Folding of protein secondary structures using a novel polarizable force field method

Yip, Yew Mun

2017

Yip, Y. M. (2017). Folding of protein secondary structures using a novel polarizable force field method. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/72948>

<https://doi.org/10.32657/10356/72948>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**FOLDING OF PROTEIN SECONDARY
STRUCTURES USING A NOVEL
POLARIZABLE FORCE FIELD METHOD**

YIP YEW MUN

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2017

**FOLDING OF PROTEIN SECONDARY
STRUCTURES USING A NOVEL
POLARIZABLE FORCE FIELD METHOD**

YIP YEW MUN

YIP YEW MUN

School of Physical and Mathematical Sciences

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2017

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my supervisor, Distinguished Professor Zhang Dawei for his guidance ever since I joined his lab group in 2008. It has been a wonderful journey for me, and I am extremely grateful to have such an open and understanding supervisor. He formally introduced me to what research is all about, and has been very patient with me as I blundered throughout the years. Also, to Associate Professor Li Tianhu, who has kindly agreed to take over the supervisor role when Prof Zhang had to leave to another university, thank you.

I am also grateful to have very helpful colleagues and friends throughout my postgraduate studies. Thank you especially to Duan Rui, who has helped me tremendously for the coding and programming; Yang Hong and Jinglan for making the lab a much more joyous environment to work in. Regardless of physical or emotional support, all the help was greatly appreciated.

To my parents and brother, thank you. I wouldn't have gone this far and achieved this much if it was not for all of you. Thank you for being there for me always and believing in me. Thank you for your unwavering support. Thank you for having faith in me all these year. Thank you for all the time and effort you have invested in me.

Last but not least, to my wife, Wai Yin. From my QE to the completion of my PhD, you have always been there for me and believing in me. When I am frustrated and down, you are always there to cheer me up. When I lose motivation, you assured me that everything is going to be okay. When I was

busy with the thesis, you took great care of our home, ensuring that it is a conducive place for me to work in.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	3
ABSTRACT	8
LIST OF FIGURES	9
LIST OF PUBLICATIONS	15
PART I INTRODUCTION AND LITERATURE REVIEW	16
1. Introduction	16
1.1 Understanding the importance of protein folding	16
1.3 Introduction to commonly used force fields in MD simulations	19
1.4 Why is it necessary to include the polarization effect in force fields	20
1.5 Viewing of the hydrogen bond at the quantum-mechanical level	21
1.6 What is the polarization effect and how does it affect the hydrogen bond	21
1.7 Introducing the various polarizable force field models	22
1.7.1 Fluctuating Charge Model	23
1.7.2 The Drude Oscillator	24
1.7.3 The Inducible Point Dipole Model	26
1.7.4 The Polarized Protein-Specific Charge Model	26
1.7.5 The Effective Polarizable Bond Model	27
1.7.6 The Polarizable Hydrogen Bond Model	28
1.8 Introducing the Polarized Structure-Specific Backbone Charge Model	30
1.9 Outline of Thesis	31
References	34
PART II EXPERIMENTAL	43

2.1 Polarizable Structure-specific Backbone Charge (PSBC) Model	44
References	46
PART III RESULTS AND DISCUSSION	47
3. <i>Ab initio</i> folding of mixed-fold FSD-EY protein using formula-based	
polarizable hydrogen bond (PHB) charge model	47
3.1 Introduction	47
3.2 Simulation Details for REMD simulation.....	49
3.2.1 Simulation Details for the Minimization of the Folded FSD-EY	
under Explicit Solvent condition	52
3.2.2 Ascertaining of REMD simulation convergence.....	53
3.3 Results and Discussion	54
3.3.1 Contrasting the ability of PHB, PHB0.7 and AMBER ff99SB charge	
models in the protein folding of FSD-EY	55
3.3.2 FSD-EY folding via conventional MD simulation under explicit	
solvent condition	63
3.4 Chapter Summary	67
References	68
4. An initial application of the PSBC model on the folding of simple β-sheets –	
Tryptophan zippers 1 and 2.....	73
4.1 Introduction	73
4.2 Simulation Details.....	77
4.3 Results and Discussion	79
4.3.1 Comparison of the quality of MD simulations between the PSBC	
model and the traditional AMBER force field.....	79

4.3.2 Folding mechanisms of trpzip1 & 2	85
4.4 Chapter Summary	89
References	90
5. Application of the Polarized Structure-Specific Backbone Charge Scheme	
on the Folding of Beta3s.....	97
5.1 Introduction	97
5.2 Simulation Details.....	100
5.3 Results and Discussion	100
5.4 Chapter Summary	105
References	106
6. Application of the PSBC model onto the folding of α-helices with varying	
helical propensities.....	112
6.1 Introduction	112
6.2 Simulation Details.....	116
6.3 Results and Discussion	118
6.3.1 Peptide with high helicity.....	121
6.3.2 Peptides with low helicity	129
6.3.3 Folding of E5 peptide.....	132
6.3.4 Folding of ES peptide	139
6.4 Chapter Summary	142
References	144
7. Conclusion.....	149

ABSTRACT

Protein folding is a complicated process that bridges the understanding between the protein structure and its biological function. The “protein folding problem” is continuously being pursued in the protein community up till today and there have been many principal studies conducted to comprehend protein folding. While numerous experimental efforts have been devised to probe the understanding of the protein folding process, these methods faced limitations. With technology advancing rapidly, the MD simulation is becoming an essential tool to understand the protein folding process dynamically at the atomic level. Modelling of the exact nature of these interactions makes theoretical prediction of native structures a complex task as the correctness of the prediction depends heavily on the quality of the force field used. However, the absence of the polarization effect in force fields has reduced the accuracy of MD simulations when studying protein folding. Therefore, we developed a novel atomic charge update method known as the Polarized Structure-specific Backbone Charge (PSBC) model. The PSBC model accounts for polarizability by periodically updating the charges of atoms involved in backbone hydrogen bonding during the MD simulation. In addition, the model does not employ computationally expensive quantum-mechanical calculations on-the-fly during MD simulations. The model is applied to peptides of various challenging secondary structures to rigorously test its capability, effectiveness and accuracy.

LIST OF FIGURES

Figure 1 Schematic illustration of the PSBC model.

Figure 2 Hierarchical cluster analyses of the simulations performed under the (A) PHB, (B) PHB0.7 and (C) AMBER ff99SB force field at 267 K.

Figure 3 C α -RMSDs probability distribution curves of the FSD-EY protein folding at 267 K (left) and 296 K (right), obtained from the REMD simulations conducted using the (A) PHB, (B) PHB0.7 and (C) AMBER force field.

Figure 4 (A) Evolution of the α -helical content over simulation time for the REMD simulation at 267 K of the FSD-EY protein folding conducted with PHB (black), PHB0.7 (red) and the AMBER force field (green). (B) Probability distribution plots of the α -helical content in FSD-EY at 267 K for PHB (black), PHB0.7 (red) and the AMBER force field (green).

Figure 5 Temperature-dependent profiles for REMD simulations performed under PHB (black), PHB0.7 (red) and AMBER (green) charges.

Figure 6 2D-FELs of FSD-EY obtained for the (A) PHB model, (B) the PHB0.7 model and (C) the non-polarizable AMBER ff99SB force field. The reaction coordinates used are the number of backbone hydrogen bonds (Hb Count) and the COM distance of the Y7 and F12 side chains. The represented structures for the regions 1-4 are shown on the right of the 2D-FELs in (A) to (C). (D) A schematic illustration of the backbone hydrogen bonds that are formed between the i and $i + 4$ residues in the α -helical region of FSD-EY and the location of the Y7 and F12 residues in the β -sheet domain.

Figure 7 Folding of the FSD-EY mixed-fold protein under explicit solvent condition: (A) Evolution of the RMSD value for the full protein (black), α -helix (red) and β -sheet (blue) with reference to the FSD-EY NMR structure. (B)

Representative structures obtained from the cluster analysis of the conventional MD simulation at different periods of the simulation. The percentages shown below the structures represents the probability of the structures occurring during that period of the MD simulation.

Figure 8 Amino acid sequences for tryptophan zippers 1 (left) and 2(right). The turn residues are highlighted in red for trpzip1 and green for trpzip 2.

Figure 9 Simulated melting curves of trpzijs 1 and 2 (top). Melting temperature values of trpzijs 1 and 2 (bottom).

Figure 10 Backbone (top) and tryptophan side-chains (bottom) root-mean-square deviations of trpzijs 1 and 2 in the MD simulations.

Figure 11 Free energy landscapes of the trpzijs using traditional AMBER (top) and PSBC-incorporated AMBER (bottom) force fields.

Figure 12 Lowest energy and lowest RMSD structures of the trpzijs under the PSBC model and the traditional AMBER force field.

Figure 13 DSSP diagrams of trpzijs 1 and 2.

Figure 14 (a) Hydrogen bond pairs found in NMR structures of trpzijs 1 and 2. (b) Percentages of frames where the hydrogen bond lengths are less than or equal to 3 Å for trpzijs 1 and 2 in the last 12.5 ns of the simulations.

Figure 15 Normalized frequency diagrams of each amino acid residue's ϕ and ψ angles in trpzijs 1 and 2 simulated using traditional AMBER force field (black) and PSBC model (red), compared with NMR structure.

Figure 16 Schematic representation of beta3s. The hydrogen bonds are indicated in green dotted lines, while the purple arrows indicate C α H protons that give rise to NOEs.

Figure 17 Most populated structures from the 3 runs in AMBER and PSBC.

Figure 18 DSSP analysis of run 3 in AMBER and PSBC. It is clearly observed that the AMBER force field is unable to fold a stable three-stranded anti-parallel β -sheet in the absence of the PSBC model.

Figure 19 NOE distances of the 4 residue pairs proposed experimentally to confirm the successful folding of beta3s under AMBER (left) and PSBC (right).

Figure 20 Hydrogen bond distances of various residue pairs in beta3s. The PSBC model (shown in red) is able to stabilize the hydrogen bonds to a greater extent than the AMBER force field (shown in black).

Figure 21 Free energy landscapes of the beta3s folding process using the radius of gyration and the $C\alpha$ - $C\alpha$ distances of the residue pairs that give rise to the NOEs. The lowest energy basin is indicated in red while the highest energy basin is indicated in blue.

Figure 22 (A) Helical content of ER, E5 and ES peptide determined through simulations conducted as a function of the experimental helicity of the three peptides. (B) Summary of the helical content predicted through simulations and the experimental helicity of ER, E5 and ES peptide. (C) Amino acid sequences of ER, E5 and ES peptide.

Figure 23 (A) Schematic representation of (*i*+4)-ER peptide with salt bridges (red dotted lines) formed between Glu and Arg residues illustrated. (B) Secondary structure assignment for ER peptides simulated using PSBC (*top*) and AMBER (*bottom*) charges. (C) RMSF of ER peptide during the last 50 ns of the simulations conducted using PSBC (black) and AMBER (red) charges.

Figure 24 Secondary structure variation of the ER peptide during the folding simulation performed using PSBC and AMBER charges. DSSP algorithm was used to monitor the changes in secondary structures.

Figure 25 (A) Free energy landscape of ER folding simulations conducted using PSBC with structures representing populations on the free energy landscape included. Overlap of the theoretical structure of ER (yellow) and folded ER (cyan) acquired through simulations conducted using PSBC. (B) Free energy landscape of ER folding simulations conducted using AMBER charges with structures representing populations on the free energy landscape included. Overlap of the theoretical structure of ER (yellow) and folded ER (cyan) acquired through simulations conducted using AMBER charges. (C) Lowest energy conformations (grey) acquired through the four simulations conducted using PSBC and AMBER charges overlapped with the theoretical structure of ER (black) for comparison.

Figure 26 (A) Schematic representation of 13 possible backbone hydrogen bonds in ER peptide. (B) (top) Average occupancy (bottom) and average length of 13 backbone hydrogen bonds formed in folded ER during the four simulations performed each using PSBC and AMBER charges. (C) Probability distribution of backbone hydrogen bonds formed during all simulations conducted using PSBC (blue) and AMBER (red) charges.

Figure 27 Secondary structure assignment for (A) E5 and (B) ES peptides simulated using PSBC (*top*) and AMBER (*bottom*) charges. Representative structures of the three most populated clusters acquired through cluster

analyses performed for MD trajectories acquired for the folding of (C) E5 and (D) ES peptides under PSBC (*top*) and AMBER ff03 force field (*bottom*). Red represents 310- helix, blue represents α -helix and cyan represents turn.

Figure 28 (A) Schematic representation of four backbone hydrogen bonds formed when E5 peptide was folded using standard AMBER charges. (B) Occupancy of the four possible backbone hydrogen bonds formed in E5 during the simulation performed using PSBC (blue) and AMBER (red) charges. (C) Variation in the length of backbone hydrogen bonds possibly formed between hydrogen bond pairs (7@O-11@H, 8@O-12@H, 9@O-13@H and 10@O-14@H) in E5 during simulations performed under AMBER force field and PSBC.

