are able to capture folding-unfolding for proteins that have homogeneous secondary structures.

As a further demonstration for the application of *ProMPT*, we examine the dimerization of GpA WT and mutants. For the WT, parallel dimers are the most populated. The GxxxG contacts that have been observed from both experiments and previous simulations are also reproduced. In addition, the crossing angles also agree with experiments. For the mutant T87L, anti-parallel dimers are most populated. In this case, since the T87-T87 lock is absent, which is considered to stabilize the WT dimer, it could lead to a change in population ratio of parallel and anti-parallel dimers. For the second mutant G79L that has a mutated point located at the GxxxG region, no dimers are formed. Helices are observed to bend and have a hinge near the GxxxG region for G79L and no specific contacts are observed for this mutant. In this case, the alteration of secondary structure could possibly lead to change of protein behavior, not merely dimer stability. At last, we mark the novelty and potential application of *ProMPT* to be able to capture the secondary structure change due to mutations without constraints in a heterogeneous environment, which is not previously feasible for other CG models.

In chapter 4, we apply *ProMPT* and an in-house CG IL model to study the effects of $[\text{TEA}]^+[\text{Ms}]^-$ on Aβ_{16-22} peptide aggregation. Both the β-sheet fraction and the aggregation size decrease as the IL concentration raises up, which agree with previous experiments and simulations [50, 251, 252]. However, the morphology is different in water and in the presence of ILs. In water, a cross-β structure observed in previous experiments can form with a nice hydrophobic core [253, 254]. The β-sheet interactions are intact and results in a stable aggregate. In IL, the aggregates can still form but without a stable hydrophobic core. In addition, the Phe-Phe contacts are less than those in water and the formed aggregate can break into several oligomers. In general, in the presence of $[\text{TEA}]^+[\text{Ms}]^-$ it is more difficult to form β-sheets and even the β-sheet network can form, it is not stable, mainly due to the disruption caused by $[\text{TEA}]^+$-Phe contacts. A positive correlation between the helical fraction and the IL concentration is observed, in agreement with previous literature [50, 251, 252]. This observation could be explained by the disruption in inter-peptide contacts at high IL contents that provides higher probability for the peptide to form a helix, since the peptide do not need to build any inter-peptide connection. In this work, we explore how an IL can modulate the degree of protein aggregation through the tuning of IL concentration.

A contribution of this thesis is the construction of a novel CG MD protein force field. *ProMPT* can be applied to a variety of proteins in water, lipids, ILs, and detergent micelles. The novelty of *ProMPT* is the implementation of peptide backbone structural polarization. The idea is that the formation of protein backbone hydrogen bonds is critical in generating particular orientation of dipoles in secondary structures. We use a polar bead to represent the BB for each amino acid. For the polar bead, there is a central bead with two dummies carrying opposite charges. This set up results in an induced-dipole effect and adds a directionality to peptide backbone-backbone interaction which helps fold secondary structures. This model has been tested against small peptides and proven to have the ability to simulate folding and aggregation without adding external constraints. Furthermore, since this model has the roots in the popular CG MARTINI model, it is highly transferrable to study systems that have been previously parametrized. I have also extended this model to be able to apply to systems consisting of ILs. This model has the potential in the application of studying the change of protein secondary structure in various solvent environments and should be of good use to the Chemistry, Biophysics, and Materials Science computational communities.

5.2 Future work

The ultimate goal of understanding the interplay between different ILs and proteins is to unravel unique mechanisms in order to help design suitable ILs for protein preservation. Many of these studies start from globular proteins such as lysozyme, ribonuclease A, lipases, and cytochrome c, to name a few [17, 19, 27, 28, 29, 30, 32, 33]. The focuses of these studies are generally on the effects of ILs on specific tertiary structures or on certain long range non-bonded interactions, which could be difficult to pin point the effects of ILs on a specific secondary structure that can be then deduced to other proteins as a general rule. Therefore, it becomes desiring, and necessary to look at how certain type of ILs interact with mini proteins that persist unique secondary structures. We are currently working on the effects of $[\text{choline}]^+[\text{Cl}]^-$ and $[\text{EMIM}]^+[\text{Cl}]^-$ on two mini-proteins: Trp-cage and Trpzip4. Trp-cage, a 20-residue mini-protein, has an α-helix and a 3-10 helix. It also has a hydrophobic core formed by aromatic residues and Pro residues. Trpzip4, a 16-residue peptide, is a β-hairpin itself. The β-hairpin is intact due to the strong hydrophobic interactions from four Trp residues. These two mini-proteins have been studied broadly as benchmark systems due to their representative secondary structure and their small size that allows for a robust sampling of their configurational space.

Trp-cage, as a benchmark protein for the study of helix secondary structure, has been found both stabilized and destabilized with different ILs. $[C_4mpy]^+[Tf_2N]^-$ and amino acid based ILs with [EMIM] cations are found to stabilize Trp-cage [38, 257], while $[EMIM]^+[Act]^-$ denatures Trp-cage in a cold denaturation fashion. In this case, the helix is mostly intact, while the whole tertiary structure is loosen [258]. Trpzip4 and its natural occurred form - GB1 hairpin often serve as benchmarks for stable β-hairpin. Neat $[C_4mpy]^+[Tf_2N]^-$ is observed to denature Trpzip4 from experiments. On the other hand, [EMIM] based ILs with different anions can exert different degree of denaturation on GB1 hairpin [259]. In the case of $[EMIM]^+[acetate]^-$, the anions are determined to be the driving force for denaturation, not the cations [260]. It occurs that, similar to large globular proteins, the stabilization or destabilization of these mini proteins also depend on the combination of IL cation and anion, and more comprehensive studies combining both experiments and simulations are needed.

ProMPT is under constant update with our lab. Currently its application has been extended to include environments that consist of divalent cations, lipids, and sterol varieties. In addition, our next step is to modify the shape and size of the amino acid representation in the model to better describe the residue topology. From our previous experiences with proteins consisting of heterogeneous secondary structures we found that improved shape and size of the amino acids are needed in order to produce correct core packing for large proteins. A systematic CG model for non-traditional solvents, which include but are not limited to ILs, is also needed for the science community. Many biological processes take place in water or in membrane lipids and therefore these environments are usually the focus points most biological research concern. However, non-traditional solvents such as ILs and deep eutectic solvents start to emerge as new solvent environments that are able to modulate protein behavior in the past decade [261, 262, 263, 264]. These non-traditional solvents are shown by experiments to have a strong potential in the applications of vaccine additives, bio-product solvents and protein stabilizers and thus draw many research interests. In order to understand the mechanisms behind how these non-traditional solvents modulate protein behaviors, computational simulations are critical to obtain the molecular insights. Current computational models for non-traditional solvents are in general costly in computational resource and are made for only a small number of solvent categories. Therefore, it seems desiring for a novel multi-scale computational model for these non-traditional solvents, keeping in mind both the accuracy and transferability. In addition, a handy platform for researchers to generate materials needed for proteins in non-traditional solvent simulations can also be constructed. This new model for non-traditional solvents can be applied in other than the biological field since non-traditional solvents have also been reported to have superior potential in electrolytes, catalysis, and nuclear fuel reprocessing. Success in building such a model for non-traditional solvents can elucidate the molecular mechanisms behind the effects of these solvents to bio-systems and lead to a rational design for biochemical products. Moreover, in the case of disease, these studies will help develop alternative systems for vaccine preservation that are more economic and safe.

Appendix A: Experimental results for evaluating the role of $[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$ and $[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$ on lysozyme stability and activity

A.1 Methods

A.1.1 Materials

$[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$ ($\geq$ 95%), $[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$ ($\geq$ 98%), and lysozyme ($\geq$ 90%) from chicken egg white were purchased from Sigma-Aldrich. All chemicals were used as received. *M. lysodeikticus* lyophilized cells were purchased from Worthington biochemical corporation. Potassium phosphate monobasic ($> 99\%$) and potassium phosphate dibasic (99.5%) were purchased from Fisher Scientific.

A.1.2 Sample preparation

Lysozyme solutions (10 wt%) were prepared by dissolving 30 mg of lysozyme in 270 mg of various IL–water (17 to 75 wt%) or buffer (50 mM potassium phosphate, pH 6.2), and stored at room temperature. Lysozyme samples were subsequently rehydrated for circular dichroism, activity assays, and differential scanning calorimetry measurements. For lysozyme samples prepared in phosphate buffer this same buffer was used for rehydrations; for samples prepared in IL-water mixtures, water was used for rehydrations.

A.1.3 Calorimetry

Differential scanning calorimetry (DSC) measurements were performed with a TA Q100 with refrigerated cooling system. Approximately 10 mg of solution was hermetically sealed in an aluminum pan and an empty pan used as a reference. For measuring the unfolding temperature T_m, samples were heated from 20 to 100 °C with a heating rate of 2 °C/min. The instrument was calibrated using sapphire and indium. Every experiment was repeated thrice, and average value reported.

DSC experiments for the samples having 0.1 wt% lysozyme were performed with a MicroCal VP-DSC. All samples were degassed with a MicroCal ThermoVac for 10 minutes before DSC runs. Sample cell reservoir was filled with lysozyme-IL-water solutions while reference cell reservoir was filled with equal volume of IL-water solution. Temperature scans were run from 20 to 100 °C with a scan rate of 90 °C /hour, prescan thermostat was set at 5 min and post cycle thermostat was 25 °C.

A.1.4 Fluorescence Spectroscopy

A photoluminescence spectrometer from Photon Technology International was used for the fluorescence measurements. The intrinsic fluorescence of lysozyme was monitored with an excitation wavelength of 280 nm in quartz cuvette having a path length of 10 mm. The emission spectra were recorded between 285 to 485 nm keeping the excitation and emission slit widths at 0.5 mm. All above experiments were performed at room temperature.

A.1.5 Activity assay

In one activity assay, we followed the manufacturer's instructions (Sigma-Aldrich), based on the lysis of *Micrococcus lysodeikticus*. First, a suspension of *M. lysodeikticus* was prepared at 0.15 mg/ml in 50 mM potassium phosphate buffer, pH 6.2. Stock lysozyme solutions at 10 wt% were diluted to 0.001 wt% (400 units/ml), the recommended concentration, by using phosphate buffer. To begin the activity assay 80 μL of the lysozyme solution were mixed with 2 mL of the *M. lysodeikticus* suspension. Activities were determined by measuring the decrease in turbidity at 450 nm over the initial 5 min at 20 °C. Turbidity was recorded on a UV-3600 Shimadzu UV-VIS-NIR spectrophotometer with quartz cuvette having path length of 10 mm.

In another activity assay the aim was to maintain a constant solvent composition. To do so, all samples were separately prepared using identical IL-buffer concentration (e.g., 0.5 – 17 wt%). *Micrococcus lysodeikticus* lyophilized cells were prepared at 0.15 mg/ml and lysozyme solutions were prepared at 0.001 wt% (400 units/ml). To begin the activity assay 80 μL of the lysozyme solution were mixed with 2 mL of the *M. lysodeikticus* suspension. As in the first assay, activities were determined by measuring the decrease in turbidity at 450 nm over the initial 5 min at 20°C. Turbidity was recorded on a UV-3600 Shimadzu UV-VIS-NIR spectrophotometer with quartz cuvette having path length of 10 mm. Both activity assays are schematically depicted in Figure A.1.