Figure 29 (A) Number of water molecules within 3.4 Å radii of the atoms in the four hydrogen bond pairs namely 7O-11H, 8O-12H, 9O-13H and 10O-14H. (B) Schematic representation of backbone hydrogen bonds formed between 8O and 11H ($i+3 \rightarrow i$ hydrogen bonding), 8O and 12 H ($i+4 \rightarrow i$ hydrogen bonding) and 8O and 13H ($i+5 \rightarrow i$ hydrogen bonding) for the representative structure of the most populated cluster achieved through PSBC-based simulation. (C) Variation in the length of backbone hydrogen bonds possibly formed between Ala8 and Ala11 and Ala8 and Ala13 in E5 folded using the PSBC model.

Figure 30 (A) Schematic representation of the representative structure of the most populated cluster acquired through PSBC-based simulation of E5 with possible backbone-sidechain hydrogen bond pairings between Gln 4 and Gln 9 (D1 and D2) and Glu 6 and Gln 9 (D3) labelled accordingly. (B) Schematic representation of the representative structure of the most populated cluster acquired through AMBER-based simulation of E5 with Glu4, Glu6 and Glu9

labelled accordingly. (C) Variation in the distances, D1, D2 and D3 as the simulation progresses for E5 folded using PSBC (blue) and AMBER (red) charges.

Figure 31 (A) Schematic representation of three backbone hydrogen bonds formed when ES peptide was folded using standard AMBER charges. (B) Occupancy of the three possible backbone hydrogen bonds formed in ES during the simulation performed using AMBER charges. Zero occupancy of these backbone hydrogen bonds were recorded for PSBC-based simulation.

Figure 32 (A) Variation in the length of backbone hydrogen bonds possibly formed between 1O-5H, 2O-6H and 3O-7H when ES was folded using AMBER force field. The lengths between the aforementioned atom pairs was also calculated for the simulation performed under PSBC model for comparison. (B) Average hydrogen bond energy calculated for hydrogen bonds formed between 1O-5H, 2O-6H and 3O-7H in ES folded under AMBER force field (red) and PSBC model (blue).

LIST OF PUBLICATIONS

1. Hartono, Y. D.; Yip, Y. M.; Zhang, D. Adsorption and Folding Dynamics of MPER of HIV-1 gp41 in the Presence of Dpc Micelle. *Proteins Struct. Funct. Bioinforma.***2013**, *81* (6), 933–944.
2. Liu, T.; Yip, Y. M.; Song, L.; Feng, S.; Liu, Y.; Lai, F.; Zhang, D.; Huang, D. Inhibiting Enzymatic Starch Digestion by the Phenolic Compound Diboside A: A Mechanistic and in Silico Study. *Food Res. Int.***2013**, *54* (1), 595–600.
3. Toh, Z. S.; Wang, H.; Yip, Y. M.; Lu, Y.; Lim, B. J. A.; Zhang, D.; Huang, D. Phenolic Group on A-Ring Is Key for Dracoflavan B as a Selective Noncompetitive Inhibitor of α -Amylase. *Bioorg. Med. Chem.***2015**, *23* (24), 7641–7649.
4. Mohamed Rashid, A.; Lu, K.; Yip, Y. M.; Zhang, D. Averrhoa Carambola L. Peel Extract Suppresses Adipocyte Differentiation in 3T3-L1 Cells. *Food Funct.***2016**, *7* (2), 881–892.
5. Zhang, D.; Yip, Y. M.; Li, L. In Silico Construction of HK2-VDAC1 Complex and Investigating the HK2 Binding-Induced Molecular Gating Mechanism of VDAC1. *Mitochondrion***2016**, *30*, 222–228.
6. Zhang, D.; Lazim, R.; Yip, Y. M. Ab Initio Folding of Mixed-Fold FSD-EY Protein Using Formula-Based Polarizable Hydrogen Bond (PHB) Charge Model. *Chem. Phys. Lett.***2017**, *684*, 24–29.

INTRODUCTION AND LITERATURE REVIEW

Chapter 1

1. Introduction

1.1 Understanding the importance of protein folding

The biological function of proteins is associated with its functional shape or conformation. To achieve the desired functional conformation, the proteins undergo protein folding, a physical process by which a protein chain acquires its native 3-dimensional structure, a conformation that is usually biologically functional, in an efficient and reproducible manner.

When a protein misfolds or folds incorrectly, certain structural features that are involved in the protein's biological functions are lost. This leads to a breakdown in its biological function and gives rise to diseases. On the other hand, misfolding may also cause aggregation with other misfolded proteins. This aggregation mostly results in the gain of toxicity within the cell.

Some of the more commonly known diseases associated with protein misfolding include Alzheimer's disease, cystic fibrosis and amyotrophic lateral sclerosis (ALS).

The Alzheimer's disease is associated with the 'A β hypothesis'. Under this hypothesis, the amyloid- β 42 (A β ₄₂) monomers undergo misfolding and aggregate into damaging oligomers and fibrils within the proximity of A β ₄₂ plaques. These aggregates then cause certain inflammatory responses. While the mechanisms remain unknown, the aggregation eventually lead to

aggregation, phosphorylation and propagation of the tau protein, which is associated with microtubules constitutes as the main substance of harmful tangles. This then causes the neurons and synapses to be dysfunctional, leading to additional inflammatory responses. The progressive degeneration of affected neurons and synapses leads to cognitive decline, Alzheimer's disease and disability.¹⁻⁵

Cystic fibrosis is caused by mutations of the cystic fibrosis transmembrane conductance regulator (CFTR), a plasma membrane chloride channel. The most common mutation is the deletion of a phenylalanine residue at position 508 in CFTR. This mutation causes misfolding and targets the protein for degradation.⁶⁻⁹

In ALS, it is hypothesized that the misfolding of the prion-like protein is responsible for cellular toxicity as well as the spreading of pathology. Studies have shown that mutations in the human copper-zinc superoxide dismutase (SOD1) and other genes cause misfolding and induce a pathologic protein conformation in ALS.¹¹⁻¹⁴

Therefore, it is crucial to understanding the folding mechanism¹⁵ and the interaction of the amino acids that are involved during the folding process, in order to design inhibitors that are able to effectively prevent aggregation¹⁶ or rescue the misfolding¹⁷.

1.2 How MD simulation is used in understanding the protein folding mechanism

There have been numerous methods developed over the years to seek answers to some important questions in protein folding, such as the crucial

amino acids required for successful folding; the main intra-peptide interactions that are required for successful folding; the intermediates formed during the folding process. On the other hand, studying factors that leads to misfolding are also important to elucidating a solution to rescue the misfolding. These factors are similar to the ones studied for successful protein folding.

There are quite a number of experimental methods that are used to study protein folding, such as the far-UV circular dichroism (CD); the Fourier transform infra-red (FTIR) spectroscopy; the pulse labelling NMR and pulse labelling mass spectrometry (MS).¹⁸ While the far-UV CD and FTIR spectroscopy measure the average backbone conformation and hydrogen bond properties^{18,19,22}, the pulse labelling NMR and MS sequence specific hydrogens bonds and hydrogen bonds formed in intermediates.^{18,20,21}

Although these methods are effective to a certain extent in studying protein folding, there are shortcomings pertaining to these methods. The far-UV CD is unable to provide high resolution structural information since it is only able to obtain the averaged structure.^{18,22} If the peptide is soluble in aqueous solution, the FTIR spectroscopy would observe an overlapping of the water and the amide I band absorbance signals.^{18,19,23} The NMR method is only applicable to protons that are strongly protected in the native state, resulting in a potential bias towards detecting native-like structure. In addition, non-native interactions formed during the folding process are less likely to be detected by NMR.^{18,21} For the MS method, any residual level information about the folding are unable to be provided.^{18,20}

While mutagenesis studies can identify the residues that are critical to the misfolding process²⁴, it is unable to identify the transient non-covalent interactions made by these residues. This is where *in silico* studies come into play. *In silico* studies refer to molecular dynamics (MD) simulations where Newton's laws of motions are integrated over time to locate the position and velocity of particles.^{25,26} While MD simulations are unable to simulate the exact experimental conditions of the protein folding process, they are able to provide insights on the plausible folding mechanism.²⁷⁻²⁹ With the plausible folding mechanism identified, experimental studies can then be conducted for verification. This iteration between *in silico* and experimental studies not only saves time and cost in designing small molecular inhibitors, but it is also able to provide a deeper understand about the protein of interest so that we can make a more rational choice when designing inhibitors.

1.3 Introduction to commonly used force fields in MD simulations

Molecular mechanics (MM) simulations have been widely used to study biological systems. These simulations make use of force fields, which are mathematical functions that calculates the energies of the simulated system. Ever since the first MD simulation of the bovine pancreatic trypsin inhibitor in 1977, there have been many force fields developed over the past 30 years, such as AMBER, CHARMM, GROMOS, OPLS and others.³⁰ These force fields generally contain the following energy terms, the bonded and the non-bonded energy terms. The bonded energy term contains the bonds, angles and dihedrals energies while the non-bonded energy term contains the electrostatic and van der Waals energies.

1.4 Why is it necessary to include the polarization effect in force fields

Current force fields used in MD simulations do not account for the polarization effect dynamically.³¹ Some of these well-known force fields include AMBER³² and CHARMM³³, where they have been used in many simulation studies of protein-ligand systems³⁴ and protein folding^{35,36}. Taking AMBER as an example, it is unable to account for the polarization effect dynamically as all atoms in the protein are given their atomic charges at the start of the simulation and these charges remain fixed throughout the simulation.^{37,38} However, proteins do not remain in the same conformation for a long time, they constantly change conformation to seek out the lowest energy conformation that is believed to be the native state as explained by the Levinthal paradox.^{39,40} The paradox states that if a protein of a certain amino acid length were to be folded into its native structure, mathematically it would take an infinite amount of time to do so. However, proteins in biological systems fold in a matter of seconds or faster, therefore the protein must have found a more efficient way to fold into its native structure, which is to constantly locate the lowest energy conformation. Since proteins change conformation often, it is then easily understood that at the atomic level, different atoms are in different microenvironments constantly.⁴¹⁻⁴³ From fundamental chemistry, it is learnt that when atoms of different polarity are close to each other, electron densities of both atoms will undergo changes as compared to when they are infinitely far away from each other. Therefore, it is important to consider this polarization effect during MD simulations so as to increase the accuracy of the simulation for analytical purposes.⁴⁴⁻⁴⁶

1.5 Viewing of the hydrogen bond at the quantum-mechanical level

In most force fields, the hydrogen bond is treated as an electrostatic interaction between two oppositely charged atoms, since the hydrogen atom of the donor group is positively charged due to its bonding with an electronegative atom and the acceptor group is another electronegative element. Force fields assign fixed charges to atoms during a molecular dynamics (MD) simulation. At the quantum-mechanical level, physical bonds do not exist and are simply an interaction of electron clouds. When a bond length changes, these electron clouds overlap each other for interaction. Therefore, one would expect that a long bond length would result in minimal electron cloud overlapping as compared to a short bond length.

1.6 What is the polarization effect and how does it affect the hydrogen bond

On the other hand, during the overlap, electrons repel each other as well as attracted to the positively charged nucleus. The strengths of these attractions and repulsions is dependent on the proximity of the interacting electron clouds, which is basically the bond length. This is known as the polarization effect, where one electron cloud influences the other, and the extent of influence is dependent on the bond length. However, since fixed charges are assigned fixed to atoms during the MD simulation, it is unable to account for the polarization effect when bond lengths change during the MD simulation. When the polarization effect is not accounted for a change in bond length, the atomic charges of the hydrogen bond pair will not be accurate. When that happens, the energy term for hydrogen bonding will also be inaccurate. Putting this perspective at a macroscopic level, this will affect the

energy of the intermediate states formed during protein folding and finally, the locating of the native structure.

1.7 Introducing the various polarizable force field models

Recognizing the importance to include the polarization effect in molecular mechanical force fields, a number of polarizable force field models have been developed over the past decade, such as the fluctuating charge (FQ), Drude oscillator and inducible point dipole models. Discussions of each model and their applications will be presented in the following sections.

While there are several polarization models available^{45,47-49}, most of them are either too difficult to implement or they are too computationally costly. Computationally costly in this sense refers to not only the computational resources required to do the calculations during the simulations, but it also refers to the monetary cost of such calculations. In addition, it is possible that after running long time scales simulations, due to the inaccuracy of the force field used, the experimental results are unable to be reproduced in the simulations. Therefore, it is important to derive a polarization model that is able to achieve a decent balance between computational cost and accuracy.

There are many polarization models developed to address this problem, such as the polarized protein specific (PPC) model²⁹, the effective polarizable bond (EPB) model³¹ and the polarizable hydrogen bond (PHB) model.¹⁶ All 3 models re-derive and update the atomic charges periodically throughout the simulation.

1.7.1 Fluctuating Charge Model

The fluctuating charge (FQ) model works by equalizing the electronegativities between two atoms, where it assumes that charges flow between the two atoms until the electronegativities are equalized. This model allows the partial charge of an atom to change and adapt to different electrostatic environments.⁵⁰ The variable partial charges are calculated by minimizing the electrostatic energy for a given geometry of the molecule. The number of parameters are minimal in the FQ model and easier to apply because the polarization effect is only kept at point charges without any introduction of new interactions.⁵¹

Banks et al. incorporated the FQ model into biomolecular simulations by using an alanine dipeptide. Several conformations of the alanine dipeptide were used to fit the parameters of the FQ model to the atom types of the dipeptide. The electrostatic model constructed for the alanine dipeptide allowed the energies to be accurately reproducible. The model was then assimilated into the OPLS-AA force field, where it was found to perform well during the ranking of the alanine tetrapeptide conformational energies. Banks et al. then used the backbone and side chain parameters from the alanine dipeptide and small molecules to construct a similar electrostatic model for the serine dipeptide.⁵²

However, the FQ model possesses its own sets of inadequacies as well. In the case of planar or linear molecules or chemical moieties, the FQ model is unable to reproduce the out-of-plane polarization since the model is based on the equalization of the electronegativity along the bonds between atoms.⁵¹

1.7.2 The Drude Oscillator

In the Drude oscillator model, the electronic polarization effect is represented by an atom or ion as a two-particle system: a shell with the charge $q_{i,D}$ and a core with the charge $q_{i,0}$. Together, the charged shell and core make up a Drude particle and they are connected via a harmonic spring. In order to mimic the electronic polarization effect due to an external electrostatic field, the magnitude of the shell's and core's charges are fixed. The quantification of the atomic polarizability, α_i , is then linked to the force constant of the harmonic spring where

$$\alpha_i = \frac{(q_{i,D})^2}{k}$$

and k is the force constant. Therefore, by fitting molecular polarizability data and experimental intermolecular interaction energies along with other energetic properties, the atomic charge magnitudes and harmonic force constants are able to be obtained.

Drude-1 was the first protein model that was based on the Drude oscillator model. It was constructed by simply a combination of the N-methyl amide (NMA) peptide cap and ethane with the alanine dipeptide, which is the most commonly used model during the parameterization of protein force fields. The internal energetic terms were optimized and based on high-level gas phase quantum-mechanical data, two-dimensional cross term energy correction maps (CMAP) were generated. However, the oversimplified Drude-1 model was unable to correlate with the J-coupling data in good agreement when it was applied on the polyalanine peptide model. Instead, it biased the

peptide towards extended conformations. Further analysis indicated that this observation was caused by inadequate treatment of the polarization effect when electrostatic parameters derived from NMA was directly applied onto the polypeptide.

With the lessons learnt from the Drude-1 model, the Drude-2 model was constructed. The Drude-2 model was developed by generating the electrostatic potential maps individually of the alanine dipeptide in 5 different conformations, which was followed by averaging of these parameter sets. However, the Drude-2 model was unable to reproduce the gas-phase quantum-mechanical dipole moments accurately, while the biasness towards extended conformations remained.

Learning from the failures observed in the Drude-1 and Drude-2 models, the Drude-3 model was constructed by using a Monte Carlo/simulated annealing (MC/SA) protocol to refine the Drude-2 model. The construction of the model targeted a wide variety of data, including the quantum-mechanical interaction energies between the alanine dipeptide and water, the dipole moments, polarizability and relative energies of the dipeptide in different conformations. In addition, the gas phase polarizabilities were scaled by a factor of 0.85 in order to strike a balance between solvent and intrapeptide interactions. This force field was later renamed as the Drude-2013 force field. Significant emphasis was placed on the model's transferability between different compounds when it was developed for small molecules. The final aim was to eventually assimilate all the parameter sets so that the Drude oscillator model can be applied on biomolecules.⁵³

However, the Drude oscillator model possesses a limitation known as the “polarization catastrophe”. When a Drude particle is displaced too far away from its parent atom, where the displacement is comparable to inter-atomic separations, this results in unphysically large interaction energies which eventually leads to simulation failure. The “polarization catastrophe” is most likely to occur in a Drude polarizable force field because the Drude particles are not situated at the center of the Lennard-Jones (LJ) repulsive sphere, and is therefore deshielded from other charged ions or atoms.^{54,55}

1.7.3 The Inducible Point Dipole Model

In the induced dipole model, numerous isotropic inducible point dipoles are placed on each atom. In the absence of hyperpolarization effects which occur due to strong electric fields, the induced dipole produces a linear response with respect to the electric field, where the induced dipole of an atom is the product of the total electric field and a scalar atomic polarizability.

The total electric field comprises of the contribution from other induced dipoles and the external electric field that arises from the permanent charge sources. Therefore, to reproduce the polarization effect using isotropic atomic polarizabilities, the induced dipoles residing in the same molecule are expected to interact with each other.^{56,57,58}

1.7.4 The Polarized Protein-Specific Charge Model

The Polarized Protein-specific Charge (PPC) model is developed by Zhang et al.⁶⁰ This model targets the AMBER force field where atomic charges are fixed during biomolecular simulations, clearly demonstrating the lack of

inclusion of the polarization effect. Learning from the previous polarizable charge models, the PPC model aims to introduce more exclusiveness to each biomolecular system simulated instead of developing a model that is based on a dipeptide. However, performing quantum-mechanical calculations and charge fitting on the whole protein is too time-consuming and computationally expensive. Therefore, the Molecular Fractionation of Conjugate Caps (MFCC) method is developed to supplement the PPC model.

The MFCC method decomposes a protein into amino acid-based fragments and caps the ends of the fragments properly with protein caps.⁵⁹ After which, the fragment undergoes electronic calculation to obtain its electron density distribution. The electron density distribution is then used to generate the electrostatic potential of the fragment, and followed by charge fitting of the electrostatic potential to obtain the atomic charges of the fragment. During the whole calculation and charge fitting process of the fragment, the rest of the protein is treated as background charge. Although this model is able to account for the polarization effect dynamically when the protein simulation is running, it is computationally costly when the protein of interest or the fragmental molecules become larger in size.

1.7.5 The Effective Polarizable Bond Model

In the presence of an external electric field, the electron distribution of a gas phase molecule undergoes distortion. During this distortion or polarization, two opposite energetic effects are observed. The polarization of the molecule enhances the interaction energy between the molecule and external electric field, so as to lower the total energy of the system. On the

other hand, the distortion brings about an increase in the internal energy of the molecule. This energy that is required for the electron distribution of the molecule to be distorted is known as the polarization cost or distortion energy. Via electronic structure calculation, a quantitative relationship between the polarization state and its energetic cost can be derived.

The Effective Polarizable Bond (EPB) model is based on the rationale that the polarization effect is a balance between two opposing energetic effect, namely the electrostatic and internal energies. After the model has calculated the amount of charge transferred during the polarization, it will then be added to the atomic charge value prior to the polarization. In the EPB model, all the polar moieties of the amino acids are considered polarizable, with the relevant parameters relating to polarizability, derived from the fitting of electrostatic properties via quantum-mechanical calculations. Similarly, the charge update is conducted periodically throughout the biomolecular simulation of interest.⁶¹

1.7.6 The Polarizable Hydrogen Bond Model

The polarizable hydrogen bond (PHB) model is a model derived by Gao et al.⁶² It is based on the concept that backbone hydrogen bonds are crucial to the formation of secondary motifs in protein folding. The backbone hydrogen bond is formed between the N & H atoms of one amino acid and the C & O atoms of another amino acid. Since these 4 atoms contain electronegative atoms (N & O), it is therefore expected that polarization will occur as the hydrogen bond distance changes during the MD simulation. By locating a relationship between the hydrogen bond distance and the charge transferred

between these 4 atoms, the charges of these atoms can then be simply updated during the course of the simulation to account for the polarization effect directly.

Compared to the EPB model, the PHB model is a simpler concept to implement in simulation studies. The PHB model derives a relationship between the amount of charge transferred during polarization and the distance between the atoms. By varying the hydrogen bond distance between backbone atoms and calculating the amount of charge transferred between these atoms, functions of charge transferred and hydrogen bond distance is obtained and implemented in simulation runs. Therefore, during the simulation, as the distance between the backbone atoms vary, the amount of charge transferred between these atoms will vary and similar to the EPB model, the amount of charge is added to the atomic charge value prior to the polarization.

While this model has been successfully used to fold various peptides^{62,63}, these peptides are mainly alpha helices. However, many of the protein misfolding diseases are related to beta sheets.^{14,64} Therefore, there is a need to verify if the PHB model is capable of folding beta sheets. In order to capture the model's strength in folding alpha helices while probing its capability in folding beta sheets, the FSD-EY protein⁴⁶ is chosen. The FSD-EY protein belongs to the FSD-1 protein family where the amino acid sequence is completely designed computationally. It consists of two secondary motifs, an alpha helix and two anti-parallel beta sheets linked by a loop (Figure 1). Not only is this protein small in size (28 residues), but also there are experimental

studies⁶⁶⁻⁶⁸ readily available, rendering it as an ideal choice to test the PHB model. This is covered further in Chapter 3.

1.8 Introducing the Polarized Structure-Specific Backbone Charge Model

The initial polarizable charge models (FQ, Drude oscillator and induced dipole) recognize the absence of the polarization effect in force fields, and implemented different interpretations of the polarization effect. Through the use of numerous complicated mathematical functions, these models were developed into force fields for biomolecular simulations. However, due to the wide variety of biomolecular systems, it is evident that these models remain inadequate to account for the polarization effect dynamically throughout the MD simulation.

The more recent models (PPC, EPB and PHB) improved the situation by deriving atomic charges from quantum-mechanical calculations and performing charge fitting using the RESP method. In the case of PPC, the periodic quantum-mechanical calculations proved to be computationally expensive when it is applied to a large protein. The EPB and PHB models circumvented this problem by using two alanine amino acids to mimic the hydrogen bonding networking in secondary structure and only performed calculations on the atoms involved in the hydrogen bond. By varying the distance of the hydrogen bond, a relationship between the charge transferred between the atoms and the hydrogen bond distance was derived. This relationship took the form of much simpler mathematical functions which reduced the computational cost while maintaining the dynamicity of charge updating in MD simulations. As the EPB and PHB models were derived from

two alanine amino acids in an α -helix conformation, most of the successful protein folding simulations are α -helices even though there are a few cases of β -sheets. Therefore, a possible biasness towards α -helices may be inherent in these models.

Therefore, to circumvent this biasness, the Polarized Structure-specific Backbone Charge (PSBC) model was proposed and constructed. The PSBC model builds on the insights obtained from the initial models (FQ, Drude oscillator and induced dipole) where it recognizes that current force fields do not account for the polarization effect explicitly and electronegativity plays a key role in the polarization of the electron clouds. The PSBC model also builds on the simplification of the mathematical functions that describes polarization dynamically in MD simulations from the PPC, EPB and PHB models. In order to increase the robustness of the PSBC model, two alanine amino acids in the β -conformation were used to derive the mathematical functions that best describe the polarization effect during the folding of β -sheets. In addition, the PSBC model uses the Dictionary of Protein Secondary Structure (DSSP), which is based on the Φ and Ψ angles of amino acids in various secondary structure, to determine when to apply the functions for the α -helix and β -sheet. This eliminates any inherent biasness during the application of the PSBC model in MD simulations.

1.9 Outline of Thesis

Chapter 1 introduces the concept of protein folding and the problems facing protein folding, both experimentally and computationally. The chapter then establishes the areas that experimental methods are lacking in and how

computational methods are able to study protein folding further in depth. Following that, the chapter brings forth the problem that the absence of describing the polarization effect in MD simulations is hindering the accuracy and progress of protein folding simulations. On the same note, the chapter continues with a critical review of the various polarizable charge models developed over the years, emphasizing on: the basis of these models, how the newer models learnt from the older models, and the problems that these models are facing.

Chapter 2 introduces the methodology of the polarizable structure-specific backbone charge (PSBC) model.

Chapter 3 explores the application of PHB model on the mixed-fold protein, FSD-EY. Since the PHB model is the latest model to be based on for the construction of the PSBC model, the application of the PHB model on the mixed-fold protein will provide insights to the following questions: 1. Is there inherent biasness of the PHB model to fold α -helices? 2. Is the PHB model capable of folding β -sheets? 3. Is the PHB model capable of folding the mixed-fold protein, which remains as of current a challenging protein moiety?

Chapter 4 demonstrates the first application of the PSBC model to fold the well-known simple β -sheets – Tryptophan zippers (trpzips 1 and 2). The trpzips were designed to understand the relationship between β -turn residues and how they affect the β -propensity. Therefore, the trpzips are relatively well-studied experimentally and provide as a suitable peptide subject to test the capability of the PSBC model in folding β -sheets.