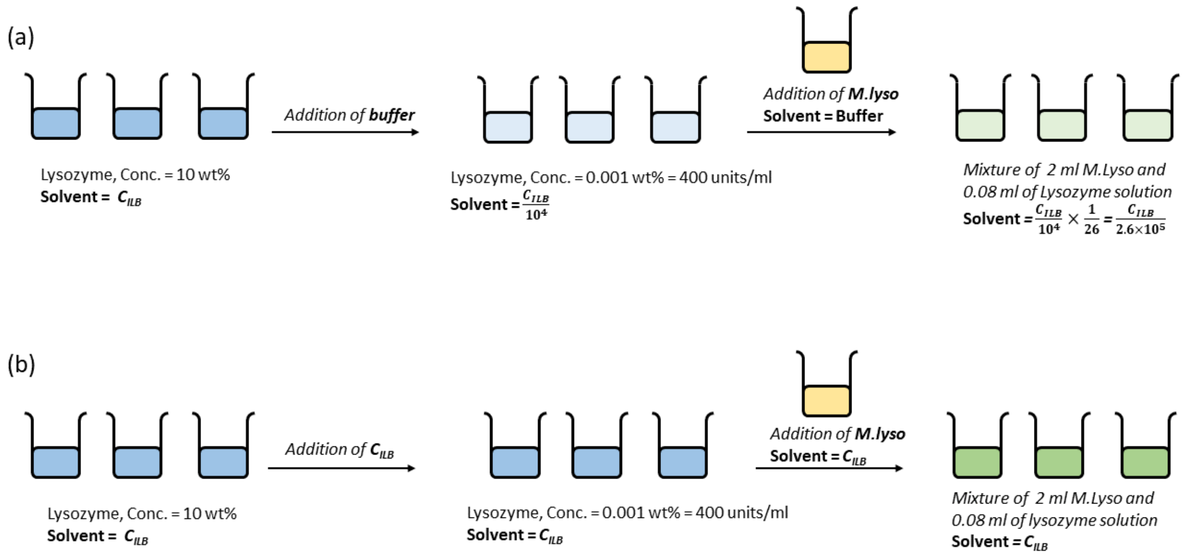


Figure A.1: Schematic of activity assays. (a) ”Traditional” assay where samples are diluted with phosphate buffer to achieve required final enzyme concentration of 400 U/mL. (b) ”Modified” assay where samples are prepared and mixed at a constant composition, C_{ILB} refers to IL-buffer mixture from 5% to 17% IL.

A.1.6 Circular dichroism (CD) Spectroscopy

The CD spectra were recorded in the far-UV region (190–260 nm) on a Jasco J-1500 spectrophotometer at 20±1 °C, with a quartz cuvette having path length of 1 mm. Both the data pitch and bandwidth were set to 1 nm with scanning speed of 100 nm/min.

A.2 Results and discussion

A.2.1 Evaluate the role of $[EMIM]^+[EtSO_4]^-$ and $[EMIM]^+[Et_2PO_4]^-$ on lysozyme stability

A.2.1.1 Compare the stability of lysozyme in water and in ILs/water mixtures

Calorimetric scans of lysozyme in IL-water solutions provide a great deal of information regarding the protein state and its transitions. Upon heating, lysozyme shows a well-defined endothermic peak, consistent with thermally-induced protein unfolding. The location of this peak corresponds to the melting temperature T_m. Second, with appropriate baseline correction [265], the enthalpy of the transition ΔH is directly given by the area under the curve. Third, integration of the curve allows us to determine the transition width ΔT, an important parameter of the unfolding process [266].

Figure A.2 shows how the lysozyme melting temperature T_m varies in the presence of both ILs. There is an initial linear regime, followed by a stronger decrease at higher IL content. This trend indicates a gradual effect of the ILs on lysozyme stability and is in contrast to strong denaturants, such as guanidinium chloride, which causes a rapid decrease in T_m and complete unfolding near room temperature [1, 267, 268]. Comparing the effect of the two ILs, $[EMIM]^+[Et_2PO_4]^-$ lowers the lysozyme T_m more readily than $[EMIM]^+[EtSO_4]^-$, with the difference most noticeable above 30 wt% IL.

The thermal denaturation of lysozyme in other IL-water mixtures has been previously reported [76, 82, 126, 269]. Specifically, calorimetric determinations of lysozyme T_m can either

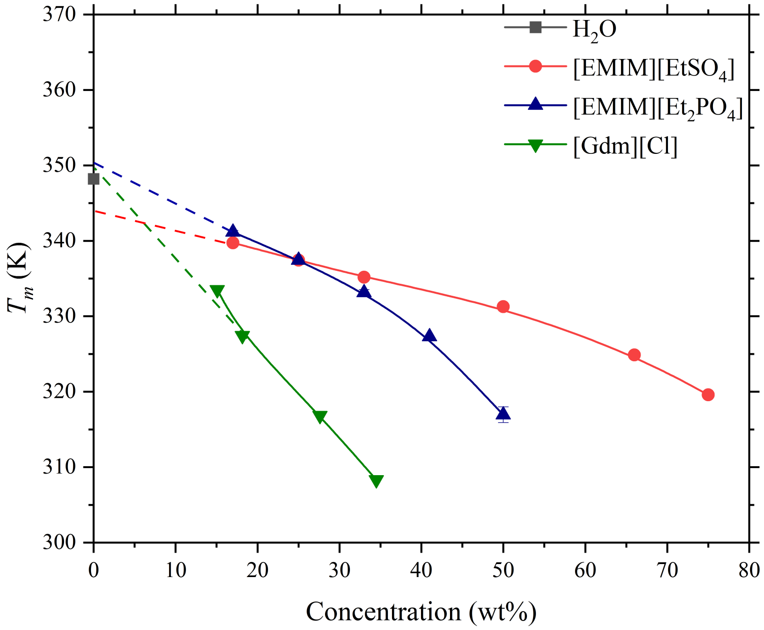


Figure A.2: Effect of IL concentration on lysozyme melting temperature T_m from calorimetry. For comparison, data for the effect of [Gdm][Cl] is also shown[1].

show a decrease or an increase with added IL. Such variation is not limited to lysozyme, as there are similar instances for RNaseA in IL-water mixtures [81, 95]. It seems clear that there is no simple rule to explain such effects and instead we must turn attention to specific interactions between the IL ions and protein functional groups.

The energy required for protein transitions can be directly measured by DSC. Figure A.3(a) shows how the lysozyme transition enthalpy ΔH varies in the presence of both ILs. With added [EMIM]$^+$[$EtSO_4$]$^-$, ΔH first gradually decreases and then reaches a plateau at intermediate concentrations. With added [EMIM]$^+$[Et_2PO_4]$^-$, ΔH decreases continuously with IL concentration. As with T_m, the difference between [EMIM]$^+$[$EtSO_4$]$^-$ and [EMIM]$^+$[Et_2PO_4]$^-$ is most noticeable above 30 wt% IL. The lower values of ΔH indicates less heat (energy) is required to unfold lysozyme in [EMIM]$^+$[Et_2PO_4]$^-$, and the relative impact of this IL is more clearly seen in ΔH as compared to T_m.

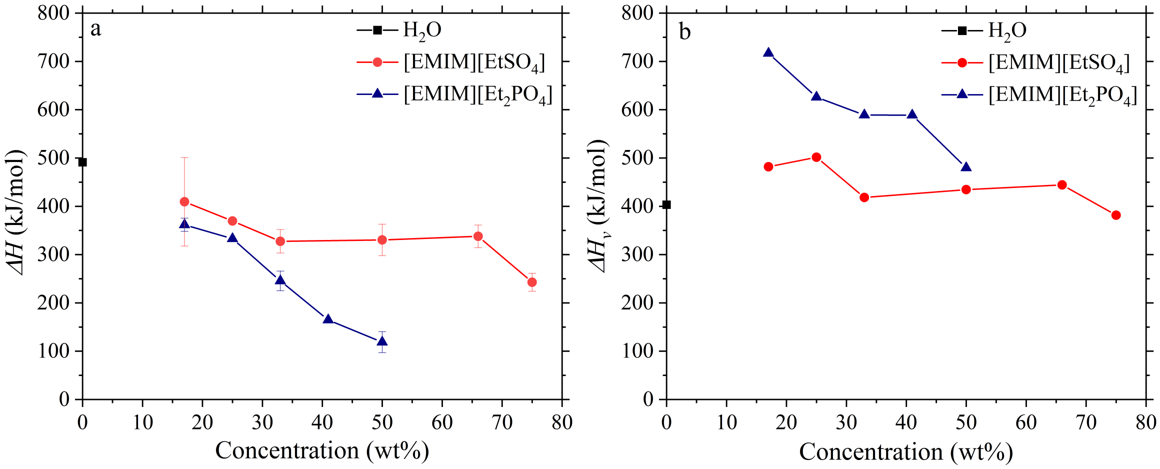


Figure A.3: Effect of IL concentration on the lysozyme (a) transition enthalpy ΔH and (b) effective transition enthalpy ΔH_v from the van't Hoff equation.

The melting temperature T_m of a protein is often quoted as a measure of thermal stability, but from the relation $T_m = \Delta H/\Delta S$ it is seen that T_m reflects two quantities. In both [EMIM]$^+$[EtSO$_4$]$^-$ and [EMIM]$^+$[Et$_2$PO$_4$]$^-$ solutions, the decrease in lysozyme ΔH (Figure A.3(a)) is accompanied by a decrease in ΔS. Specifically for [EMIM]$^+$[Et$_2$PO$_4$]$^-$, the small ΔS at high IL concentrations is consistent with a smaller difference between the initial (folded) and final (unfolded) states, and suggests a "loose" initial structure.

The shape of the melting curve determines if the thermal transition reflects "all-or-none", partial, or aggregate unfolding [266, 270, 271]. Intuitively it is understood that a small transition width ΔT is indicative of high cooperativity, whereas a large ΔT is indicative of low cooperativity, aggregation, or polymorphism. The van't Hoff equation $\Delta H_v/RT^2 = d\ln K/dT$, where R is the gas constant and K is the equilibrium constant, is used to *calculate* the transition enthalpy as $\Delta H_v = 4RT_m^2/\Delta T$. If the calculated ΔH_v equals the calorimetrically-measured ΔH, then the assumption of a two-state equilibrium is satisfied and the protein melts as a single cooperative unit. Comparing ΔH and ΔH_v (Figure A.3), we conclude that lysozyme in [EMIM]$^+$[EtSO$_4$]$^-$

melts nearly as an all-or-none process, whereas lysozyme in $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$ melts in a complex fashion, possibly as aggregates [272]. We note here that turbidity is observed for samples having very high IL content (> 66 wt% for $[\text{EMIM}]^+[\text{EtSO}_4]^-$ and > 40 wt% for $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$), however, these samples remain indefinitely stable suspensions (observed for several weeks).

A.2.1.2 Cation effects on lysozyme stability

As an experimental method to follow changes in lysozyme structure, we used intrinsic protein fluorescence. Lysozyme's fluorescence emission is predominantly due to tryptophan (Trp) residues [273]. Out of the six Trp residues, only two of these (Trp62 and Trp108) are thought to contribute [89], the others (Trp28, Trp63, Trp111, and Trp123) are either buried within the hydrophobic core or quenched by nearby cysteine residues. Trp62 is especially important for lysozyme stability, since its mutation to glycine results in a lysozyme variant with disrupted long-range interactions [123].

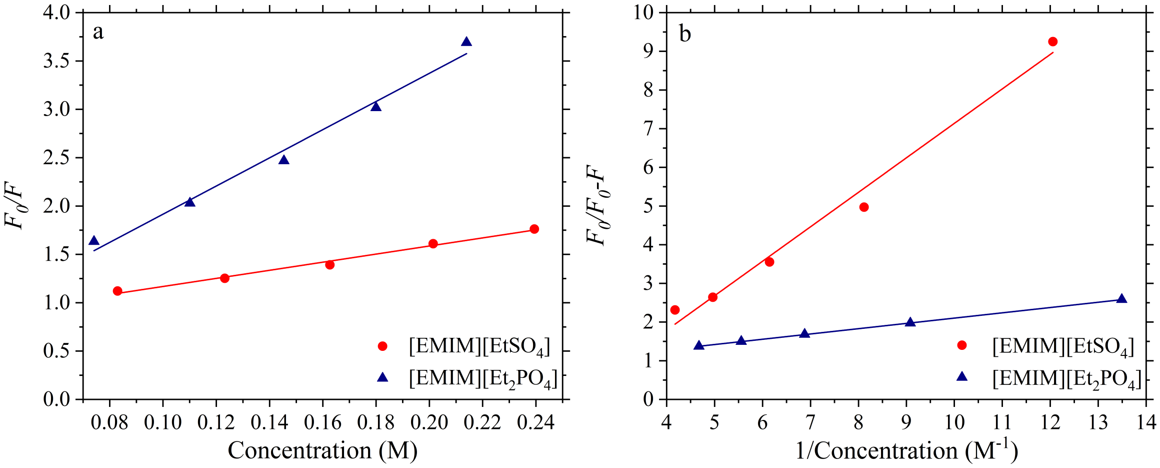


Figure A.4: (a) Stern-Volmer plot and (b) Lineweaver-Burk plot from fluorescence quenching of lysozyme with added IL.