In Chapter 5, the benchmark standard for the PSBC model was raised by applying the model to fold a 3-stranded β -sheet, Beta3s. Currently, successful studies of folding 3-stranded β -structures are very rare as the hydrogen bonds formed are much weaker than that of α -helices. Therefore, being able to fold Beta3s successfully with the PSBC model would indicate that there is no inherent biasness of the model to fold α -helices since it was constructed by referencing the PHB model. In addition, the successful folding would also indicate that the PSBC model is capable of discerning and capturing weak hydrogen bonds.

After verifying the capability of the PSBC model in folding β -structures, Chapter 6 puts the PSBC model to the test by determining its ability to fold α -helices of various lengths. There have been many simulation studies that reported the successful folding of long α -helices, but none were able to adequately fold weak or short α -helices. Given the tremendous number of structures in the Protein Data Bank (PDB) that have short α -helices, it is therefore important to test the robustness of the PSBC model in folding α -helices of different lengths.

Chapter 7 recaps on the importance of understanding protein folding and emphasizes on the importance of accounting for the polarization effect dynamically in MD simulations so as to improve the accuracy of protein folding simulations, which eventually aid the in-depth understanding of the protein folding mechanisms, and therefore allowing potential therapeutics for diseases caused by protein misfolding.

References

1. Selkoe, D. J. Cell Biology of Protein Misfolding: The Examples of Alzheimer's and Parkinson's Diseases. *Nat. Cell Biol.***2004**, 6 (11), 1054–1061.
2. Reiman, E. M. Alzheimer's Disease: Attack on Amyloid- β Protein. *Nature***2016**, 537 (7618), 36–37.
3. Tanzi, R. E. The Synaptic A β Hypothesis of Alzheimer Disease. *Nat. Neurosci.***2005**, 8 (8), 977–979.
4. Glenner, G. G.; Wong, C. W. Alzheimer's Disease: Initial Report of the Purification and Characterization of a Novel Cerebrovascular Amyloid Protein. *Biochem. Biophys. Res. Commun.***1984**, 120 (3), 885–890.
5. Selkoe, D. J. The Molecular Pathology of Alzheimer's Disease. *Neuron***1991**, 6 (4), 487–498.
6. Lukacs, G. L.; Verkman, A. S. CFTR: Folding, Misfolding and Correcting the Δ F508 Conformational Defect. *Trends Mol. Med.***2012**, 18 (2), 81–91.
7. Fraser-Pitt, D.; O'Neil, D. Cystic Fibrosis – a Multiorgan Protein Misfolding Disease. *Futur. Sci. OA***2015**, 1 (2), fso.15.57.
8. Qu, B. H.; Strickland, E. H.; Thomas, P. J. Localization and Suppression of a Kinetic Defect in Cystic Fibrosis Transmembrane Conductance Regulator Folding. *J. Biol. Chem.***1997**, 272 (25), 15739–15744.
9. Valastyan, J. S.; Lindquist, S. Mechanisms of Protein-Folding Diseases at a Glance. *Dis. Model. Mech.***2014**, 7 (1), 9–14.

10. Renton, A. E.; Chiò, A.; Traynor, B. J. State of Play in Amyotrophic Lateral Sclerosis Genetics. *Nat. Neurosci.***2014**, *17* (1), 17–23.
11. Silverman, J. M.; Fernando, S. M.; Grad, L. I.; Hill, A. F.; Turner, B. J.; Yerbury, J. J.; Cashman, N. R. Disease Mechanisms in ALS: Misfolded SOD1 Transferred Through Exosome-Dependent and Exosome-Independent Pathways. *Cell. Mol. Neurobiol.***2016**, *36* (3), 377–381.
12. Parakh, S.; Atkin, J. D. Protein Folding Alterations in Amyotrophic Lateral Sclerosis. *Brain Res.***2016**, *1648*, 633–649.
13. Teilum, K.; Smith, M. H.; Schulz, E.; Christensen, L. C.; Solomentsev, G.; Oliveberg, M.; Akke, M. Transient Structural Distortion of Metal-Free Cu/Zn Superoxide Dismutase Triggers Aberrant Oligomerization. *Proc. Natl. Acad. Sci.***2009**, *106* (43), 18273–18278.
14. Soto, C. Unfolding the Role of Protein Misfolding in Neurodegenerative Diseases. *Nat. Rev. Neurosci.***2003**, *4* (1), 49–60.
15. Pace, C. N.; Fu, H.; Lee Fryar, K.; Landua, J.; Trevino, S. R.; Schell, D.; Thurlkill, R. L.; Imura, S.; Scholtz, J. M.; Gajiwala, K.; et al. Contribution of Hydrogen Bonds to Protein Stability. *Protein Sci.***2014**, *23* (5), 652–661.
16. Gao, Y.; Lu, X.; Duan, L. L.; Zhang, J. Z. H.; Mei, Y. Polarization of Intraprotein Hydrogen Bond Is Critical to Thermal Stability of Short Helix. *J. Phys. Chem. B***2012**, *116* (1), 549–554.
17. Li, Y.; Gao, Y.; Zhang, X.; Wang, X.; Mou, L.; Duan, L.; He, X.; Mei, Y.; Zhang, J. Z. H. A Coupled Two-Dimensional Main Chain Torsional

- Potential for Protein Dynamics: Generation and Implementation. *J. Mol. Model.***2013**, 19 (9), 3647–3657.
18. Dobson, C. M. Experimental Investigation of Protein Folding and Misfolding. *Methods***2004**, 34 (1), 4–14.
 19. Barth, A. Infrared Spectroscopy of Proteins. *Biochim. Biophys. Acta - Bioenerg.***2007**, 1767 (9), 1073–1101.
 20. Wani, A. H.; Udgaonkar, J. B. Mass Spectrometry Studies of Protein Folding. *Curr. Sci.***2012**, 102 (2), 245–265.
 21. Juneja, J.; Udgaonkar, J. B. NMR Studies of Protein Folding. *Curr. Sci.***2003**, 84 (2), 157–172.
 22. Kelly, S. M.; Price, N. C. The Use of Circular Dichroism in the Investigation of Protein Structure and Function. *Curr. Protein Pept. Sci.***2000**, 1 (4), 349–384.
 23. Venyaminov SYu; Prendergast, F. G. Water (H₂O and D₂O) Molar Absorptivity in the 1000-4000 Cm⁻¹ Range and Quantitative Infrared Spectroscopy of Aqueous Solutions. *Anal. Biochem.***1997**, 248 (2), 234–245.
 24. Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry***1990**, 29 (31), 7133–7155.
 25. Dill, K. A.; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat. Struct. Biol.***1997**, 4 (1), 10–19.
 26. Karplus, M. The Levinthal Paradox: Yesterday and Today. *Fold. Des.***1997**, 2, S69–S75.

27. Mobley, D. L.; Dumont, É.; Chodera, J. D.; Dill, K. A. Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B***2007**, *111* (9), 2242–2254.
28. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.***2003**, *24* (16), 1999–2012.
29. Ji, C.; Mei, Y.; Zhang, J. Z. H. Developing Polarized Protein-Specific Charges for Protein Dynamics: MD Free Energy Calculation of pKa Shifts for Asp26/Asp20 in Thioredoxin. *Biophys. J.***2008**, *95* (3), 1080–1088.
30. Lopes, P. E. M.; Guvench, O.; Mackerell, A. D. Current Status of Protein Force Fields for Molecular Dynamics Simulations; **2015**; pp 47–71.
31. Xiao, X.; Zhu, T.; Ji, C. G.; Zhang, J. Z. H. Development of an Effective Polarizable Bond Method for Biomolecular Simulation. *J. Phys. Chem. B***2013**, *117* (48), 14885–14893.
32. Zhang, D. W.; Zhang, J. Z. H. Molecular Fractionation with Conjugate Caps for Full Quantum Mechanical Calculation of Protein–molecule Interaction Energy. *J. Chem. Phys.***2003**, *119* (7), 3599–3605.
33. Bruch, M. D.; Dhingra, M. M.; Gierasch, L. M. Side Chain-Backbone Hydrogen Bonding Contributes to Helix Stability in Peptides Derived from an α -Helical Region of Carboxypeptidase A. *Proteins Struct. Funct. Genet.***1991**, *10* (2), 130–139.

34. Garcia, A. E.; Sanbonmatsu, K. Y. Alpha-Helical Stabilization by Side Chain Shielding of Backbone Hydrogen Bonds. *Proc. Natl. Acad. Sci.***2002**, *99* (5), 2782–2787.
35. Osterhout, J. J.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Dyson, H. J.; Wright, P. E. Proton NMR Studies of the Solution Conformations of an Analog of the C-Peptide of Ribonuclease A. *Biochemistry***1989**, *28* (17), 7059–7064.
36. Marqusee, S.; Baldwin, R. L. Helix Stabilization by Glu...Lys⁺ Salt Bridges in Short Peptides of de Novo Design. *Proc. Natl. Acad. Sci. U. S. A.***1987**, *84* (24), 8898–8902.
37. Zerella, R.; Williams, D. H.; Chen, P.-Y.; Evans, P. A.; Raine, A. Structural Characterization of a Mutant Peptide Derived from Ubiquitin: Implications for Protein Folding. *Protein Sci.***2000**, *9* (11), 2142–2150.
38. Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G. *Science***1991**, *253* (5020), 657–661.
39. Ulmschneider, J. P.; Jorgensen, W. L. Polypeptide Folding Using Monte Carlo Sampling, Concerted Rotation, and Continuum Solvation. *J. Am. Chem. Soc.***2004**, *126* (6), 1849–1857.
40. Jang, S.; Shin, S.; Pak, Y. Molecular Dynamics Study of Peptides in Implicit Water: Ab Initio Folding of β -Hairpin, β -Sheet, and $B\beta\alpha$ -Motif. *J. Am. Chem. Soc.***2002**, *124* (18), 4976–4977.

41. Zhou, R.; Berne, B. J. Can a Continuum Solvent Model Reproduce the Free Energy Landscape of a β -Hairpin Folding in Water? *Proc. Natl. Acad. Sci.***2002**, *99* (20), 12777–12782.
42. Zhou, R.; Berne, B. J.; Germain, R. The Free Energy Landscape for Hairpin Folding in Explicit Water. *Proc. Natl. Acad. Sci.***2001**, *98* (26), 14931–14936.
43. Dinner, A. R.; Lazaridis, T.; Karplus, M. Understanding Beta-Hairpin Formation. *Proc. Natl. Acad. Sci.***1999**, *96* (16), 9068–9073.
44. Pande, V. S.; Rokhsar, D. S. Molecular Dynamics Simulations of Unfolding and Refolding of a Beta-Hairpin Fragment of Protein G. *Proc. Natl. Acad. Sci. U. S. A.***1999**, *96* (16), 9062–9067.
45. Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C.-K.; Li, M. S. Free Energy Landscape and Folding Mechanism of a Beta-Hairpin in Explicit Water: A Replica Exchange Molecular Dynamics Study. *Proteins***2005**, *61* (4), 795–808.
46. Sarisky, C. A.; Mayo, S. L. The $\beta\beta\alpha$ Fold: Explorations in Sequence Space. *J. Mol. Biol.***2001**, *307* (5), 1411–1418.
47. Lucent, D.; Vishal, V.; Pande, V. S. Protein Folding under Confinement: A Role for Solvent. *Proc. Natl. Acad. Sci.***2007**, *104* (25), 10430–10434.
48. Lucent, D.; Vishal, V.; Pande, V. S. Protein Folding under Confinement: A Role for Solvent. *Proc. Natl. Acad. Sci.***2007**, *104* (25), 10430–10434.
49. Prabhu, N.; Sharp, K. Protein–Solvent Interactions. *Chem. Rev.***2006**, *106* (5), 1616–1623.

50. Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization Effects in Molecular Mechanical Force Fields. *J. Phys. Condens. Matter***2009**, *21* (33), 333102.
51. Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J.-P. Polarizable Force Fields for Biomolecular Modeling. In *Rev. Comp. Chem.*; 2015; pp 51–86.
52. Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. Parametrizing a Polarizable Force Field from Ab Initio Data. I. The Fluctuating Point Charge Model. *J. Chem. Phys.***1999**, *110* (2), 741–754.
53. Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.***2016**, *116* (9), 4983–5013.
54. Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.***2013**, *9* (12), 5430–5449.
55. Antila, H. S.; Salonen, E. Polarizable Force Fields. In *Biomolecular Simulations*; 2013; pp 215–241.
56. Elking, D.; Darden, T.; Woods, R. J. Gaussian Induced Dipole Polarization Model. *J. Comput. Chem.***2007**, *28* (7), 1261–1274.
57. van Duijnen, P. T.; Swart, M. Molecular and Atomic Polarizabilities: Thole’s Model Revisited. *J. Phys. Chem. A***1998**, *102* (14), 2399–2407.
58. Applequist, J.; Carl, J. R.; Fung, K.-K. Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and

- Determination of Atom Polarizabilities. *J. Am. Chem. Soc.***1972**, *94* (9), 2952–2960.
59. Zhang, D. W.; Zhang, J. Z. H. Molecular Fractionation with Conjugate Caps for Full Quantum Mechanical Calculation of Protein–molecule Interaction Energy. *J. Chem. Phys.***2003**, *119* (7), 3599–3605.
60. Ji, C.; Mei, Y.; Zhang, J. Z. H. Developing Polarized Protein-Specific Charges for Protein Dynamics: MD Free Energy Calculation of pKa Shifts for Asp26/Asp20 in Thioredoxin. *Biophys. J.***2008**, *95* (3), 1080–1088.
61. Xiao, X.; Zhu, T.; Ji, C. G.; Zhang, J. Z. H. Development of an Effective Polarizable Bond Method for Biomolecular Simulation. *J. Phys. Chem. B***2013**, *117* (48), 14885–14893.
62. Gao, Y.; Lu, X.; Duan, L. L.; Zhang, J. Z. H.; Mei, Y. Polarization of Intraprotein Hydrogen Bond Is Critical to Thermal Stability of Short Helix. *J. Phys. Chem. B***2012**, *116* (1), 549–554.
63. Gao, Y.; Lu, X.; Duan, L.; Zhang, D.; Mei, Y.; Zhang, J. Z. H. Direct Folding Simulation of a Long Helix in Explicit Water. *Appl. Phys. Lett.***2013**, *102* (19), 193706.
64. Ross, C. A.; Poirier, M. A. Protein Aggregation and Neurodegenerative Disease. *Nat. Med.***2004**, *10* (7), S10–S17.
65. Stefani, M.; Dobson, C. M. Protein Aggregation and Aggregate Toxicity: New Insights into Protein Folding, Misfolding Diseases and Biological Evolution. *J. Mol. Med.***2003**, *81* (11), 678–699.

66. Sadqi, M.; de Alba, E.; Perez-Jimenez, R.; Sanchez-Ruiz, J. M.; Munoz, V.
A Designed Protein as Experimental Model of Primordial Folding. *Proc. Natl. Acad. Sci.***2009**, *106* (11), 4127–4132.
67. Feng, J. A.; Kao, J.; Marshall, G. R. A Second Look at Mini-Protein Stability: Analysis of FSD-1 Using Circular Dichroism, Differential Scanning Calorimetry, and Simulations. *Biophys. J.***2009**, *97* (10), 2803–2810.
68. Li, W.; Zhang, J.; Wang, W. Understanding the Folding and Stability of a Zinc Finger-Based Full Sequence Design Protein with Replica Exchange Molecular Dynamics Simulations. *Proteins Struct. Funct. Bioinforma.***2007**, *67* (2), 338–349.

PART II

EXPERIMENTAL

Chapter 2

2. Methodology

2.1 Polarizable Structure-specific Backbone Charge (PSBC) Model

Flowchart of the PSBC model

Figure 1 Schematic illustration of the PSBC model.

The PSBC model was formulated based on the understanding that the formation of hydrogen bonds lead to the reorganization in the electronic density of the hydrogen bond pairs as a result of mutual polarization. This phenomenon lowers the electrostatic interaction energy between the hydrogen bond donor and acceptor hence favoring the development of hydrogen bonds. Since, hydrogen bond plays a significant role in determining the secondary and tertiary structures of proteins, we considered the changes in the atomic charges of atoms involved in backbone hydrogen bonds to provide a more realistic portrayal of backbone hydrogen bonds during the MD simulation. The strategy in the PSBC scheme is similar to the PHB model proposed by Gao et al.¹ in that both are formula-based polarization models. In both models, a formula was derived to account for the charge transferred between backbone hydrogen bond donor (N) and acceptor (O) during the formation of a hydrogen bond. To model the charge transferred between backbone hydrogen bond pairs, dialanine peptides conforming to geometries of backbone hydrogen bonds formed in α -helices and anti-parallel β -sheets were subjected to quantum mechanical (QM) calculations using Gaussian03 at HF/6-31G* level of theory to derive the electrostatic potential of the dialanine peptides. Atomic charges of the backbone hydrogen bond donor and acceptor were subsequently fitted using RESP method during which only backbone atoms namely C, O, N and H were allowed to vary with the rest of dialanine atoms fixed at AMBER94 charges. The QM calculations and RESP fitting processes were repeated with dialanine models with distance between donor (N) and acceptor (O) atoms varied between 2.5 Å and 6.5 Å to acquire a

function that relates the charge transferred between hydrogen bond pairs and the distance between O and N atoms. Only the distance between O and N atoms was varied while other degrees of freedom were kept constant. In a situation where the dialanine pair is far apart, the charges of backbone C, O, N and H atoms of an isolated alanine residue are fitted using the same procedures outlined above with the partnering alanine pair present in the form of ghost atoms at a distance where no hydrogen bonding occurs. This eliminates the presence of background charges hence distinguishing the PSBC model from the PHB model. The difference between the distance-dependent charges of the dialanine model and the charge obtained from the isolated alanine residue was calculated to acquire the charge transferred between hydrogen bond pairs. The charge transferred calculated was plotted against the distance between O and N atoms and the following functional form for α -helix and β -sheet were acquired respectively (Figure 1) where r is the distance between the hydrogen bond donor (N) and acceptor (O). Since only the charges of C, O, N and H atoms are varied, the charge transferred for C and H atoms is negative of the values computed for the O and N atoms respectively.

References

1. Gao, Y.; Lu, X.; Duan, L. L.; Zhang, J. Z. H.; Mei, Y. Polarization of Intraprotein Hydrogen Bond Is Critical to Thermal Stability of Short Helix. *J. Phys. Chem. B* **2012**, *116* (1), 549–554.

PART III

RESULTS AND DISCUSSION

Chapter 3

3. *Ab initio* folding of mixed-fold FSD-EY protein using formula-based polarizable hydrogen bond (PHB) charge model

3.1 Introduction

There have been a number of studies that look at the role of polarizability in protein modelling. These studies have revealed the importance of the polarization effect at stabilizing the native state of the folded proteins, as well as more accurately depicting the backbone hydrogen bond strength that were underestimated by non-polarizable force fields¹⁻⁹. Since the folding of protein secondary structures are highly dependent on the formation of backbone hydrogen bonds^{10,11}, describing the polarization effect during the formation of hydrogen bonds accurately is significant to enhancing the protein folding accuracy during MD simulations. Recently, the polarizable hydrogen bond (PHB) model, which was designed by Gao et al., integrated the polarizability of backbone hydrogen bonds with traditional non-polarizable force fields^{2,8}. The PHB model periodically updates the atomic charges of the participating backbone hydrogen bonds (C, O, N, H) by applying an exponential mathematical function that relates the charge transferred and the hydrogen bond distance during hydrogen bond occurrence⁸. Although the PHB model was able to fold long and short α -helices effectively^{8,9}, there remains a need to

ascertain the ability of the PHB model to fold other secondary structures besides α -helices.

In this chapter, the PHB model is applied to the folding of FSD-EY, a $\beta\beta\alpha$ mixed-fold protein¹². The FSD-EY mini-protein was originated from FSD-1, which was a *in silico* designed protein that was based on the zinc finger subdomain^{12,13}. In order to obtain FSD-EY, there were two mutations (Q1E and I7Y) on FSD-1 that was intended to prevent the cyclization at the N-terminus so as to enhance the stability of the β -turn in the β -sheet region¹³. Given the small protein size of only 28 residues, while containing two types of secondary structures, which are the α -helix and β -sheet, the FSD-EY mixed fold protein is a suitable subject to evaluate the ability of the PHB model in folding the protein from the linear state^{12,13}.

In this study, the FSD-EY protein was folding via replica-exchange molecular dynamics (REMD) simulations. REMD simulation is commonly used in protein folding studies because it renders more efficient sampling of phase space as compared to isothermally conventional MD simulation^{14,15}. By making use of high temperature replicas in REMD simulations, it prevents the system simulated from being trapping in multitudes of local energy minima that occur at low temperatures, therefore attaining a higher probability of reaching folded states^{14,15}. Altogether, there are three sets of REMD simulations conducted with the traditional non-polarizable AMBER ff99SB force field, the PHB model and a scaled-down PHB model^{8,16,17}. The scaled-down PHB model will be termed as PHB0.7 in this chapter, as it represents the reduction in polarizability during the formation of backbone hydrogen bonds to 70% of the

value originally derived by the PHB model. Since the internal dielectric constant of the protein could reach as low as 2, the factor of 0.7 is therefore chosen as a relative approximate reciprocal of the dielectric constant in order to reduce the polarizability manually, so as to observe the effect of this reduced polarizability on the folding of the secondary structures present in FSD-EY. There are no other simulation tests in addition to the 30% reduction in polarizability conducted in this chapter.

3.2 Simulation Details for REMD simulation

In order to evaluate the efficiency of the PHB model in folding the mixed-fold protein, FSD-EY (PDB id: 1FME) was folded from the linear conformation via REMD simulation¹². The linear structure of the FSD-EY protein was generated using the LEaP module in AmberTools 1.2¹⁸. The non-polarizable AMBER ff99SB force field and the generalized Born solvation model (igb=7) was applied to the protein folding and modelling of the solvation effect respectively¹⁷⁻¹⁹. The salt concentration for the REMD simulation was set to 0.1 M. Minimization of the FSD-EY linear structure was conducted prior to the REMD simulation, with the steepest descent method for 5000 steps and was followed by the conjugate gradient method for 5000 steps, in order to eliminate any severe steric clashes. In this chapter, a total of 20 replicas ranging from 250 K to 473 K were used in the REMD simulations. The FSD-EY linear structure was first heated to 250, 259, 268, 277, 286, 296, 306, 317, 328, 339, 350, 362, 375, 387, 400, 414, 428, 443, 458 and 473 K for 1 ns to obtain 20 different starting conformations before equilibration in the REMD simulation. The distribution of the 20 temperatures was pre-determined in

order to make sure that there is a targeted acceptance ratio of 0.2. Regulation of the temperatures during the REMD simulation was done by applying the Langevin thermostat^{20,21} using a collision frequency of 4 ps^{-1} .

Temperature exchange between the 20 replicas was attempted every 10 ps (10000 steps) during the REMD simulation. The integration time step of 1 fs was used. 100 ns of REMD simulation was conducted for each temperature replica. All non-bonded interactions are calculated with the Particle Mesh Ewald method using a cutoff of 12 Å. The SHAKE algorithm was applied to constrain all bonds that involve hydrogen atoms^{22,23}. All the simulation parameters applied for the AMBER ff99SB force field, were also used for the REMD simulations of the PHB and PHB0.7 models. In addition, the SIMULAIID program^{8,24} was employed to determine the atomic charges of the backbone hydrogen bond atoms (C, O, N, H) during the occurrence of backbone hydrogen bonds. Periodic atomic charge updating and checking of backbone hydrogen bonds during the REMD simulation was conducted every 0.5 ps.

Ref 8 contains a detailed description of the PHB model. This chapter will be providing a brief outline of the PHB model protocol. The basis of the PHB model lies with the distortion of the electron clouds when a hydrogen bond is formed, due to a difference in electronegativity between the hydrogen bond acceptors and donors. This distortion reduces the electrostatic energy of the hydrogen bond pair, therefore increasing the hydrogen bond strength. This inclusion of the polarization effect in hydrogen bond pairs may enhance the energetics of protein folding in MD simulations. The derivation of the PHB

model utilizes a pair of dialanine residues confined to an α -helical configuration as a simplified model to represent the state of hydrogen bonding in α -helices. The atomic charges of the backbone hydrogen bond atoms in this dialanine model were fitted using the restrained electrostatic potential (RESP) method²⁵. Iteration of the charge fitting was continued till convergence was achieved. In order to remain consistent with the AMBER ff99SB charges, all quantum-mechanical (QM) calculations were conducted with Gaussian09 at the HF/6-31G* level of theory²⁶. The above-mentioned procedure was repeated with the dialanine model while the hydrogen bond acceptor-donor distance was varied from 2.5 to 6.5 Å. This yielded exponential mathematical functions that relate the charge transferred between the acceptor and donor atoms to the acceptor-donor distance:

$$\Delta q_N = -0.843 \times e^{-0.455(R+1.18)}$$

$$\Delta q_O = -0.250 \times e^{-0.466(R-0.623)}$$

where R is acceptor-donor distance. During the charge fitting procedure using the RESP method, only the atomic charges of the backbone atoms were varied while the rest of the protein atoms were assigned with fixed AMBER ff99SB charges. For the C and H atoms, the charge transferred is the exact negative value of that of the O and N atoms respectively. In the PHB0.7 model, the charge transferred was reduced by 30% of the original value derived by the original PHB model.