The fluorescence data are presented in two ways: Stern-Volmer and Lineweaver-Burk plots (Figure A.4(a) and (b), respectively). In this manner we can discriminate between dynamic and static quenching. Both the Stern-Volmer equation $F_0/F = 1 + K_{SV}[C]$ and the Lineweaver-Burk equation $1/(F_0 - F) = 1/F_0 + 1/K_{LB}F_0[C]$ give their corresponding binding constants from the slopes of appropriate fits to the data. The fits to the experimental data are good, and Table A.1 reveals that $K_{SV} > K_{LB}$, indicating a predominantly dynamic binding interaction between the ILs and lysozyme. This result is in contrast to a related study by [89], who used $[\mathrm{BMIM}]^+[\mathrm{BF_4}]^-$ and $[\mathrm{BMIM}]^+[\mathrm{X}]^-$ where $[\mathrm{X}]^-$ are halides. They found predominantly static binding with lysozyme, presumably due to the increased cation hydrophobicity in their study.

Table A.1: Binding constants obtained from Stern-Volmer and Lineweaver-Burk plots.

System	K_{SV}, M^{-1}	K_{LB}, M
$[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$ + lysozyme	4.3	1.1
$[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$ + lysozyme	14.5	7.3

Table A.1 shows that $[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$ has a larger binding constant than $[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$.

A.2.2 Evaluate the role of $[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$ and $[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$ on lysozyme activity

To test the ability of ILs to preserve lysozyme over time, this protein was dissolved in various IL-water mixtures and maintained at room temperature. At selected times, aliquots were taken and rehydrated to different extents by the addition of water (or buffer in the case of control samples). The extent of rehydration is given by the ”dilution factor” df, i.e., the ratio of final to initial volume.

A.2.2.1 Lysozyme secondary structure is preserved after rehydration

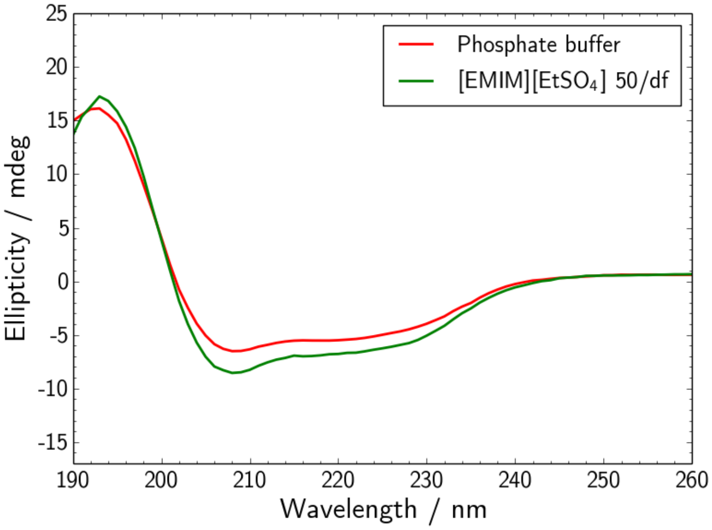


Figure A.5: Far-UV CD spectra of lysozyme. Freshly prepared sample in phosphate buffer (red), freshly prepared in 50 wt% $[\text{EMIM}]^+[\text{EtSO}_4]^-$ and rehydrated with water (green) The dilution factor df $= 1000$.

To probe the secondary structure of lysozyme, we used far-UV circular dichroism (Figure A.5). We find that lysozyme maintains its secondary structure after rehydration from high IL content, showing characteristic features near 210 and 230 nm [274]. Prior to rehydration, circular dichroism cannot be performed at high IL content due to absorbance by the imidazolium cation [275].

A.2.2.2 Lysozyme activity is restored after rehydration

To determine the functional capabilities of lysozyme, we measured enzymatic activity using a well-known turbidity assay [276]. Following the manufacturer's instructions corresponds to a dilution factor df $= 26$, and we realized that this large change in solvent environment may explain literature reporting on ”increased protein activity in the presence of all ionic liquids examined" [91], among similar results [39, 80, 94, 105]. In our case this dilution was even more significant because of the high lysozyme content, such that an additional dilution was needed to bring the lysozyme concentration to the suggested 400 units/mL, for an overall df $= 2.6 \times 10^5$. Nevertheless with this typical activity assay and its corresponding rehydration, activity values similar to controls are obtained (Table A.2). Therefore lysozyme activity is *recovered* upon sufficient rehydration with water, but we remained curious as to the effect of the IL on activity.

Table A.2: Normalized lysozyme activity using the typical activity assay, under different conditions. Prior to rehydration all samples here showed no detectable activity, i.e., activity $= 0\%$. The dilution factor df $= 2.6 \times 10^5$.

	$[\text{EMIM}]^+[\text{EtSO}_4]^-$	$[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$
IL conc., wt%	activity, %	activity, %
17/df $= 6.5 \times 10^{-5}$	115	110
33/df $= 1.3 \times 10^{-4}$	116	98
50/df $= 1.9 \times 10^{-4}$	114	87

To address the effect of the IL on activity more directly, we performed a separate activity assay where the IL content was maintained at a constant level throughout the assay. With this modified activity assay, there is a *complete absence* of lysozyme activity at even moderate IL levels (greater than or equal to 10 wt%). Figure A.6 shows a steep change in activity as $[\text{EMIM}]^+[\text{EtSO}_4]^-$ is increased, with the 50% activity level corresponding to approximately 3 wt% IL. Similar results were obtained for $[\text{EMIM}]^+[\text{Et}_2\text{PO}_4]^-$, with the 50% activity level corresponding to approximately 5 wt% IL (data not shown). In short, the activity $= 0\%$ in the presence of sufficient IL.

In a separate set of experiments with samples aged more than 1 year in IL-water mixtures, lysozyme activity can be restored to greater than 75% relative to control samples, irrespective of

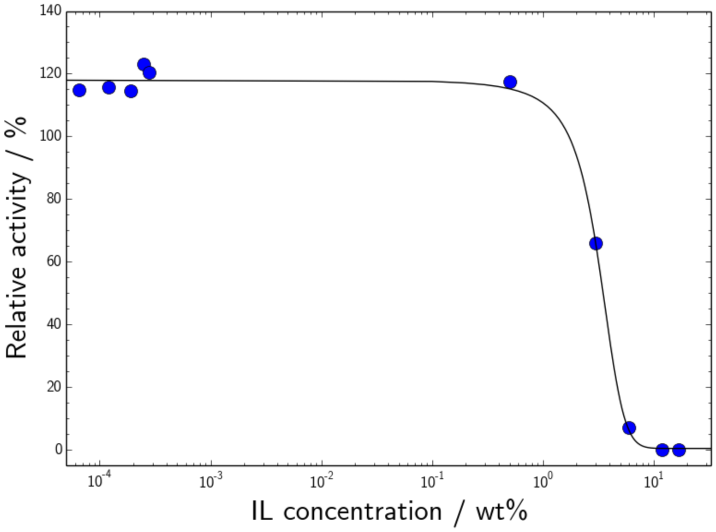


Figure A.6: Normalized lysozyme activity as a function of $[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$ content using a modified activity assay to avoid dilution steps. The curve fit is a Boltzmann-type sigmodial function.

the initial IL content. Therefore the presence of these two imidazolium ILs, even after prolonged incubation, can be easily displaced by water and the lysozyme activity recovered.

A.2.2.3 T_m returns to initial value after rehydration

Table A.3: Lysozyme melting temperature T_m under different conditions. All samples here are aged in their respective solvents for more than 1 year. The dilution factor df $= 100$. ^{a} DSC. ^{b} VP-DSC.

	lysozyme 10 wt%		lysozyme 10/df wt%	
	IL conc., wt%	T_m, °C	IL conc., wt%	T_m, °C
Buffer	0	75.3^a	0	74.6^b
$[\mathrm{EMIM}]^+[\mathrm{EtSO_4}]^-$	50	58.1^a	50/df	75.9^b
$[\mathrm{EMIM}]^+[\mathrm{Et_2PO_4}]^-$	33	59.9^a	33/df	75.3^b

Others [36, 126] and the first half of this section have shown variation of the lysozyme melting temperature T_m in an IL-dependent manner. In the second half of the chapter we find that even when T_m is reduced by the ILs, it remains far above room temperature (Table A.3), indicating that a large majority of the lysozyme protein is still in its native folded state when in presence of these ILs at room temperature. With the addition of a relatively modest amount of water (df = 100), the T_m return to their original values (Table A.3). From the combined experimental and simulation results we suggest that the process observed here is *not refolding*, but rather a rehydration that returns lysozyme to its native solvation state. Part of the reason for this behavior is because of the unusually high stability of lysozyme [129].

Bibliography

[1] Kyoko Ogasahara and Kozo Hamaguchi. Structure of lysozyme: XII. effect of pH on the stability of lysozyme. *J. Biochem.*, 61(2):199–210, 1967.

[2] Raymond P Welch, Hamilton Lee, Michael A Luzuriaga, Olivia R Brohlin, and Jeremiah J Gassensmith. Protein–polymer delivery: Chemistry from the cold chain to the clinic. *Bioconjug. Chem.*, 29(9):2867–2883, 2018.

[3] Alessandra Russo, Paola Bonci, and Paolo Bonci. The effects of different preservation processes on the total protein and growth factor content in a new biological product developed from human amniotic membrane. *Cell Tissue Bank*, 13(2):353–361, 2012.

[4] Jeney Zhang, Eleanor Pritchard, Xiao Hu, Thomas Valentin, Bruce Panilaitis, Fiorenzo G Omenetto, and David L Kaplan. Stabilization of vaccines and antibiotics in silk and eliminating the cold chain. *Proc. Natl. Acad. Sci. U.S.A.*, 109(30):11981–11986, 2012.

[5] Liem Bui-Le, Alex PS Brogan, and Jason P Hallett. Solvent-free liquid avidin as a step toward cold chain elimination. *Biotechnol. Bioeng.*, 118(2):592–600, 2021.

[6] Robert Alcock, Matthew G Cottingham, Christine S Rollier, Julie Furze, Samodh D De Costa, Marian Hanlon, Alexandra J Spencer, Jared D Honeycutt, David H Wyllie, Sarah C Gilbert, et al. Long-term thermostabilization of live poxviral and adenoviral vaccine vectors at supraphysiological temperatures in carbohydrate glass. *Sci. Transl. Med*, 2(19):19ra12–19ra12, 2010.

[7] Maria A Croyle, Blake J Roessler, Beverly L Davidson, John M Hilfinger, and Gordon L Amidon. Factors that influence stability of recombinant adenoviral preparations for human gene therapy. *Pharm Dev Technol*, 3(3):373–383, 1998.

[8] Robert K Evans, Denise K Nawrocki, Lynne A Isopi, Donna M Williams, Danilo R Casimiro, Stephen Chin, Minchun Chen, De-Min Zhu, John W Shiver, and David B Volkin. Development of stable liquid formulations for adenovirus-based vaccines. *J. Pharm. Sci.*, 93(10):2458–2475, 2004.

[9] Chander Kanta Gupta, Jeanne Leszczynski, Rajesh K Gupta, and George R Siber. Stabilization of respiratory syncytial virus (rsv) against thermal inactivation and freeze-thaw cycles for development and control of rsv vaccines and immune globulin. *Vaccine*, 14(15):1417–1420, 1996.

[10] Jason Rexroad, Christopher M Wiethoff, Anthony P Green, Timothy D Kierstead, Miller O Scott, and C Russell Middaugh. Structural stability of adenovirus type 5. *J. Pharm. Sci.*, 92(3):665–678, 2003.

[11] M Stewart, SJ Ward, and J Drew. Use of adenovirus as a model system to illustrate a simple method using standard equipment and inexpensive excipients to remove live virus dependence on the cold-chain. *Vaccine*, 32(24):2931–2938, 2014.

[12] Yun-Chu Chen, Tristan Smith, Robert H Hicks, Aswin Doekhie, Francoise Koumanov, Stephen A Wells, Karen J Edler, Jean van den Elsen, Geoffrey D Holman, Kevin J Marchbank, et al. Thermal stability, storage and release of proteins with tailored fit in silica. *Sci. Rep*, 7(1):1–8, 2017.

[13] Hermann Weingärtner, Chiara Cabrele, and Christian Herrmann. How ionic liquids can help to stabilize native proteins. *Phys. Chem. Chem. Phys.*, 14(2):415–426, 2012.

[14] Hua Zhao. Protein stabilization and enzyme activation in ionic liquids: specific ion effects. *J. Chem. Technol. Biotechnol.*, 91(1):25–50, 2016.

[15] Shashi Kant Shukla and Jyri-Pekka Mikkola. Use of ionic liquids in protein and dna chemistry. *Front. Chem.*, 8:598662, 2020.

[16] Marco Klähn, Geraldine S Lim, and Ping Wu. How ion properties determine the stability of a lipase enzyme in ionic liquids: A molecular dynamics study. *Phys. Chem. Chem. Phys.*, 13(41):18647–18660, 2011.

[17] Diana Constantinescu, Hermann Weingärtner, and Christian Herrmann. Protein denaturation by ionic liquids and the hofmeister series: a case study of aqueous solutions of ribonuclease a. *Angew. Chem. Int. Ed. Engl.*, 46(46):8887–8889, 2007.

[18] Richard J Simpson. Stabilization of proteins for storage. *Cold Spring Harb. Protoc*, 2010(5):pdb–top79, 2010.

[19] Kyoko Fujita, Douglas R MacFarlane, and Maria Forsyth. Protein solubilising and stabilising ionic liquids. *ChemComm*, (38):4804–4806, 2005.

[20] Sheila N Baker, T Mark McCleskey, Siddharth Pandey, and Gary A Baker. Fluorescence studies of protein thermostability in ionic liquids. *ChemComm*, (8):940–941, 2004.

[21] Jason P Mann, Adam McCluskey, and Rob Atkin. Activity and thermal stability of lysozyme in alkylammonium formate ionic liquids—influence of cation modification. *Green Chem.*, 11(6):785–792, 2009.

[22] Khashayar Ghandi. A review of ionic liquids, their limits and applications. *Green sustain. chem.*, 2014, 2014.

[23] Catherine A Summers and Robert A Flowers. Protein renaturation by the liquid organic salt ethylammonium nitrate. *Protein Sci.*, 9(10):2001–2008, 2000.

[24] Anvay Ukidve, Katharina Cu, Morgan Goetz, Pavimol Angsantikul, Alexander Curreri, Eden EL Tanner, Joerg Lahann, and Samir Mitragotri. Ionic-liquid-based safe adjuvants. *Adv Mater*, 32(46):2002990, 2020.

[25] Yoshiro Tahara, Kaho Morita, Rie Wakabayashi, Noriho Kamiya, and Masahiro Goto. Biocompatible ionic liquid enhances transdermal antigen peptide delivery and preventive vaccination effect. *Mol. Pharm.*, 17(10):3845–3856, 2020.

[26] Mohammad Nasir Uddin, Debasish Basak, Robert Hopefl, and Babak Minofar. Potential application of ionic liquids in pharmaceutical dosage forms for small molecule drug and vaccine delivery system. *J. Pharm. Pharm. Sci.*, 23:158–176, 2020.

[27] Takahiro Takekiyo, Kumiko Yamazaki, Erika Yamaguchi, Hiroshi Abe, and Yukihiro Yoshimura. High ionic liquid concentration-induced structural change of protein in aqueous solution: a case study of lysozyme. *The J. Phys. Chem. B* ., 116(36):11092–11097, 2012.

[28] Emmy C Wijaya, Frances Separovic, Calum J Drummond, and Tamar L Greaves. Activity and conformation of lysozyme in molecular solvents, protic ionic liquids (pils) and salt–water systems. *Phys. Chem. Chem. Phys.*, 18(37):25926–25936, 2016.

[29] N Byrne and CA Angell. Protein unfolding, and the “tuning in” of reversible intermediate states, in protic ionic liquid media. *J. Mol. Biol.*, 378(3):707–714, 2008.

[30] Sheila N Baker, Hua Zhao, Siddharth Pandey, William T Heller, Frank V Bright, and Gary A Baker. Fluorescence energy transfer efficiency in labeled yeast cytochrome c: a rapid screen for ion biocompatibility in aqueous ionic liquids. *Phys. Chem. Chem. Phys.*, 13(9):3642–3644, 2011.

[31] Indrani Jha and Pannuru Venkatesu. Unprecedented improvement in the stability of hemoglobin in the presence of promising green solvent 1-allyl-3-methylimidazolium chloride. *ACS Sustain. Chem. Eng.*, 4(2):413–421, 2016.

[32] FJ Deive, A Rodriguez, AB Pereiro, JMM Araújo, MA Longo, MAZ Coelho, JN Canongia Lopes, JMSS Esperança, LPN Rebelo, and IM Marrucho. Ionic liquid-based aqueous biphasic system for lipase extraction. *Green Chem.*, 13(2):390–396, 2011.

[33] Qi Han, Xungai Wang, and Nolene Byrne. Understanding the influence of key ionic liquid properties on the hydrolytic activity of thermomyces lanuginosus lipase. *ChemCatChem*, 8(8):1551–1556, 2016.

[34] Katherine D Weaver, Regina M Vrikkis, Matthew P Van Vorst, Jonathan Trullinger, Ranganathan Vijayaraghavan, David M Foureau, Iain H McKillop, Douglas R MacFarlane, Joanna K Krueger, and Gloria D Elliott. Structure and function of proteins in hydrated choline dihydrogen phosphate ionic liquid. *Phys. Chem. Chem. Phys.*, 14(2):790–801, 2012.

[35] Diana Constatinescu, Christian Herrmann, and Hermann Weingärtner. Patterns of protein unfolding and protein aggregation in ionic liquids. *Phys. Chem. Chem. Phys.*, 12(8):1756–1763, 2010.

[36] João V Rodrigues, Vesna Prosinecki, Isabel Marrucho, Luís Paulo N Rebelo, and Cláudio M Gomes. Protein stability in an ionic liquid milieu: on the use of differential scanning fluorimetry. *Phys. Chem. Chem. Phys.*, 13(30):13614–13616, 2011.

[37] Angelo Miguel Figueiredo, Joao Sardinha, Geoffrey R Moore, and Eurico J Cabrita. Protein destabilisation in ionic liquids: the role of preferential interactions in denaturation. *Phys. Chem. Chem. Phys.*, 15(45):19632–19643, 2013.

[38] Jia Lin Huang, Michael E Noss, Karson M Schmidt, Leigh Murray, and Michelle R Bunagan. The effect of neat ionic liquid on the folding of short peptides. *ChemComm*, 47(28):8007–8009, 2011.

[39] Xiaoxi Yu, Ningning Tian, Fang Huang, Xin Huang, Changlong Liu, Shang Gao, Zhi Yang, and Yining Wu. Evaluating the role of ionic liquids (ils) in the crystallization of lysozyme. *J. Mol. Liq*, 296:112018, 2019.

[40] Vance W Jaeger and Jim Pfaendtner. Structure, dynamics, and activity of xylanase solvated in binary mixtures of ionic liquid and water. *ACS Chem. Biol.*, 8(6):1179–1186, 2013.

[41] Supratim Datta, Bradley Holmes, Joshua I Park, Zhiwei Chen, Dean C Dibble, Masood Hadi, Harvey W Blanch, Blake A Simmons, and Rajat Sapra. Ionic liquid tolerant hyperthermophilic cellulases for biomass pretreatment and hydrolysis. *Green Chem.*, 12(2):338–345, 2010.

[42] Wen-Yong Lou, Min-Hua Zong, Thomas J Smith, Hong Wu, and Ju-Fang Wang. Impact of ionic liquids on papain: an investigation of structure–function relationships. *Green Chem.*, 8(6):509–512, 2006.

[43] Christopher M Dobson. Protein misfolding, evolution and disease. *Trends Biochem. Sci.*, 24(9):329–332, 1999.

[44] John A Hardy and Gerald A Higgins. Alzheimer's disease: the amyloid cascade hypothesis. *Science*, 256(5054):184–185, 1992.

[45] Christiane Reitz. Alzheimer's disease and the amyloid cascade hypothesis: a critical review. *J. Alzheimer's Dis.*, 2012, 2012.

[46] Fred Van Rantwijk and Roger A Sheldon. Biocatalysis in ionic liquids. *Chem.*, 107(6):2757–2785, 2007.

[47] DR Fujita. Macfarlane, and m. forsyth. *Chem. Commun*, pages 4804–4806, 2005.

[48] Hamid Reza Kalhor, Mostafa Kamizi, Jafar Akbari, and Akbar Heydari. Inhibition of amyloid formation by ionic liquids: ionic liquids affecting intermediate oligomers. *Biomacromolecules*, 10(9):2468–2475, 2009.

[49] Nolene Byrne and C Austen Angell. Formation and dissolution of hen egg white lysozyme amyloid fibrils in protic ionic liquids. *ChemComm*, (9):1046–1048, 2009.

[50] Natalie Debeljuh, Colin J Barrow, and Nolene Byrne. The impact of ionic liquids on amyloid fibrilization of aβ16-22: tuning the rate of fibrilization using a reverse hofmeister strategy. *Phys. Chem. Chem. Phys.*, 13(37):16534–16536, 2011.

[51] Heejin Hwang, Hyunghun Choi, Hyun-kyung Kim, T Doohun Kim, et al. Ionic liquids promote amyloid formation from α-synuclein. *Anal. Biochem*, 386(2):293–295, 2009.

[52] Hongxing Lei and Yong Duan. Improved sampling methods for molecular simulation. *Curr. Opin. Struct. Biol*, 17(2):187–191, 2007.

[53] Xiaolin Cheng, Guanglei Cui, Viktor Hornak, and Carlos Simmerling. Modified replica exchange simulation methods for local structure refinement. *The J. Phys. Chem. B* ., 109(16):8220–8230, 2005.

[54] Thur Zar Lwin and Ray Luo. Overcoming entropic barrier with coupled sampling at dual resolutions. *J. Chem. Phys.*, 123(19):194904, 2005.

[55] Massimiliano Bonomi and Michele Parrinello. Enhanced sampling in the well-tempered ensemble. *Phys. Rev. Lett.*, 104(19):190601, 2010.

[56] Pu Liu, Byungchan Kim, Richard A Friesner, and BJ Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.*, 102(39):13749–13754, 2005.

[57] Mookyung Cheon, Iksoo Chang, and Carol K Hall. Extending the prime model for protein aggregation to all 20 amino acids. *Proteins: Struct. Funct. Genet.*, 78(14):2950–2960, 2010.

[58] Aram Davtyan, Nicholas P Schafer, Weihua Zheng, Cecilia Clementi, Peter G Wolynes, and Garegin A Papoian. Awsem-md: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *The J. Phys. Chem. B* ., 116(29):8494–8503, 2012.

[59] Fabio Sterpone, Simone Melchionna, Pierre Tuffery, Samuela Pasquali, Normand Mousseau, Tristan Cragnolini, Yassmine Chebaro, Jean-Francois St-Pierre, Maria Kalimeri, Alessandro Barducci, et al. The opep protein model: from single molecules, amyloid formation, crowding and hydrodynamics to dna/rna systems. *Chem. Soc. Rev*, 43(13):4871–4893, 2014.

[60] Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H De Vries. The martini force field: coarse grained model for biomolecular simulations. *The J. Phys. Chem. B* ., 111(27):7812–7824, 2007.

[61] Semen O Yesylevskyy, Lars V Schäfer, Durba Sengupta, and Siewert J Marrink. Polarizable water model for the coarse-grained martini force field. *PLoS Comput. Biol.*, 6(6):e1000810, 2010.

[62] Yanting Wang and Gregory A Voth. Tail aggregation and domain diffusion in ionic liquids. *The J. Phys. Chem. B .*, 110(37):18601–18608, 2006.

[63] Tianying Yan, Christian J Burnham, Mario G Del Pópolo, and Gregory A Voth. Molecular dynamics simulation of ionic liquids: The effect of electronic polarizability. *The J. Phys. Chem. B .*, 108(32):11877–11881, 2004.

[64] Yanting Wang, WEI Jiang, Tianying Yan, and Gregory A Voth. Understanding ionic liquids through atomistic and coarse-grained molecular dynamics simulations. *Acc. Chem. Res.*, 40(11):1193–1199, 2007.

[65] Yanting Wang, Sergei Izvekov, Tianying Yan, and Gregory A Voth. Multiscale coarse-graining of ionic liquids. *The J. Phys. Chem. B .*, 110(8):3564–3575, 2006.

[66] Shulu Feng and Gregory A Voth. Molecular dynamics simulations of imidazolium-based ionic liquid/water mixtures: Alkyl side chain length and anion effects. *Fluid Ph. Equilibria*, 294(1-2):148–156, 2010.