3.2.1 Simulation Details for the Minimization of the Folded FSD-EY under

Explicit Solvent condition

The lowest RMSD structure relative to the FSD-EY NMR structure was obtained from the REMD simulation performed under the PHB0.7 model. The octahedral TIP3P water box was used to solvate the obtained protein structure with a minimum distance of 10 Å between the edge of the water box and the protein. Chloride ions were added to the water box for the neutralization of the solvated protein system. Minimization of the solvated protein system was conducted in two steps. During the first step of minimization, the water molecules were minimized using the steepest descent method for 10000 steps, and then followed by the conjugate gradient method for 10000 steps, while the protein was restrained with a harmonic force constant of 500 kcal/mol/Å². During the second step of minimization, the protein-water system was minimized for 5000 steps using the steepest descent method, followed by the conjugate gradient method for 15000 steps. After making sure that there are no severe steric clashes, the entire system underwent heating for 100 ps to 280 K using the Langevin thermostat with a collision frequency of 4 ps⁻¹.^{20,21} During the heating, a weak restraint using a harmonic constant of 5 kcal/mol/Å² was applied to the protein. After heating, the production run for the entire system was performed for 50 ns under the non-polarizable AMBER ff99SB force field.¹⁷⁻¹⁹ The time step for the production run is 1 fs. All non-bonded interactions are calculated with the Particle Mesh Ewald method using a cutoff of 12 Å. The SHAKE algorithm was applied to constrain all bonds that involve hydrogen atoms^{22,23}.

3.2.2 Ascertaining of REMD simulation convergence

One of the aims in this chapter was to evaluate the efficiency of the PHB model in folding FSD-EY to its native conformation from its unfolded linear structure. Before obtaining the folded structure from the REMD simulations, it is important to ascertain that the simulations were converged so to as ensure that folded conformation obtained is in the equilibrium state. One of the methods to ascertain whether the simulations have converged is the plotting of the time-dependent changes in terms of cluster size. For this study, hierarchical cluster analyses²⁴ were conducted in blocks of 10 ns for the 100 ns REMD trajectories obtained. The cluster analyses for the simulations performed with the AMBER ff99SB force field, the PHB and PHB0.7 models are displayed in Figure 2. Figure 2 shows a sharp decrease in the number of clusters, followed by a plateau after 70 ns for all the simulation types, indicating that equilibrium is reached for the FSD-EY folding. In addition, the application of the PHB and PHB0.7 models resulted in an earlier attainment of the equilibrium state than the AMBER ff99SB force field.

Figure 2 Hierarchical cluster analyses of the simulations performed under the (A) PHB, (B) PHB0.7 and (C) AMBER ff99SB force field at 267 K.

3.3 Results and Discussion

Folding of the $\beta\beta\alpha$ mixed-fold moiety is a daunting task in the computational field that many studies have attempted using both REMD and conventional MD simulations²⁸⁻³¹. These studies commonly carried out intricate modifications of the non-polarizable force fields in order to manage the different propensities of the different types of secondary structures that

reside within the same protein²⁸⁻³¹. This chapter explores the folding of FSD-EY via a combination of simulation techniques, which are the REMD simulations under implicit solvation and conventional MD simulations under explicit solvation^{12,30}. Three sets of REMD simulations were conducted using the non-polarizable AMBER force field and the polarizable PHB and PHB0.7 models to explore the protein folding of FSD-EY^{8,16,17}.

3.3.1 Contrasting the ability of PHB, PHB0.7 and AMBER ff99SB charge models in the protein folding of FSD-EY

After ascertaining that the simulations have converged as displayed in Figure 2, the structural similarity between the folded structure and the NMR structure of FSD-EY was observed via the C α -RMSDs distributions (Figure 3) calculated from the protein conformations obtained in the last 50 ns of all three types of REMD simulations. At 267 K and 296 K, the α -helical domain of FSD-EY was observed to have folded effectively for all three types of REMD simulations performed as displayed by the tallest distribution peaks attained at approximately 1 Å for the α -helical domain. While the PHB and PHB0.7 models only displayed one peak, the AMBER ff99SB force field was observed to produce another distribution peak at approximately 3.5 Å, which is plausibly corresponding to a partially folded α -helix in Figure 3C. These observations may indicate the applicability of the PHB and PHB0.7 models in favouring the folding of the α -helical domain, while at the same time demonstrated the tendency of including the polarization effect in enhancing the α -helix stability at higher temperatures as compared to the non-polarizable AMBER force field.

Figure 3 Ca -RMSDs probability distribution curves of the FSD-EY protein folding at 267 K (left) and 296 K (right), obtained from the REMD simulations conducted using the (A) PHB, (B) PHB0.7 and (C) AMBER force field.

Next, the chapter will be examining the folding of the β -sheet region in FSD-EY. As opposed to the observation made for the α -helical domain, the β -sheet region was unable to fold satisfactorily in all the REMD simulations performed. The distribution curves at 267 K for the β -sheet domain were shifted to the right with the peaks at RMSD values of approximately 6 Å, 7 Å and 5 Å for the REMD simulations performed with AMBER, PHB and PHB0.7 atomic charges respectively. Even though the values of the Ca -RMSD for the β -sheet domain

are large, the PHB0.7 model was able to fold the β -sheet domain of FSD-EY slightly better than the PHB model and the AMBER ff99SB force field. Additionally, the distribution curve peak for the simulation conducted under the PHB0.7 model possesses a higher occurrence probability (approximately 0.58) as compared to the other two simulations. Therefore, applying the PHB0.7 model to REMD simulations in the folding of the FSD-EY protein has presented as a plausibly better option since the β -sheet domain was observed to fold better under this model, while at the same time maintain a comparable folding efficiency of the α -helical domain to that of the PHB model. The C α -RMSD distribution curves in Figure 3 displayed a glimpse of the folding efficiency of the three different charge models. The chapter proceeds with more analyses in order to obtain deeper insights on the difference between the three different charge models.

Since all the charge models were able to fold the α -helical region in FSD-EY, the variation in the α -helical content was monitored at 267 K. An amino acid was considered to be in an α -helical conformation if its dihedral angles are within the range of $(\phi, \psi) = (-60 \pm 20, -40 \pm 20)^{31}$. The FSD-EY NMR structure shows that the residues 15 to 24 form the α -helix, indicating the α -helical content of FSD-EY to be approximately 35%¹². When comparing the α -helical content obtained from the three different charge models, the PHB and PHB0.7 models are able to reproduce the α -helical content that resemble closer to the experimental value as compared to the non-polarizable AMBER ff99SB force field (Figure 4A). This observation demonstrates the significance of including polarizability during the folding of α -helices while at the same

time, emphasizes the ability of the PHB models in integrating the polarization effect into the current non-polarizable AMBER force field to enhance the folding of α -helices. The probability distribution curves for the α -helical content were plotted using the last 50 ns of the REMD simulations conducted (Figure 4B). The probability distribution curve of the PHB model was shifted to the right of the experimental value of the NMR structure with a modal value of 0.4 (Figure 4B). This indicates that the PHB model possesses a tendency to overestimate the helicity of the folded α -helical domain. On the other hand, the probability distribution curve of the non-polarizable AMBER force field (green) was shifted to the left of the experimental value of the NMR structure with a modal value of 0.3. Only the PHB0.7 model resulted in a probability distribution curve that is very similar to that of a Gaussian distribution and having a modal value that closely matches the experimental value.

Figure 4 (A) Evolution of the α -helical content over simulation time for the REMD simulation at 267 K of the FSD-EY protein folding conducted with PHB (black), PHB0.7 (red) and the AMBER force field (green). (B) Probability distribution plots of the α -helical content in FSD-EY at 267 K for PHB (black), PHB0.7 (red) and the AMBER force field (green).

The REMD simulation also enables the extraction of the protein melting temperature. In this chapter, the temperature-dependent profiles for the three different types of charge models were constructed in order to perform a comparison for the melting temperature obtained from the different charge models. Figures 3 and 4 have indicated that the α -helical region in FSD-EY was folded successfully, therefore the fraction of the folded α -helical conformation at different temperatures was calculated (Figure 5). At approximately 380 K, the α -helix was completely unfolded for all the three different charge models. For the non-polarizable AMBER force field, the α -helical domain in FSD-EY was observed to have the highest rate of unfolding. This demonstrated that the non-polarizable AMBER force field was unable to stabilize the folded α -helix as strong as the PHB and PHB0.7 models. Comparing the curves in Figure 5, the PHB model displayed the highest stability for the folding of the α -helical region as indicated by the higher temperature that corresponded to the fraction of the helix folded at 50% i.e. the simulated melting temperature. The critical transition temperature of the α -helical region in FSD-1 was experimentally determined to be 314 K by Feng et al. Since the Q1E and I7Y mutations that resulted in FSD-EY increase the protein stability, the α -helical region in FSD-EY is predicted to be at a higher temperature than 314 K. The extra stabilization brought about by the mutations is portrayed in the REMD simulation performed under the PHB model, where the simulated melting temperature is higher than 320 K. On the other hand, the simulated melting temperature for the PHB0.7 model resides closely to 314 K.

Figure 5 Temperature-dependent profiles for REMD simulations performed under PHB (black), PHB0.7 (red) and AMBER (green) charges.

After analysing the different secondary structure domains in FSD-EY, the following analysis examines the protein folding pathway of FSD-EY during the REMD simulation performed under the three different charge models. Kim et al. performed an *ab initio* folding simulation of the FSD-EY mixed-fold protein under implicit solvation via the multiplexed Q-REMD simulation and has proposed a plausible protein folding pathway for FSD-EY³⁰. In the study, it was suggested that FSD-EY folds via an initial formation of a loop structure, followed by the eventual folding of the α -helix and the partial aggregation of the hydrophobic core³⁰. The protein folding was completed after the hydrophobic core has assembled via the folding of the β -sheet domain³⁰.

Therefore, to decipher the protein folding pathway of FSD-EY derived from the three types of REMD simulations performed in this chapter, 2-dimensional free energy landscapes (2D-FELs) were plotted to compare and contrast the differences in the protein folding pathways via the weighted histogram analysis method^{33,34}. (Figure 6) The reaction coordinates employed

in the construction of the 2D-FELs are the number of backbone hydrogen bonds formed in the α -helix and the center-of-mass (COM) distance between the Y7 and F12 side chains. This distance represented the distance between the two strands of the β -sheet and therefore was measured to observe the closeness of the two strands of the β -sheet during the REMD simulations. (Figure 6D) The PHB0.7 model resulted in a folding conformation that bore resemblance to the native state as displayed at Region 4 (Figure 6B). Examining the 2D-FELs in Figure 6 indicated that sampling of the region around (Hb Count, Dist) = (8.0, 8.0) only existed when the polarization effect was included in the backbone hydrogen bonds formed during the REMD simulation. In addition, a larger population of at the region (Hb Count, Dist) = (8.0, 8.0) was observed for the PHB0.7 charge model (Region 4, Figure 6B). This further supports the hypothesis that the reduction in the polarization effect of the charge model may help in increasing the mixed-fold protein folding efficiency. The PHB model resulted in a significant population at (Hb Count, Dist) = (8.0, 12.9) (Region 4) that represents the lowest energy state attained for the FSD-EY protein folding performed under the PHB model. (Figure 6A) This observation supported the results obtained so far, which highlighted a tendency of the PHB model to over favour the folding of a stable α -helix while disfavours the β -sheet folding.

Figure 6 2D-FELs of FSD-EY obtained for the (A) PHB model, (B) the PHB0.7 model and (C) the non-polarizable AMBER ff99SB force field. The reaction coordinates used are the number of backbone hydrogen bonds (Hb Count) and the COM distance of the Y7 and F12 side chains. The represented structures for the regions 1-4 are shown on the right of the 2D-FELs in (A) to (C). (D) A schematic illustration of the backbone hydrogen bonds that are formed between the i and $i + 4$ residues in the α -helical region of FSD-EY and the location of the Y7 and F12 residues in the β -sheet domain.

3.3.2 FSD-EY folding via conventional MD simulation under explicit solvent condition

From the results obtained from the various analyses so far, the PHB0.7 charge model has displayed its ability to fold a conformation that resembles the closest to the FSD-EY native structure as compared to the PHB model and the

non-polarizable AMBER force field. However, the inability of the various charge models to fold the β -sheet domain effectively, resulting in the incomplete assembly of the hydrophobic core, remains unsatisfactory. Therefore, since the representative structure obtained from Region 4 (Figure 6B) under the PHB0.7 model bears the greatest resemblance to the NMR structure, the structure was further subjected to conventional MD simulations under the explicit solvent condition for 50 ns to probe any possibility of folding the β -sheet domain closer to the native conformation.