[67] Jurgis Ruza, Wujie Wang, Daniel Schwalbe-Koda, Simon Axelrod, William H Harris, and Rafael Gómez-Bombarelli. Temperature-transferable coarse-graining of ionic liquids with dual graph convolutional neural networks. *J. Chem. Phys.*, 153(16):164501, 2020.

[68] Luis Itza Vazquez-Salazar, Michele Selle, Alex H De Vries, Siewert J Marrink, and Paulo CT Souza. Martini coarse-grained models of imidazolium-based ionic liquids: from nanostructural organization to liquid–liquid extraction. *Green Chem.*, 22(21):7376–7386, 2020.

[69] Onkar Singh, Pei-Yin Lee, Silvina Matysiak, and Harry Bermudez. Dual mechanism of ionic liquid-induced protein unfolding. *Phys. Chem. Chem. Phys.*, 22(35):19779–19786, 2020.

[70] Pei-Yin Lee, Onkar Singh, Harry Bermudez, and Silvina Matysiak. Recovery of enzyme structure and activity following rehydration from ionic liquid. *Phys. Chem. Chem. Phys.*, 24(17):10365–10372, 2022.

[71] Hua Zhao. Methods for stabilizing and activating enzymes in ionic liquids—a review. *J. Chem. Technol. Biotechnol.*, 85(7):891–907, 2010.

[72] Christian Lange, Ganesh Patil, and Rainer Rudolph. Ionic liquids as refolding additives: N-alkyl and n-(ω-hydroxyalkyl) n-methylimidazolium chlorides. *Protein Sci.*, 14(10):2693–2701, 2005.

[73] Robert Buchfink, Alexander Tischer, Ganesh Patil, Rainer Rudolph, and Christian Lange. Ionic liquids as refolding additives: variation of the anion. *J. Biotechnol.*, 150(1):64–72, 2010.

[74] Fred van Rantwijk and Roger A. Sheldon. Biocatalysis in Ionic Liquids. *Chem.*, 107(6):2757–2785, 2011/10/24 2007.

[75] J R Swartz. Advances in Escherichia coli production of therapeutic proteins. *Curr. Opin. Biotechnol.*, 12(2):195–201, April 2001.

[76] Hermann Weingärtner, Chiara Cabrele, and Christian Herrmann. How ionic liquids can help to stabilize native proteins. *Phys. Chem. Chem. Phys.*, 14(2):415–26, January 2012.

[77] Alexandra Schindl, Matthew L. Hagen, Shafaq Muzammal, Henadira A. D. Gunasekera, and Anna K. Croft. Proteins in ionic liquids: Reactions, applications, and futures. *Front. Chem.*, 7:1–31, 2019.

[78] John Lloyd, Patrick Lydon, Ramzi Ouhichi, and Michel Zaffran. Reducing the loss of vaccines from accidental freezing in the cold chain: the experience of continuous temperature monitoring in Tunisia. *Vaccine*, 33(7):902–907, February 2015.

[79] Takahiro Takekiyo, Kumiko Yamazaki, Erika Yamaguchi, Hiroshi Abe, and Yukihiro Yoshimura. High ionic liquid concentration-induced structural change of protein in aqueous solution: a case study of lysozyme. *J. Phys. Chem. B* ., 116(36):11092–7, September 2012.

[80] Emmy C. Wijaya, Frances Separovic, Calum J. Drummond, and Tamar L. Greaves. Activity and conformation of lysozyme in molecular solvents, protic ionic liquids (pils) and salt-water systems. *Phys. Chem. Chem. Phys.*, 18(37):25926–25936, 2016.

[81] Diana Constantinescu, Hermann Weingärtner, and Christian Herrmann. Protein denaturation by ionic liquids and the hofmeister series: A case study of aqueous solutions of ribonucleasea. *Angew. Chem. Int. Ed. Engl.*, 46(46):8887–8889, 2007.

[82] N Byrne and C A Angell. Protein unfolding, and the ”tuning in” of reversible intermediate states, in protic ionic liquid media. *J. Mol. Biol.*, 378(3):707–714, 2008.

[83] Kyoko Fujita, Douglas R MacFarlane, and Maria Forsyth. Protein solubilising and stabilising ionic liquids. *ChemComm*, (38):4804–6, October 2005.

[84] Sheila N Baker, Hua Zhao, Siddharth Pandey, William T Heller, Frank V Bright, and Gary A Baker. Fluorescence energy transfer efficiency in labeled yeast cytochrome c: a rapid screen for ion biocompatibility in aqueous ionic liquids. *Phys. Chem. Chem. Phys.*, 13(9):3642–4, 2011.

[85] Indrani Jha and Pannuru Venkatesu. Unprecedented improvement in the stability of hemoglobin in the presence of promising green solvent 1-allyl-3-methylimidazolium chloride. *ACS Sustain. Chem. Eng.*, 4(2):413–421, 2015.

[86] F. J. Deive, A. Rodríguez, A. B. Pereiro, J. M. M. Araújo, M. A. Longo, M. A. Z. Coelho, J. N. Canongia Lopes, J. M. S. S. Esperança, L. P. N. Rebelo, and I. M. Marrucho. Ionic liquid-based aqueous biphasic system for lipase extraction. *Green Chem.*, 13(2):390–396, 2011.

[87] Qi Han, Xungai Wang, and Nolene Byrne. Understanding the influence of key ionic liquid properties on the hydrolytic activity of thermomyces lanuginosus lipase. *ChemCatChem*, 8(8):1551–1556, Mar 2016.

[88] Upendra K. Singh, Meena Kumari, Sabab H. Khan, Himadri B. Bohidar, and Rajan Patel. Mechanism and dynamics of long-term stability of cytochrome c conferred by long-chain imidazolium ionic liquids at low concentration. *ACS Sustain. Chem. Eng.*, 6(1):803–815, 2018.

[89] Yun Guo, Bo Zhang, Chao Lu, Xiaoxue Liu, Qing Li, Hua Zhang, and Zhanzhong Wang. Locating the binding domains of lysozyme with ionic liquids in aqueous solution via spectroscopic studies. *Spectrochim. Acta A Mol. Biomol.*, 214:239–245, 2019.

[90] Meena Kumari, Upendra Kumar Singh, Ilyas Beg, Amer M. Alanazi, Azmat Ali Khan, and Rajan Patel. Effect of cations and anions of ionic liquids on the stability and activity of lysozyme: Concentration and temperature effect. *J. Mol. Liq*, 272:253–263, 2018.

[91] Jason P. Mann, Adam McCluskey, and Rob Atkin. Activity and thermal stability of lysozyme in alkylammonium formate ionic liquids-influence of cation modification. *Green Chem.*, 11(6):785, 2009.

[92] Michael Senske, Diana Constantinescu-Aruxandei, Martina Havenith, Christian Herrmann, Hermann Weingärtner, and Simon Ebbinghaus. The temperature dependence of the hofmeister series: thermodynamic fingerprints of cosolute–protein interactions. *Phys. Chem. Chem. Phys.*, 18(43):29698–29708, 2016.

[93] Nolene Byrne, Li-Min Wang, Jean-Philippe Belieres, and C Austen Angell. Reversible folding-unfolding, aggregation protection, and multi-year stabilization, in high concentration protein solutions, using ionic liquids. *ChemComm*, (26):2714–6, July 2007.

[94] Katherine D Weaver, Regina M Vrikkis, Matthew P Van Vorst, Jonathan Trullinger, Ranganathan Vijayaraghavan, David M Foureau, Iain H McKillop, Douglas R MacFarlane, Joanna K Krueger, and Gloria D Elliott. Structure and function of proteins in hydrated choline dihydrogen phosphate ionic liquid. *Phys. Chem. Chem. Phys.*, 14(2):790–801, 2012.

[95] Diana Constatinescu, Christian Herrmann, and Hermann Weingärtner. Patterns of protein unfolding and protein aggregation in ionic liquids. *Phys. Chem. Chem. Phys.*, 12:1756–1763, 2010.

[96] João V Rodrigues, Vesna Prosinecki, Isabel Marrucho, Luís Paulo N Rebelo, and Cláudio M Gomes. Protein stability in an ionic liquid milieu: on the use of differential scanning fluorimetry. *Phys. Chem. Chem. Phys.*, 13(30):13614–6, 2011.

[97] Tony Hunter. Why nature chose phosphate to modify proteins. *Philos Trans R Soc Lond B Biol Sci*, 367(1602):2513–2516, 2012.

[98] Vadim V Mozhaev, Yuri L Khmelnitsky, Mariya V Sergeeva, Alla B Belova, Nataliya L Klyachko, Andrey V Levashov, and Karel Martinek. Catalytic activity and denaturation of enzymes in water/organic cosolvent mixtures: α-chymotrypsin and laccase in mixed water/alcohol, water/glycol and water/formamide solvents. *Eur. J. Biochem.*, 184(3):597–602, 1989.

[99] Miguel Arroyo, Raquel Torres, Isabel de la Mata, M Pilar Castillón, and Carmen Acebal. Interaction of penicillin v acylase with organic solvents:: Catalytic activity modulation on the hydrolysis of penicillin v. *Enzyme Microb. Technol.*, 25(3-5):378–383, 1999.

[100] Dibyendu Kumar Sasmal, Tridib Mondal, Supratik Sen Mojumdar, Aparajita Choudhury, Rajat Banerjee, and Kankan Bhattacharyya. An fcs study of unfolding and refolding of cpm-labeled human serum albumin: role of ionic liquid. *The J. Phys. Chem. B .*, 115(44):13075–13083, 2011.

[101] Upendra Kumar Singh, Meena Kumari, Farooq Ahmad Wani, Mehraj ud din Parray, Juhi Saraswat, Pannuru Venkatesu, and Rajan Patel. Refolding of acid denatured cytochrome c by anionic surface-active ionic liquid: Choice of anion plays key role in refolding of proteins. *Colloids Surf. A Physicochem. Eng*, 582:123872, 2019.

[102] Pankaj Attri and Pannuru Venkatesu. Thermodynamic characterization of the biocompatible ionic liquid effects on protein model compounds and their functional groups. *Phys. Chem. Chem. Phys.*, 13(14):6566–6575, 2011.

[103] Le-Ping Dang, Wen-Zhi Fang, Yan Li, Qian Wang, Hua-Zhi Xiao, and Zhan-Zhong Wang. Ionic liquid-induced structural and activity changes in hen egg white lysozyme. *Appl. Biochem. Biotechnol.*, 169(1):290–300, 2013.

[104] Konstantza Tonova. Long-term preservation of α-amylase activity in highly concentrated aqueous solutions of imidazolium ionic liquid. *Green Process. Synth.*, 7(2):106–113, 2018.

[105] Navin Kumar Mogha, Niketa Yadav, Anamika Sindhu, and Pannuru Venkatesu. Does poly (ionic liquid) modulate the non-covalent interactions of chicken egg white lysozyme? elucidation of biomolecular interactions between biomolecules and macromolecular solvents. *New J Chem*, 43(42):16759–16766, 2019.

[106] Stephen H. Strassburg, Harry Bermudez, and David A. Hoagland. Lysozyme Solubility and Conformation in Two Neat Ionic Liquids and their Mixtures with Water. *Biomacromolecules*, 17(6):2233–2239, 2016.

[107] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4(3):435–447, 2008.

[108] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

[109] José N Canongia Lopes, Agílio AH Pádua, and Karina Shimizu. Molecular force field for ionic liquids iv: Trialkylimidazolium and alkoxycarbonyl-imidazolium cations; alkylsulfonate and alkylsulfate anions. *The J. Phys. Chem. B .*, 112(16):5039–5046, 2008.

[110] Leela S Dodda, Israel Cabeza de Vaca, Julian Tirado-Rives, and William L Jorgensen. Ligpargen web server: an automatic opls-aa parameter generator for organic ligands. *Nucleic Acids Res. Spec. Publ.*, 45(W1):W331–W336, 2017.

[111] David AC Beck, Roger S Armen, and Valerie Daggett. Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. *Biochemistry*, 44(2):609–616, 2005.

[112] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[113] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys*, 52(12):7182–7190, 1981.

[114] Tadeusz Hofman, Andrzej Gołdon, Ashley Nevines, and Trevor M Letcher. Densities, excess volumes, isobaric expansivity, and isothermal compressibility of the (1-ethyl-3-methylimidazolium ethylsulfate+ methanol) system at temperatures (283.15 to 333.15) k and pressures from (0.1 to 35) mpa. *J. Chem. Thermodyn.*, 40(4):580–591, 2008.