To make sure that the FSD-EY protein is folded to its native state, the RMSD value referencing the FSD-EY NMR structure was calculated. (Figure 7A) In the first 30 ns of the conventional MD simulation, the RMSD of the full FSD-EY protein experienced drastic fluctuations due to the significant structural changes occurring at the β -sheet domain of FSD-EY. Cluster analyses conducted on the trajectory in 15 ns blocks for the first 30 ns of the conventional MD simulation supported the drastic conformational changes that occurred at the β -sheet domain of FSD-EY. The representative structures obtained from the cluster analyses demonstrated the high β -sheet flexibility in contrast to the α -helix which remained folded. (Figure 7B) At approximately 30 ns, the RMSD value of the β -sheet domain was observed to stabilize at around 2 Å, also indicating that the RMSD of the full FSD-EY protein has stabilized. Cluster analysis performed in the last 20 ns of the conventional MD simulation resulted in representative folded structures that possess both the α -helix and the β -sheet moieties. (Figure 7B) Even though the conventional MD simulation was conducted under explicit solvent condition in contrast to the REMD simulations performed under implicit solvent condition, little fluctuation is

observed in the RMSD of the α -helix, indicating that the conformation integrity (RMSD < 2 Å) was retained during the conventional MD simulation. This supported the proposed protein folding pathway of the FSD-EY mixed-fold protein, where the α -helical region completes its folding before proceeding to complete the folding of the β -sheet region. At the same time, the conventional MD simulation provided insights on the role of water molecules in promoting and aiding the folding of the β -sheet region and finally resulting in the complete packing of the hydrophobic core.

Figure 7 Folding of the FSD-EY mixed-fold protein under explicit solvent condition: (A) Evolution of the RMSD value for the full protein (black), α -helix (red) and β -sheet (blue) with reference to the FSD-EY NMR structure. (B) Representative structures obtained from the cluster analysis of the conventional MD simulation at different periods of the simulation. The

percentages shown below the structures represents the probability of the structures occurring during that period of the MD simulation.

3.4 Chapter Summary

Gao et al. designed the PHB model with the intention to incorporate polarizability into backbone hydrogen bonds in traditional non-polarizable force fields, via exponential mathematical forms that relate the amount of charge transferred between the acceptor and donor atoms to the acceptor-donor distance⁸. This model successfully folded α -helices of various amino acid lengths^{8,9}. Hence, to determine the ability of the PHB model in folding proteins that possess other types of secondary structures, the model was applied to the folding of the mixed-fold $\beta\beta\alpha$ protein, FSD-EY, from its linear structure. In this chapter, a total of three REMD simulations were performed under the PHB model, the PHB0.7 model and the non-polarizable AMBER ff99SB force field to fold FSD-EY. The α -helical domain was observed to have folded successfully in all the three simulations, in contrast to the β -sheet domain. Therefore, a representative structure that bears the closest resemblance to the FSD-EY NMR structure was obtained from the REMD simulation performed with the PHB0.7 model. Minimization of the protein-water system was conducted, and followed by 50 ns of conventional MD simulation. The resulting folded structure had a lowest RMSD value of 1.29 Å with reference to the NMR structure, indicating the successful folding of the FSD-EY protein. Although the combination of REMD simulations and conventional MD simulation resulted in the successful folding of the mixed-fold protein, the inability of the PHB model to fold the β -sheet region indicates the importance for further refinement of the PHB model to increase its versatility for α -helices and β -sheets. As demonstrated in this chapter, better protein folding of FSD-EY was observed

when there was a reduction of the polarizability in the PHB model that resulted in the PHB0.7 model. However, this is not a flawless solution and prompted the construction of the Polarized Structure-specific Backbone Charge (PSBC) model to cater to both α -helices and β -sheets.

References

1. Ji, C.; Mei, Y. Some Practical Approaches to Treating Electrostatic Polarization of Proteins. *Acc. Chem. Res.***2014**, *47* (9), 2795–2803.
2. Duan, L. L.; Mei, Y.; Zhang, D.; Zhang, Q. G.; Zhang, J. Z. H. Folding of a Helix at Room Temperature Is Critically Aided by Electrostatic Polarization of Intraprotein Hydrogen Bonds. *J. Am. Chem. Soc.***2010**, *132* (32), 11159–11164.
3. Tong, Y.; Ji, C. G.; Mei, Y.; Zhang, J. Z. H. Simulation of NMR Data Reveals That Proteins' Local Structures Are Stabilized by Electronic Polarization. *J. Am. Chem. Soc.***2009**, *131* (24), 8636–8641.
4. Duan, L. L.; Mei, Y.; Zhang, Q. G.; Zhang, J. Z. H. Intra-Protein Hydrogen Bonding Is Dynamically Stabilized by Electronic Polarization. *J. Chem. Phys.***2009**, *130* (11), 115102.
5. Xu, Z.; Mei, Y.; Duan, L.; Zhang, D. Hydrogen Bonds Rebuilt by Polarized Protein-Specific Charges. *Chem. Phys. Lett.***2010**, *495* (1–3), 151–154.
6. Xu, Z.; Lazim, R.; Mei, Y.; Zhang, D. Stability of the β -Structure in Prion Protein: A Molecular Dynamics Study Based on Polarized Force Field. *Chem. Phys. Lett.***2012**, *539–540*, 239–244.
7. Lazim, R.; Wei, C.; Sun, T.; Zhang, D. Ab Initio Folding of Extended α -Helix: A Theoretical Study about the Role of Electrostatic Polarization in

- the Folding of Helical Structures. *Proteins Struct. Funct. Bioinforma.***2013**, *81* (9), 1610–1620.
8. Gao, Y.; Lu, X.; Duan, L. L.; Zhang, J. Z. H.; Mei, Y. Polarization of Intraprotein Hydrogen Bond Is Critical to Thermal Stability of Short Helix. *J. Phys. Chem. B***2012**, *116* (1), 549–554.
 9. Gao, Y.; Lu, X.; Duan, L.; Zhang, D.; Mei, Y.; Zhang, J. Z. H. Direct Folding Simulation of a Long Helix in Explicit Water. *Appl. Phys. Lett.***2013**, *102* (19), 193706.
 10. Lodish, H., Berk, A., Zipursky, S. L., Matsudaira P., Baltimore D., Darnell J. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; **2000**. Section 3.1, Hierarchical Structure of Proteins. Available from <http://www.ncbi.nlm.nih.gov.ezlibproxy1.ntu.edu.sg/books/NBK21581/>.
 11. Vogt, G.; Woell, S.; Argos, P. Protein Thermal Stability, Hydrogen Bonds, and Ion Pairs. *J. Mol. Biol.***1997**, *269* (4), 631–643.
 12. Sarisky, C. A.; Mayo, S. L. The B β α Fold: Explorations in Sequence. *J. Mol. Biol.***2001**, *307* (5), 1411–1418.
 13. Kim, E.; Jang, S.; Pak, Y. All-Atom Ab Initio Native Structure Prediction of a Mixed Fold (1FME): A Comparison of Structural and Folding Characteristics of Various B β α Miniproteins. *J. Chem. Phys.***2009**, *131* (19), 195102.
 14. Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.***1999**, *314* (1–2), 141–151.

15. Nymeyer, H. How Efficient Is Replica Exchange Molecular Dynamics? An Analytic Approach. *J. Chem. Theory Comput.***2008**, 4 (4), 626–636.
16. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Bioinforma.***2006**, 65 (3), 712–725.
17. Wickstrom, L.; Okur, A.; Simmerling, C. Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling Data. *Biophys. J.***2009**, 97 (3), 853–856.
18. D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, P.A. Kollman **Amber 10** University of California, San Francisco (2008)
19. Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.***2007**, 3 (1), 156–169.
20. Pastor, R. W.; Brooks, B. R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.***1988**, 65 (6), 1409–1419.
21. Uberuaga, B. P.; Anghel, M.; Voter, A. F. Synchronization of Trajectories in Canonical Molecular-Dynamics Simulations: Observation,

- Explanation, and Exploitation. *J. Chem. Phys.***2004**, *120* (14), 6363–6374.
22. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.***1993**, *98* (12), 10089–10092.
23. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. . Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.***1977**, *23* (3), 327–341.
24. Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3* (6), 2312–2334.
25. Mezei, M. Simulaid: A Simulation Facilitator and Analysis Program. *J. Comput. Chem.***2010**, *31* (14), 2658–2668.
26. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.***1993**, *115* (21), 9620–9631.
27. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A. et al., Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford CT, **2010**.

28. Mohanty, S.; Hansmann, U. H. E. Folding of a Miniprotein with Mixed Fold. *J. Chem. Phys.***2007**, *127* (3), 35102.
29. Lei, H.; Wang, Z.-X.; Wu, C.; Duan, Y. Dual Folding Pathways of an A/ β Protein from All-Atom Ab Initio Folding Simulations. *J. Chem. Phys.***2009**, *131* (16), 165105.
30. Kim, E.; Jang, S.; Pak, Y. All-Atom Ab Initio Native Structure Prediction of a Mixed Fold (1FME): A Comparison of Structural and Folding Characteristics of Various B β α Miniproteins. *J. Chem. Phys.***2009**, *131* (19), 195102.
31. Lee, I.-H.; Kim, S.-Y.; Lee, J. Folding Models of Mini-Protein FSD-1. *J. Phys. Chem. B***2012**, *116* (23), 6916–6922.
32. Vila, J.; Williams, R. L.; Grant, J. A.; Wójcik, J.; Scheraga, H. A. The Intrinsic Helix-Forming Tendency of L-Alanine. *Proc. Natl. Acad. Sci. U. S. A.***1992**, *89* (16), 7821–7825.
33. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.***1992**, *13* (8), 1011–1021.
34. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.***1995**, *16* (11), 1339–1350.

Chapter 4

4. An initial application of the PSBC model on the folding of simple β -sheets –

Tryptophan zippers 1 and 2

4.1 Introduction

The protein structure directly correlates with its biological function¹⁻⁴, therefore it is essential to decipher the protein folding process, so as to understand where are the areas during the process that can be engineered in order to ensure a higher success rate of folding to the desired protein structure.

The protein folding process was initially thought to be a very long process where the protein or peptide has to sample all conformations possible within the configurational space. However, Levinthal's paradox found that it is impossible in a physically meaningful time for a peptide to fold to the native structure if it were to sample every conformation possible.⁵⁻⁸ In addition, studies have shown that the folding rates range from microseconds to an hour, depending on the protein size.⁹⁻¹¹ Therefore, there must exist some governing rules that will enable to peptide to fold to its native structure in a very short time without sampling every conformation possible.

Via studies such as alanine or glycine mutagenesis, it is found that energetics of the conformations and intermolecular interactions play a major role in guiding the peptide to fold to its native structure.¹²⁻¹⁴ Hydrogen bonds, being one of the stronger intermolecular interactions, is one of the key factors to the folding of secondary structures such as α -helices and β -sheets.¹⁵⁻¹⁷

While in vitro studies are able to distinguish the important amino acids within the peptide that determine its successful folding to the native structure, it is unable to illustrate the intermolecular interactions between them explicitly. Therefore, molecular dynamics (MD) simulations are employed to serve this purpose.

Levinthal's paradox inspired further studies on the protein folding process, leading to the understanding that energetically favorable folding intermediates are key to guiding the peptide to fold to its native structure successfully.⁵⁻⁸ Since these intermediates exist dynamically during the folding process, the MD simulation will be able to capture this dynamicity and reveal the folding intermediates.

MD simulations are a computational tool that make use of force fields and algorithms to simulate chemical processes at the atomic level. Outputs of these simulations are then translated to meaningful physical properties to identify molecular reasons in order to explain the experimental observations.

Force fields are a set of energy functions and parameters sets that MD simulations use to calculate the potential energies of a system of atoms or coarse-grained particles. Employing MD simulations to fold a peptide to its native structure relies heavily on the accuracy of force fields.

Hydrogen bonds are a key factor to folding secondary structures. Accurate calculation of the hydrogen bond potential energy in force fields greatly enhances the success rate of protein folding. The textbook definition of a hydrogen bond is an interaction between an electronegative atom and a hydrogen attached to another electronegative atom. However, *ab initio*

studies have shown that when atoms of different electronegativities approach each other in close proximity, each atom will distort the electron cloud of the partnering atom. This is known as the polarization effect.¹⁸⁻¹⁹

In popular force fields such as CHARMM, GROMOS and AMBER, the hydrogen bond energy is calculated by employing fixed atomic charges throughout the simulation. Therefore, the force fields do not account for the polarization effect explicitly and this greatly affects the total hydrogen bond energy of the peptide during the folding process in the MD simulation.²⁰⁻²²

Since then, there have been an increasing number of research in developing novel methods of incorporating the polarization effect in force fields. However, there were a few more problems. One, to incorporate the polarization effect between hydrogen bond pairs, *ab initio* methods have to be employed. While a single hydrogen bond does not utilize hefty computational resources when calculating the atoms' electronegativities with the polarization effect taken into account, a peptide with approximate 20 to 30 amino acid residues would require either long calculation time or a significant number of computational resources. Two, because atoms are constantly in motion during the MD simulation, there is a need to dynamically incorporate the polarization effect at every frame of the simulation. When running long time-scale simulations, dynamic *ab initio* calculations would also consume hefty computational resources.