[115] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

[116] Charles Tanford. Protein Denaturation. In John T. Edsall, C.B. Anfinsen, M.L. Anson, and Frederic M. Richards, editors, *Adv. Protein Chem.*, volume 23, pages 121–282. Academic Press, 1968.

[117] P L Privalov. Stability of proteins: small globular proteins. *Adv. Protein Chem.*, 33:167–241, 1979.

[118] O. B. Ptitsyn. Molten Globule and Protein Folding. In John T. Edsall, C.B. Anfinsen, Frederic M. Richards, and David S. Eisenberg, editors, *Adv. Protein Chem.*, volume 47, pages 83–229. Academic Press, 1995.

[119] Elliott F. Osserman, Robert E. Canfield, and Sherman Beychok, editors. *Lysozyme*. Academic Press, 1974.

[120] P. Jolles, editor. *Lysozymes: Model Enzymes in Biochemistry and Biology*. Birkäuser, 1996.

[121] Steven L Kazmirski and Valerie Daggett. Non-native interactions in protein folding intermediates: molecular dynamics simulations of hen lysozyme. *J. Mol. Biol.*, 284(3):793–806, 1998.

[122] Valerie Daggett. Molecular dynamics simulations of the protein unfolding/folding reaction. *Acc. Chem. Res.*, 35(6):422–429, 2002.

[123] J. Klein-Seetharaman, M. Oikawa, S.B. Grimshaw, E. Duchardt, T. Ueda, T. Imoto, L.J. Smith, C.M. Dobson, and H. Schwalbe. Long-range interactions within a nonnative protein. *Science*, 295(5560):1719–1722, 2002.

[124] Ruhong Zhou, Maria Eleftheriou, Ajay K Royyuru, and Bruce J Berne. Destruction of long-range interactions by a single mutation in lysozyme. *Proc. Natl. Acad. Sci. U.S.A.*, 104(14):5824–5829, 2007.

[125] Maria Eleftheriou, Robert S Germain, Ajay K Royyuru, and Ruhong Zhou. Thermal denaturing of mutant lysozyme with both the oplsaa and the charmm force fields. *J. Am. Chem. Soc.*, 128(41):13388–13395, 2006.

[126] Robert Buchfink, Alexander Tischer, Ganesh Patil, Rainer Rudolph, and Christian Lange. Ionic liquids as refolding additives: variation of the anion. *J. Biotechnol.*, 150(1):64–72, October 2010.

[127] Yann Desfourgères, Thomas Croguennec, Valérie Lechevalier, Saïd Bouhallab, and Françoise Françoise Nau. Charge and size drive spontaneous self-assembly of oppositely charged globular proteins into microspheres. *The J. Phys. Chem. B .*, 114(12):4138–4144, 2010.

[128] Hui Yan, Anup Nykanen, Janne Ruokolainen, David Farrar, Julie E. Gough, Alberto Saiani, and Aline F. Miller. Thermo-reversible protein fibrillar hydrogels as cell scaffolds. *Faraday Discuss.*, 139:71, 2008.

[129] Kozo Hamaguchi. Studies on the denaturation of lysozyme: I. kinetics of heat denaturation. *J. Biochem.*, 44(11):695–706, 1957.

[130] Samantha R Summers, Kayla G Sprenger, Jim Pfaendtner, Jan Marchant, Michael F Summers, and Joel L Kaar. Mechanism of competitive inhibition and destabilization of acidothermus cellulolyticus endoglucanase 1 by ionic liquids. *The J. Phys. Chem. B .*, 121(48):10793–10803, 2017.

[131] Abhilash Sahoo, Pei-Yin Lee, and Silvina Matysiak. Transferable and polarizable coarse grained model for proteins prompt. *J. Chem. Theory Comput.*, 2022.

[132] A D McLachlan. Protein Structure and Function. *Annu. Rev. Phys. Chem.*, 23(1):165–192, October 1972.

[133] Joseph A. Marsh and Sarah A. Teichmann. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.*, 84(1):551–575, June 2015.

[134] Freddie R Salsbury Jr. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr Opin Pharmacol*, 10(6):738–744, December 2010.

[135] M. Karplus and J. Kuriyan. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.*, 102(19):6679–6685, May 2005.

[136] Ron O. Dror, Robert M. Dirks, J.P. Grossman, Huafeng Xu, and David E. Shaw. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.*, 41(1):429–452, June 2012.

[137] Paul Maragakis, Kresten Lindorff-Larsen, Michael P. Eastwood, Ron O. Dror, John L. Klepeis, Isaiah T. Arkin, Morten. Jensen, Huafeng Xu, Nikola Trbovic, Richard A. Friesner, Arthur G. Palmer, and David E. Shaw. Microsecond Molecular Dynamics Simulation Shows Effect of Slow Loop Dynamics on Backbone Amide Order Parameters of Proteins. *J. Phys. Chem. B*, 112(19):6155–6158, May 2008.

[138] Peter L. Freddolino, Feng Liu, Martin Gruebele, and Klaus Schulten. Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. *Biophys. J.*, 94(10):L75–L77, May 2008.

[139] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw. How Fast-Folding Proteins Fold. *Science*, 334(6055):517–520, October 2011.

[140] Kresten Lindorff-Larsen, Paul Maragakis, Stefano Piana, and David E. Shaw. Picosecond to Millisecond Structural Dynamics in Human Ubiquitin. *J. Phys. Chem. B*, 120(33):8313–8320, August 2016.

[141] Giovanni Bussi and Alessandro Laio. Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.*, 2(4):200–212, April 2020.

[142] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.*, 99(20):12562–12566, October 2002.

[143] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 1(5):826–843, September 2011.

[144] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1-2):141–151, November 1999.

[145] Sebastian Kmiecik, Dominik Gront, Michal Kolinski, Lukasz Wieteska, Aleksandra Elzbieta Dawid, and Andrzej Kolinski. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.*, 116(14):7898–7936, July 2016.

[146] José Nelson Onuchic, Zaida Luthey-Schulten, and Peter G. Wolynes. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.*, 48(1):545–600, October 1997.

[147] Ronald Hills and Charles Brooks. Insights from Coarse-Grained Gō Models for Protein Folding and Dynamics. *Int. J. Mol. Sci.*, 10(3):889–905, March 2009.

[148] Cecilia Clementi. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.*, 18(1):10–15, February 2008.

[149] Ram Samudrala and John Moult. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction 1 1Edited by F. Cohen. *J. Mol. Biol.*, 275(5):895–916, February 1998.

[150] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B*, 111(27):7812–7824, July 2007.

[151] Parimal Kar, Srinivasa Murthy Gopal, Yi-Ming Cheng, Alexander Predeus, and Michael Feig. PRIMO: A Transferable Coarse-Grained Force Field for Proteins. *J. Chem. Theory Comput.*, 9(8):3769–3788, August 2013.

[152] Aram Davtyan, Nicholas P. Schafer, Weihua Zheng, Cecilia Clementi, Peter G. Wolynes, and Garegin A. Papoian. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B*, 116(29):8494–8503, July 2012.

[153] Yassmine Chebaro, Samuela Pasquali, and Philippe Derreumaux. The Coarse-Grained OPEP Force Field for Non-Amyloid and Amyloid Proteins. *J. Phys. Chem. B*, 116(30):8741–8752, August 2012.

[154] Karolina Zieba, Magdalena Ślusarz, Rafał Slusarz, Adam Liwo, Cezary Czaplewski, and Adam K. Sieradzan. Extension of the UNRES Coarse-Grained Force Field to Membrane Proteins in the Lipid Bilayer. *J. Phys. Chem. B*, 123(37):7829–7839, September 2019.

[155] Henry Chan, Mathew J. Cherukara, Badri Narayanan, Troy D. Loeffler, Chris Benmore, Stephen K. Gray, and Subramanian K. R. S. Sankaranarayanan. Machine learning coarse grained models for water. *Nat. Commun.*, 10(1):379, December 2019.

[156] Jiang Wang, Simon Olsson, Christoph Wehmeyer, Adrià Pérez, Nicholas E. Charron, Gianni de Fabritiis, Frank Noé, and Cecilia Clementi. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.*, 5(5):755–767, May 2019.

[157] Cecilia Clementi, Hugh Nymeyer, and José Nelson Onuchic. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins. *J. Mol. Biol.*, 298(5):937–953, May 2000.

[158] Scott Brown, Nicolas J. Fawzi, and Teresa Head-Gordon. Coarse-grained sequences for protein folding and design. *Proc. Natl. Acad. Sci. U.S.A.*, 100(19):10712–10717, September 2003.

[159] Alex Morriss-Andrews, Frank L. H. Brown, and Joan-Emma Shea. A Coarse-Grained Model for Peptide Aggregation on a Membrane Surface. *J. Phys. Chem. B*, 118(28):8420–8432, July 2014.

[160] Payel Das, Silvina Matysiak, and Cecilia Clementi. Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes. *Proc. Natl. Acad. Sci. U.S.A.*, 102(29):10141–10146, July 2005.

[161] Hao Wu, Yamini Dalal, and Garegin A. Papoian. Binding Dynamics of Disordered Linker Histone H1 with a Nucleosomal Particle. *J. Mol. Biol.*, 433(6):166881, March 2021.

[162] David S. Yang, Arash Saeedi, Aram Davtyan, Mohsen Fathi, Michael B. Sherman, Mohammad S. Safari, Alena Klindziuk, Michelle C. Barton, Navin Varadarajan, Anatoly B. Kolomeisky, and Peter G. Vekilov. Mesoscopic protein-rich clusters host the nucleation of mutant p53 amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.*, 118(10):e2015618118, March 2021.

[163] Scott P. O. Danielsen, James McCarty, Joan-Emma Shea, Kris T. Delaney, and Glenn H. Fredrickson. Molecular design of self-coacervation phenomena in block polyampholytes. *Proc. Natl. Acad. Sci. U.S.A.*, 116(17):8224–8232, April 2019.

[164] Bogdan Barz and Brigita Urbanc. Minimal Model of Self-Assembly: Emergence of Diversity and Complexity. *J. Phys. Chem. B*, 118(14):3761–3770, April 2014.

[165] Stefan Auer, Filip Meersman, Christopher M. Dobson, and Michele Vendruscolo. A Generic Mechanism of Emergence of Amyloid Protofilaments from Disordered Oligomeric Aggregates. *PLoS Comput. Biol.*, 4(11):e1000222, November 2008.

[166] K. Yue, K. M. Fiebig, P. D. Thomas, H. S. Chan, E. I. Shakhnovich, and K. A. Dill. A test of lattice protein folding algorithms. *Proc. Natl. Acad. Sci. U.S.A.*, 92(1):325–329, January 1995.

[167] Jian Zhou, Ian F. Thorpe, Sergey Izvekov, and Gregory A. Voth. Coarse-Grained Peptide Modeling Using a Systematic Multiscale Approach. *Biophys. J.*, 92(12):4289–4303, June 2007.

[168] S. Y. Mashayak, Mara N. Jochum, Konstantin Koschke, N. R. Aluru, Victor Rühle, and Christoph Junghans. Relative Entropy and Optimization-Driven Coarse-Graining Methods in VOTCA. *PLoS One*, 10(7):e0131754, July 2015.

[169] Sergei Izvekov and Gregory A. Voth. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B*, 109(7):2469–2473, February 2005.

[170] S.M. Vaiana, M. Manno, A. Emanuele, M.B. Palma-Vittorelli, and M.U. Palma. The role of solvent in protein folding and in aggregation. *J. Biol. Phys.*, 27(2/3):133–145, 2001.

[171] Hiraku Oshima and Masahiro Kinoshita. Essential roles of protein-solvent many-body correlation in solvent-entropy effect on protein folding and denaturation: Comparison between hard-sphere solvent and water. *J. Chem. Phys.*, 142(14):145103, April 2015.

[172] C. Nick Pace, Saul Treviño, Erode Prabhakaran, and J. Martin Scholtz. Protein structure, stability and solubility in water and other solvents. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 359(1448):1225–1235, August 2004.

[173] D. K. Klimov and D. Thirumalai. Mechanisms and kinetics of beta -hairpin formation. *Proc. Natl. Acad. Sci. U.S.A.*, 97(6):2544–2549, March 2000.

[174] D. K. Klimov and D. Thirumalai. Linking rates of folding in lattice models of proteins with underlying thermodynamic characteristics. *J. Chem. Phys.*, 109(10):4119–4125, September 1998.

[175] Yuzo Ueda, Hiroshi Taketomi, and Nobuhiro G? Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. *Biopolymers*, 17(6):1531–1548, June 1978.

[176] Paulo C. T. Souza, Riccardo Alessandri, Jonathan Barnoud, Sebastian Thallmair, Ignacio Faustino, Fabian Grünewald, Ilias Patmanidis, Haleh Abdizadeh, Bart M. H. Bruininks, Tsjerk A. Wassenaar, Peter C. Kroon, Josef Melcr, Vincent Nieto, Valentina Corradi, Hanif M. Khan, Jan Domański, Matti Javanainen, Hector Martinez-Seara, Nathalie Reuter, Robert B. Best, Ilpo Vattulainen, Luca Monticelli, Xavier Periole, D. Peter Tieleman, Alex H. de Vries, and Siewert J. Marrink. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods*, 18(4):382–388, April 2021.

[177] Paulo C. T. Souza, Sebastian Thallmair, Paolo Conflitti, Carlos Ramírez-Palacios, Riccardo Alessandri, Stefano Raniolo, Vittorio Limongelli, and Siewert J. Marrink. Protein–ligand binding with the coarse-grained Martini model. *Nat Commun*, 11(1):3714, December 2020.

[178] Paulo C. T. Souza, Vittorio Limongelli, Sangwook Wu, Siewert J. Marrink, and Luca Monticelli. Perspectives on High-Throughput Ligand/Protein Docking With Martini MD Simulations. *Front. Mol. Biosci.*, 8:657222, March 2021.

[179] Cristian Gobbo, Isabelle Beurroies, David de Ridder, Rienk Eelkema, Siewert J. Marrink, Steven De Feyter, Jan H. van Esch, and Alex H. de Vries. MARTINI Model for Physisorption of Organic Molecules on Graphite. *J. Phys. Chem. C*, 117(30):15623–15631, August 2013.

[180] C. Arnarez, S. J. Marrink, and X. Periole. Molecular mechanism of cardiolipin-mediated assembly of respiratory chain supercomplexes. *Chem. Sci.*, 7(7):4435–4443, 2016.

[181] Fude Sun, Carsten F.E. Schroer, Lida Xu, Huiwei Yin, Siewert J. Marrink, and Shi-Zhong Luo. Molecular Dynamics of the Association of L-Selectin and FERM Regulated by PIP2. *Biophys. J.*, 114(8):1858–1868, April 2018.

[182] Tsjerk A. Wassenaar, Kristyna Pluhackova, Anastassiia Moussatova, Durba Sengupta, Siewert J. Marrink, D. Peter Tieleman, and Rainer A. Böckmann. High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. *J. Chem. Theory Comput.*, 11(5):2278–2291, May 2015.

[183] John Karanicolas and Charles L. Brooks. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.*, 11(10):2351–2361, April 2009.

[184] Adolfo B. Poma, Marek Cieplak, and Panagiotis E. Theodorakis. Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. *J. Chem. Theory Comput.*, 13(3):1366–1374, March 2017.

[185] Md. Iqbal Mahmood, Adolfo B. Poma, and Kei-ichi Okazaki. Optimizing Gō-MARTINI Coarse-Grained Model for F-BAR Protein on Lipid Membrane. *Front. Mol. Biosci.*, 8:619381, February 2021.

[186] Sai J. Ganesan and S. Matysiak. Role of Backbone Dipole Interactions in the Formation of Secondary and Supersecondary Structures of Proteins. *J. Chem. Theory Comput.*, 10(6):2569–2576, June 2014.

[187] Sai J. Ganesan, Hongcheng Xu, and Silvina Matysiak. Effect of lipid head group interactions on membrane properties and membrane-induced cationic β-hairpin folding. *Phys. Chem. Chem. Phys.*, 18(27):17836–17850, 2016.

[188] Sai J. Ganesan and Silvina Matysiak. Interplay between the hydrophobic effect and dipole interactions in peptide aggregation at interfaces. *Phys. Chem. Chem. Phys.*, 18(4):2449–2458, 2016.

[189] Abhilash Sahoo, Hongcheng Xu, and Silvina Matysiak. Pathways of amyloid-beta absorption and aggregation in a membranous environment. *Phys. Chem. Chem. Phys.*, 21(16):8559–8568, 2019.

[190] Abhilash Sahoo and Silvina Matysiak. Effects of applied surface-tension on membrane-assisted aβ aggregation. *Phys. Chem. Chem. Phys.*, 23(36):20627–20633, 2021.

[191] Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert-Jan Marrink. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.*, 4(5):819–834, May 2008.

[192] Isaiah T Arkin. Structural aspects of oligomerization taking place between the transmembrane α-helices of bitopic membrane proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1565(2):347–363, 2002.

[193] Jean-Luc Popot and Donald M Engelman. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry*, 29(17):4031–4037, 1990.

[194] Kevin R MacKenzie, James H Prestegard, and Donald M Engelman. A transmembrane helix dimer: structure and implications. *Science*, 276(5309):131–133, 1997.

[195] Steven O Smith, David Song, Srinivasan Shekar, Michel Groesbeek, Martine Ziliox, and Saburo Aimoto. Structure of the transmembrane dimer interface of glycophorin a in membrane bilayers. *Biochemistry*, 40(22):6553–6558, 2001.

[196] William P Russ and Donald M Engelman. The gxxxg motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.*, 296(3):911–919, 2000.

[197] Bettina Brosig and Dieter Langosch. The dimerization motif of the glycophorin a transmembrane segment in membranes: importance of glycine residues. *Protein Sci.*, 7(4):1052–1056, 1998.

[198] Mark A Lemmon, John M Flanagan, Herbert R Treutlein, Jian Zhang, and Donald M Engelman. Sequence specificity in the dimerization of transmembrane. alpha.-helixes. *Biochemistry*, 31(51):12719–12725, 1992.

[199] Karen G Fleming and Donald M Engelman. Specificity in transmembrane helix–helix interactions can define a hierarchy of stability for sequence variants. *Proc. Natl. Acad. Sci. U.S.A.*, 98(25):14340–14344, 2001.

[200] Lillian E Fisher, Donald M Engelman, and James N Sturgis. Detergents modulate dimerization, but not helicity, of the glycophorin a transmembrane domain. *J. Mol. Biol.*, 293(3):639–651, 1999.

[201] Emi Psachoulia, Philip W Fowler, Peter J Bond, and Mark SP Sansom. Helix- helix interactions in membrane proteins: Coarse-grained simulations of glycophorin a helix dimerization. *Biochemistry*, 47(40):10503–10512, 2008.

[202] Tsjerk A Wassenaar, Kristyna Pluhackova, Anastassiia Moussatova, Durba Sengupta, Siewert J Marrink, D Peter Tieleman, and Rainer A Böckmann. High-throughput simulations of dimer and trimer assembly of membrane proteins. the daft approach. *J. Chem. Theory Comput.*, 11(5):2278–2291, 2015.

[203] Thomas Lemmin, Mitko Dimitrov, Patrick C Fraering, and Matteo Dal Peraro. Perturbations of the straight transmembrane α-helical structure of the amyloid precursor protein affect its processing by γ-secretase. *J. Biol. Chem.*, 289(10):6763–6774, 2014.

[204] Alexander Götz, Nadine Mylonas, Philipp Högel, Mara Silber, Hannes Heinel, Simon Menig, Alexander Vogel, H Freyrer, Daniel Huster, Burkhard Luy, et al. Stabilization/destabilization of the app transmembrane domain by mutations in the diglycine hinge alter helical structure and dynamics, and impair cleavage by-secretase. *bioRxiv*, page 10, 2018.

[205] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L de Groot, Helmut Grubmüller, and Alexander D MacKerell. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*, 14(1):71–73, January 2017.

[206] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, July 1983.

[207] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.*, 29(11):1859–1865, August 2008.

[208] Semen O. Yesylevskyy, Lars V. Schäfer, Durba Sengupta, and Siewert J. Marrink. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput Biol*, 6(6):e1000810, June 2010.

[209] J. P. Gallivan and D. A. Dougherty. Cation-pi interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.*, 96(17):9459–9464, August 1999.

[210] Neal J. Zondlo. Aromatic–Proline Interactions: Electronically Tunable CH/π Interactions. *Acc. Chem. Res.*, 46(4):1039–1049, April 2013.

[211] Emily G Baker, Christopher Williams, Kieran L Hudson, Gail J Bartlett, Jack W Heal, Kathryn L Porter Goff, Richard B Sessions, Matthew P Crump, and Derek N Woolfson. Engineering protein stability with atomic precision in a monomeric miniprotein. *Nat Chem Biol*, 13(7):764–770, July 2017.

[212] Motohiro Nishio, Yoji Umezawa, Jacques Fantini, Manfred S. Weiss, and Pinak Chakrabarti. Ch–π hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys.*, 16(25):12648–12683, 2014.

[213] Jing Huang, Pedro E. M. Lopes, Benoît Roux, and Alexander D. MacKerell. Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model. *J. Phys. Chem. Lett.*, 5(18):3144–3150, September 2014.

[214] Haibo Yu, Troy W. Whitfield, Edward Harder, Guillaume Lamoureux, Igor Vorobyov, Victor M. Anisimov, Alexander D. MacKerell, and Benoît Roux. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J. Chem. Theory Comput.*, 6(3):774–786, March 2010.

[215] Eng-Hui Yap, Nicolas Lux Fawzi, and Teresa Head-Gordon. A coarse-grained α-carbon protein model with anisotropic hydrogen-bonding. *Proteins*, 70(3):626–638, February 2008.

[216] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, September 2015.

[217] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, June 1993.

[218] Harald A. Posch, William G. Hoover, and Franz J. Vesely. Canonical dynamics of the Nos$\tilde{e}$ oscillator: Stability, order, and chaos. *Phys. Rev. A*, 33(6):4253–4265, June 1986.

[219] Lingle Wang, Richard A Friesner, and BJ Berne. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (rest2). *J. Phys. Chem. B*, 115(30):9431–9438, 2011.

[220] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct. Funct. Genet.*, 65(3):712–725, 2006.

[221] Michael R Shirts and John D Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, 129(12):124105, 2008.

[222] Dietmar Paschek, Sascha Hempel, and Angel E García. Computing the stability diagram of the trp-cage miniprotein. *Proc. Natl. Acad. Sci. U.S.A.*, 105(46):17754–17759, 2008.

[223] Ryan Day, Dietmar Paschek, and Angel E Garcia. Microsecond simulations of the folding/unfolding thermodynamics of the trp-cage miniprotein. *Proteins: Struct. Funct. Genet.*, 78(8):1889–1899, 2010.

[224] Andrea G Cochran, Nicholas J Skelton, and Melissa A Starovasnik. Tryptophan zippers: Stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci. U.S.A.*, 98(10):5578–5583, 2001.

[225] Deguo Du, Yongjin Zhu, Cheng-Yen Huang, and Feng Gai. Understanding the key factors that control the rate of β-hairpin folding. *Proc. Natl. Acad. Sci. U.S.A.*, 101(45):15915–15920, 2004.

[226] Aimee Byrne, D Victoria Williams, Bipasha Barua, Stephen J Hagen, Brandon L Kier, and Niels H Andersen. Folding dynamics and pathways of the trp-cage miniproteins. *Biochemistry*, 53(38):6011–6021, 2014.

[227] Dietmar Paschek, Hugh Nymeyer, and Angel E García. Replica exchange simulation of reversible folding/unfolding of the trp-cage miniprotein in explicit solvent: On the structure and possible role of internal water. *J. Mol. Biol.*, 157(3):524–533, 2007.

[228] Lucas N. R. Wafer, Werner W. Streicher, and George I. Makhatadze. Thermodynamics of the Trp-cage Miniprotein Unfolding in Urea. *Proteins*, 78(6):1376–1381, May 2010.

[229] Beatrice N. Markiewicz, Lijiang Yang, Robert M. Culik, Yi Qin Gao, and Feng Gai. How Quickly Can a β-Hairpin Fold from Its Transition State? *J. Phys. Chem. B*, 118(12):3317–3325, March 2014.

[230] Andreas Reiner, Peter Henklein, and Thomas Kiefhaber. An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc. Natl. Acad. Sci. U.S.A.*, 107(11):4955–4960, 2010.

[231] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12):2577–2637, 1983.

[232] David C. Latshaw, Mookyung Cheon, and Carol K. Hall. Effects of Macromolecular Crowding on Amyloid β (16–22) Aggregation Using Coarse-Grained Simulations. *J. Phys. Chem. B*, 118(47):13513–13526, November 2014.

[233] Yiming Wang, Samuel J. Bunce, Sheena E. Radford, Andrew J. Wilson, Stefan Auer, and Carol K. Hall. Thermodynamic phase diagram of amyloid-β (16–22) peptide. *Proc. Natl. Acad. Sci. U.S.A.*, 116(6):2091–2096, February 2019.

[234] Dongdong Lin, Yin Luo, Shan Wu, Qianqian Ma, Guanghong Wei, and Xinju Yang. Investigation of the Aggregation Process of Amyloid-β-(16-22) Peptides and the Dissolution of Intermediate Aggregates. *Langmuir*, 30(11):3170–3175, March 2014.

[235] Suman Samantray, Feng Yin, Batuhan Kav, and Birgit Strodel. Different Force Fields Give Rise to Different Amyloid Aggregation Pathways in Molecular Dynamics Simulations. *J. Chem. Inf. Model.*, 60(12):6462–6475, December 2020.

[236] John J Balbach, Yoshitaka Ishii, Oleg N Antzutkin, Richard D Leapman, Nancy W Rizzo, Fred Dyda, Jennifer Reed, and Robert Tycko. Amyloid fibril formation by aβ16-22, a seven-residue fragment of the alzheimer's β-amyloid peptide, and structural characterization by solid state nmr. *Biochemistry*, 39(45):13748–13759, 2000.

[237] John J. Balbach, Yoshitaka Ishii, Oleg N. Antzutkin, Richard D. Leapman, Nancy W. Rizzo, Fred Dyda, Jennifer Reed, and Robert Tycko. Amyloid Fibril Formation by Aβ $_{16\text{-}22}$, a Seven-Residue Fragment of the Alzheimer's β-Amyloid Peptide, and Structural Characterization by Solid State NMR. *Biochemistry*, 39(45):13748–13759, November 2000.

[238] Phuong H Nguyen, Mai Suan Li, and Philippe Derreumaux. Effects of all-atom force fields on amyloid oligomerization: Replica exchange molecular dynamics simulations of the aβ 16–22 dimer and trimer. *Phys. Chem. Chem. Phys.*, 13(20):9778–9788, 2011.

[239] Clément Arnarez, Jaakko J. Uusitalo, Marcelo F. Masman, Helgi I. Ingólfsson, Djurre H. de Jong, Manuel N. Melo, Xavier Periole, Alex H. de Vries, and Siewert J. Marrink. Dry Martini, a Coarse-Grained Force Field for Lipid Membrane Simulations with Implicit Solvent. *J. Chem. Theory Comput.*, 11(1):260–275, January 2015.

[240] Sai J. Ganesan, Hongcheng Xu, and Silvina Matysiak. Influence of Monovalent Cation Size on Nanodomain Formation in Anionic–Zwitterionic Mixed Bilayers. *J. Phys. Chem. B*, 121(4):787–799, February 2017.

[241] Abhilash Sahoo and Silvina Matysiak. Microscopic Picture of Calcium-Assisted Lipid Demixing and Membrane Remodeling Using Multiscale Simulations. *J. Phys. Chem. B*, 124(34):7327–7335, August 2020.

[242] Madhusoodanan Mottamal, Jinming Zhang, and Themis Lazaridis. Energetics of the native and non-native states of the glycophorin transmembrane helix dimer. *Proteins: Struct. Funct. Genet.*, 62(4):996–1009, 2006.

[243] Kevin R MacKenzie and Donald M Engelman. Structure-based prediction of the stability of transmembrane helix–helix interactions: the sequence dependence of glycophorin a dimerization. *Proc. Natl. Acad. Sci. U.S.A.*, 95(7):3583–3590, 1998.

[244] Karen G Fleming, Anne L Ackerman, and Donald M Engelman. The effect of point mutations on the free energy of transmembrane α-helix dimerization. *J. Mol. Biol.*, 272(2):266–275, 1997.

[245] Aimee M Morris, Murielle A Watzky, and Richard G Finke. Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1794(3):375–397, 2009.

[246] Wei Wang, Sandeep Nema, and Dirk Teagarden. Protein aggregation—pathways and influencing factors. *Int. J. Pharm.*, 390(2):89–99, 2010.

[247] Wei Wang. Protein aggregation and its inhibition in biopharmaceutics. *Int. J. Pharm.*, 289(1-2):1–30, 2005.

[248] Mary EM Cromwell, Eric Hilario, and Fred Jacobson. Protein aggregation and bioprocessing. *AAPS J.*, 8(3):E572–E579, 2006.

[249] Marielle Aulikki Wälti, Francesco Ravotti, Hiromi Arai, Charles G Glabe, Joseph S Wall, Anja Böckmann, Peter Güntert, Beat H Meier, and Roland Riek. Atomic-resolution structure of a disease-relevant aβ (1–42) amyloid fibril. *Proc. Natl. Acad. Sci. U.S.A.*, 113(34):E4976–E4984, 2016.

[250] Mike LaFontaine and Medhane Cumbay. Focus: the aging brain: amyloid-beta alzheimer targets—protein processing, lipid rafts, and amyloid-beta pores. *Yale J. Biol. Med.*, 89(1):5, 2016.

[251] Natalie Debeljuh, Swapna Varghese, Colin J Barrow, and Nolene Byrne. Role of cation in enhancing the conversion of the alzheimer's peptide into amyloid fibrils using protic ionic liquids. *Aust. J. Chem.*, 65(11):1502–1506, 2012.

[252] Sathish Dasari and Bhabani S Mallik. Conformational dynamics of amyloid-β (16–22) peptide in aqueous ionic liquids. *RSC Advances*, 10(55):33248–33260, 2020.

[253] Anil K Mehta, Kun Lu, W Seth Childers, Yan Liang, Steven N Dublin, Jijun Dong, James P Snyder, Sai Venkatesh Pingali, Pappannan Thiyagarajan, and David G Lynn. Facial symmetry in protein self-assembly. *J. Am. Chem. Soc.*, 130(30):9829–9835, 2008.

[254] Yan Lu, Philippe Derreumaux, Zhi Guo, Normand Mousseau, and Guanghong Wei. Thermodynamics and dynamics of amyloid peptide oligomerization are sequence dependent. *Proteins: Struct. Funct. Genet.*, 75(4):954–963, 2009.

[255] Andrew J Doig and Robert L Baldwin. N-and c-capping preferences for all 20 amino acids in α-helical peptides. *Protein Sci.*, 4(7):1325–1336, 1995.

[256] Sai J Ganesan and S Matysiak. Role of backbone dipole interactions in the formation of secondary and supersecondary structures of proteins. *J. Chem. Theory Comput.*, 10(6):2569–2576, 2014.

[257] Guillaume Chevrot, Eudes Eterno Fileti, and Vitaly V Chaban. Protein remains stable at unusually high temperatures when solvated in aqueous mixtures of amino acid based ionic liquids. *J. Mol. Model*, 22(11):1–9, 2016.

[258] Betul Uralcan, Sang Beom Kim, Chester E Markwalter, Robert K Prud'homme, and Pablo G Debenedetti. A computational study of the ionic liquid-induced destabilization of the miniprotein trp-cage. *The J. Phys. Chem. B* ., 122(21):5707–5715, 2018.

[259] Diddo Diddens, Volker Lesch, Andreas Heuer, and Jens Smiatek. Aqueous ionic liquids and their influence on peptide conformations: denaturation and dehydration mechanisms. *Phys. Chem. Chem. Phys.*, 19(31):20430–20440, 2017.

[260] Volker Lesch, Andreas Heuer, Vasileios A Tatsis, Christian Holm, and Jens Smiatek. Peptides in the presence of aqueous ionic liquids: tunable co-solutes as denaturants or protectants? *Phys. Chem. Chem. Phys.*, 17(39):26049–26053, 2015.

[261] Mor Goldfeder and Ayelet Fishman. Modulating enzyme activity using ionic liquids or surfactants. *Appl. Microbiol. Biotechnol.*, 98(2):545–554, 2014.

[262] Debasis Saha and Arnab Mukherjee. Effect of water and ionic liquids on biomolecules. *Biophys Rev*, 10(3):795–808, 2018.

[263] Kavya Bhakuni, Niketa Yadav, and Pannuru Venkatesu. A novel amalgamation of deep eutectic solvents and crowders as biocompatible solvent media for enhanced structural and thermal stability of bovine serum albumin. *Phys. Chem. Chem. Phys.*, 22(42):24410–24422, 2020.

[264] Jennifer A Kist, Hua Zhao, Katie R Mitchell-Koch, and Gary A Baker. The study and application of biomolecules in deep eutectic solvents. *J Mater Chem B*, 9(3):536–566, 2021.

[265] William M. Jackson and John F. Brandts. Thermodynamics of protein denaturation. calorimetric study of the reversible denaturation of chymotrypsinogen and conclusions regarding the accuracy of the two-state approximation. *Biochemistry*, 9(11):2294–2301, 1970.

[266] Alexei V Finkelstein and O. B. Ptitsyn. *Protein physics: a course of lectures*. Academic Press, 2016.

[267] Kirk C. Aune and Charles Tanford. Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. I. dependence on pH at 25 degrees. *Biochemistry*, 8(11):4579–4585, 1969.

[268] Zhen Xia, Payel Das, Eugene I Shakhnovich, and Ruhong Zhou. Collapse of unfolded proteins in a mixture of denaturants. *J. Am. Chem. Soc.*, 134(44):18266–18274, 2012.

[269] Catherine A. Summers and Robert A. Flowers. Protein renaturation by the liquid organic salt ethylammonium nitrate. *Protein Sci.*, 9(10):2001–2008, 2000.

[270] Julian M. Sturtevant. Biochemical applications of differential scanning calorimetry. *Annu. Rev. Phys. Chem.*, 38(1):463–488, 1987.

[271] P L Privalov and N N Khechinashvili. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *J. Mol. Biol.*, 86(3):665–84, July 1974.

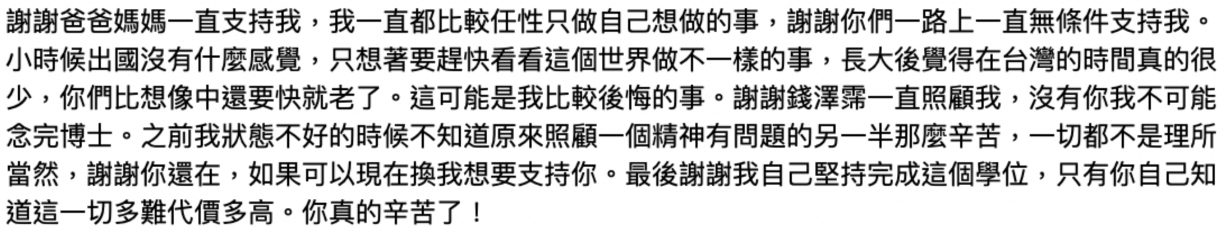
[272] P L Privalov. Intermediate states in protein folding. *J. Mol. Biol.*, 258(5):707–25, May 1996.

[273] Joseph R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Springer, 3rd edition, 2006.

[274] FUMIO Tanaka, LESLIE S Forster, PRANAB K Pal, and John A Rupley. The circular dichroism of lysozyme. *J. Biol. Chem.*, 250(17):6977–6982, 1975.

[275] Pankaj Attri, Pannuru Venkatesu, and Anil Kumar. Activity and stability of alpha-chymotrypsin in biocompatible ionic liquids: enzyme refolding by triethyl ammonium acetate. *Phys. Chem. Chem. Phys.*, 13(7):2788–96, 2011.

[276] David Shugar. The measurement of lysozyme activity and the ultra-violet inactivation of lysozyme. *Biochimica et Biophysica Acta*, 8:302–309, 1952.



謝謝爸爸媽媽一直支持我，我一直都比較任性只做自己想做的事，謝謝你們一路上一直無條件支持我。小時候出國沒有什麼感覺，只想著要趕快看看這個世界做不一樣的事，長大後覺得在台灣的時間真的很少，你們比想像中還要快就老了。這可能是我比較後悔的事。謝謝錢澤霈一直照顧我，沒有你我不可能念完博士。之前我狀態不好的時候不知道原來照顧一個精神有問題的另一半那麼辛苦，一切都不是理所當然，謝謝你還在，如果可以現在換我想要支持你。最後謝謝我自己堅持完成這個學位，只有你自己知道這一切多難代價多高。你真的辛苦了！