In order to resolve the above-mentioned problems, we developed a novel dynamic charge update method that incorporated the polarization effect

via *ab initio* calculations, known as the Polarized Structure-Specific Backbone Charge (PSBC) model.

First, we identified specifically the protein backbone atoms (N, H, C, O) as the subject of our model because the stability of secondary protein structures is heavily dependent on the strength of backbone hydrogen bonds. Second, we perform *ab initio* calculations on the different backbone hydrogen bond lengths to incorporate the polarization effect on the atoms' electronegativities. Third, a trend was obtained and a polynomial function was fitted to it. This function is then incorporated in the AMBER force field to update the backbone atomic charges dynamically at every frame of the simulation.

There have been many studies of novel or modified force fields, with the majority employing α -helices as test subjects.²³⁻²⁷ Since hydrogen bond networks are weaker in β -sheets than in α -helices, we decided to test the PSBC model vigorously to the folding of tryptophan zippers 1 and 2 (Trpzips 1 and 2).

The trpzip peptides were designed based on the C-terminal of the immunoglobulin binding domain of streptococcal protein G (GB1).²⁸ While their sequences only differ at the turn moiety, this observable difference may have seemingly attributed to the difference in melting temperatures. However, there were little studies that extensively discuss about the different factors that result in the melting temperature difference.

On the other hand, there have been at least two putative folding mechanisms proposed, namely the zipper^{29,30} and hydrophobic collapse

models^{31,32}. In addition to these two models, a mis-registered folding state with non-native backbone hydrogen bonds was also found to rearrange to the native state by a sliding motion along the peptide backbone³³.

Therefore, successful folding of the trpzip peptides with the PSBC model will not only further emphasize the importance of considering the polarization effect in backbone hydrogen bonds, but the folding process will be able to provide extensive insights on possible factors that affect the melting temperatures at the atomic level.

4.2 Simulation Details

Trpzip1		Trpzip2	
GN		NG	
E	K	E	K
W	W	W	W
T	T	T	T
W	W	W	W
S	K	S	K

Figure 8 Amino acid sequences for tryptophan zippers 1 (left) and 2(right). The turn residues are highlighted in red for trpzip1 and green for trpzip 2.

To observe the ability of the PSBC model to aid in the folding of trpzips 1 and 2, REMD simulation was performed to fold trpzips 1 (PDB id: 1LE0) and 2 (PDB id: 1LE1) from its linear conformation³⁴. The linear structure of FSD-EY was build using LEaP module in AmberTools 1.2³⁵. AMBER ff99SB and generalized Born solvation model (igb=5) by Onufriev et al. was used to model the protein and solvation effect respectively³⁵⁻³⁷. Salt concentration of the system was set to 0.01 M. Before conducting the REMD simulation,

minimization of the linear structure of trpzips 1 and 2 was performed using the steepest descent method for 10000 steps and further relaxed for another 10000 steps using the conjugate gradient method to remove severe steric clashes. For the REMD simulation performed in this study, 16 replicas were used with temperatures ranging from 260 K to 412 K. The protein system was initially heated to 260.00, 267.78, 275.84, 284.20, 292.85, 301.81, 311.10, 320.66, 330.64, 340.98, 351.69, 362.80, 374.05, 385.94, 398.27, 411.04 K in 500 ps prior to the start of the REMD simulation to obtain 16 different starting replicas. The temperature distribution was pre-determined to ensure a targeted acceptance ratio of approximately 0.2. Langevin thermostat with a collision frequency of 4 ps^{-1} was used to regulate the temperature during the simulation^{38,39}.

During the REMD simulation, exchange between replicas was attempted every 10,000 steps (10 ps). An integration time step of 1 fs was used. Each REMD simulation conducted lasted for 100 ns. All non-bonded interactions are calculated with the Particle Mesh Ewald method using an infinitely long cutoff distance and the SHAKE algorithm was utilized to constrain all bonds involving hydrogen atoms^{40,41}. For REMD simulations conducted using the PSBC model, all parameters of AMBER ff9SB were retained except for the atomic charges of backbone C, O, N and H atoms of hydrogen bonded residues that were determined dynamically during the simulations using the PSBC model. Hydrogen bond checking and charge updating were periodically conducted every 0.5 ps during the course of the REMD simulation.

4.3 Results and Discussion

4.3.1 Comparison of the quality of MD simulations between the PSBC model and the traditional AMBER force field

	Trpzip1	Trpzip2
Melting temperature (Experimental)	323.0 +/- 0.3 K	345.0 +/- 0.1 K
Melting temperature (Simulated)	319.14 K	338 K
Difference	3.86 K	7 K

Figure 9 Simulated melting curves of trpzip1 and 2 (top). Melting temperature values of trpzip1 and 2 (bottom).

A physical quantity that differentiates the stability between the trpzip1 and 2 is the melting temperature, which is defined as the temperature where the peptide is 50 % folded. The stability of trpzip1 is lower than trpzip2, as represented by their experimental melting temperatures (Figure 9, bottom). Incorporation of the polarization effect via the PSBC model is able to closely reproduce the experimental melting temperatures of the trpzip1 and 2, and more importantly, reproduce the melting temperature trend.

Figure 10 Backbone (top) and tryptophan side-chains (bottom) root-mean-square deviations of trpzip1 and 2 in the MD simulations.

One of the most common and early analyses is measuring the root-mean-square deviations (RMSD) of the simulated peptide structure against the crystal or NMR structure obtained from PDB. This analysis provides an initial quality measurement of the simulated peptide structure. Commonly, an RMSD value of 3 Å and below is regarded as the MD simulation being able to reproduce the folded structure faithfully to a large extent. Therefore, a backbone RMSD analysis was conducted for both trpzip1 and the average RMSD value taken every 2.5 ns was plotted (Figure 10, top). While the traditional AMBER force field is able to fold the trpzip1 to a low backbone RMSD value, the incorporation of the PSBC model into the AMBER force field has significantly improved the trpzip1 folding and further reduced the backbone RMSD values.

A low backbone RMSD value indicates that the simulation is able to fold the peptide from its linear state to the native secondary structure. However, studies have shown that one of the stabilizing factors is attributed to the π - π interactions between the tryptophan side-chains within the trpzips. Therefore, a tryptophan side-chain RMSD analysis was also conducted to observe the reproduction of the π - π interactions by the PSBC model (Figure 10, bottom). From the analysis, AMBER is unable to reproduce the orientation of the tryptophan side-chains very well as the average RMSD values vary between 5 to 8 Å for trpzip1 and 6 to 7 Å for trpzip2. The PSBC model, on the other hand, is able to reduce the average RMSD value to approximately 4 Å for trpzip1 and 3 to 4 Å for trpzip2. This serves as an initial indication that not only is the PSBC model able to fold the weakly hydrogen-bonded trpzip β -sheet structure, but it is only able to reproduce the π - π interactions present much better than the traditional AMBER force field.

Figure 11 Free energy landscapes of the trpzips using traditional AMBER (top) and PSBC-incorporated AMBER (bottom) force fields. Energy units are in kcal/mol.

Besides being able to reproduce the native structure to a low RMSD value, another feature of a good charge update model is the ability to reduce the free energy basin to one specific area because the native structure is regarded as the lowest energy structure during the protein folding process. Many studies have commonly constructed the free energy landscapes (FELs) with the radius of gyration (Rg) and backbone RMSD as reaction coordinates. Similarly in this study, we have also employed these two variables as reaction coordinates to construct the free energy landscapes for the trpzip peptides (Figure 11).

There are a few notable points from the FELs. One, for both types of AMBER force fields, they are unable to isolate a single lowest energy basin. However, the lowest energy basins for the PSBC model is smaller and more isolated as compared to the traditional AMBER force field. For trpzip1, there is a lowest energy basin at the lowest Rg and backbone RMSD values. For trpzip2, the PSBC model is able to produce a lowest energy basin at the lowest Rg and backbone RMSD values whereas the traditional AMBER force field is unable to do so. Two, the FELs of the trpzips under the traditional AMBER force field has a very large area of high energy intermediates as indicated by the regions coloured in green. These areas are significantly reduced when subjected to the PSBC model. While the PSBC model is not able to reduce the free energy basin to only one specific area, compared to the traditional

AMBER force field, it is able to more effectively fold the trpzip to their native structures by reducing the configurational space of high-energy intermediates and directing the folding process to the lower energy native-like structures.

Figure 12 Lowest energy and lowest RMSD structures of the trpzip under the PSBC model and the traditional AMBER force field.

From the previous analyses, low RMSD values have indicated that the simulation is able to fold the trpzip peptides close to the native structure and FELs have indicated the isolation of energy basins at lowest Rg and backbone RMSD values to contain the native structure. So, the lowest energy and RMSD structures are extracted from the simulation trajectories to provide a visualization of how close is the simulated folded structure to the experimentally obtained native structure (Figure 12).

Although the RMSD values and the FELs of trpzip1 are very close between the PSBC model and the traditional AMBER force fields, the extracted structures painted a more distinct difference between the two types of force fields. Although the lowest RMSD structures of trpzip1 are very similar to the NMR structure, AMBER is unable to reproduce a lowest energy structure that is very close to the native structure. The inaccuracy of the traditional AMBER force field becomes more obvious in the folding of trpzip2. Not only is AMBER unable to isolate a single lowest energy structure, but one of its structures possess a 3_{10} -helical moiety, which is an entirely different secondary structure from a β -sheet. The PSBC model, on the other hand, has produced a lowest energy structure that resembles a mis-registered state.

All these analyses have strongly indicated that the incorporation of the PSBC model in traditional force fields such as AMBER, which treats the hydrogen bond as simply an electrostatic interaction with fixed atomic charges, greatly improves the folding accuracy of peptide sequences in MD simulations. In addition, the PSBC model also circumvents the problem of doing computationally expensive *ab initio* calculations at every time-step of the MD simulation by employing simple polynomial functions that are able to display the effects of polarization on the atoms' electronegativities and hydrogen bond energy.

4.3.2 Folding mechanisms of trpzip1 & 2

Figure 13 DSSP diagrams of trpzip1 and 2 that are obtained from the lowest temperature replica trajectory.

There are quite a number of folding mechanisms proposed for trpzip1 and 2. As mentioned earlier, these mechanisms are the zipper model, the hydrophobic collapse model and the sliding of the mis-registered state to the native structure. The zipper model assumes that the folding of trpzip starts from the turn and propagates toward the tails by forming hydrogen bonds sequentially and the hydrophobic cluster forms later. This folding mechanism is supported by subsequent experiments and some simulations.^{42,43} The hydrophobic collapse model assumes that the hydrophobic core forms first and then the structure propagates from there. This mechanism was supported by most simulation studies using different methods.⁴⁴⁻⁴⁹ There is also a proposed reptation-like mechanism for β -hairpin folding, where a misregistered state with non-native backbone hydrogen bonds rearranges to the native structure by a sliding motion of the backbone along its length.⁵⁰ The reptation mechanism was first identified in simulation studies using a simplified model of the protein.

The dictionary of secondary structure of proteins (DSSP) analysis allows the depiction of folding events that occur during the course of the MD simulation. We conducted the DSSP analysis on both trpzip2 to identify if there is a common dominant folding mechanism that governs the folding process, which may provide first-hand insights on the difference in their stabilities (Figure 13). Residues 1 and 14 are not shown on the DSSP diagrams because they are acetyl and N-methylamide caps. From the DSSP diagrams, the β -sheet formation appears to initiate from the peptide caps, indicating clearly the appearance of frequent mis-registered states, before ending the formation at the hairpin turn. Although there are frequent occurrences of mis-registered states, the higher colour density between the peptide amino acid residues indicate that backbone hydrogen bonds remain as the dominant hydrogen bonds in the β -sheet formation. Our simulations have shown no signs of the zipper model where the β -sheet formation initiates from the hairpin turn.

Figure 14 (a) Hydrogen bond pairs found in NMR structures of trpzip 1 and 2. (b) Percentages of frames where the hydrogen bond lengths are less than or equal to 3 Å for trpzip 1 and 2 in the last 12.5 ns of the simulations.

From the NMR structures of trpzip 1 and 2, multiple hydrogen bond pairs are observed (Figure 14a). At first glance, the presence of an additional hydrogen bond from the NMR structures appear to correlate with the extra stability in trpzip 2 that resulted in a higher melting temperature. The MD simulations results, however, shows otherwise. While the 4 hydrogen bond pairs that are present along the β -sheet are well-maintained in the simulations, the extra hydrogen bond unexpectedly also achieved similar stability during the simulations.

There are a few reasons that may have attributed to this observation. One, NMR structures are static snapshots of the peptide, therefore it is possible that insufficient time was allowed for the trpzip to fold completely. Two, the hairpin turn is a very flexible secondary structure moiety, so dynamic conformational changes may have resulted in transient hydrogen bonds. This accounts for the missing 6O-9H hydrogen bond in the trpzip 1 NMR structure and the lower percentage occupancy of the 6O-9H hydrogen bond observed in simulations.

Figure 15 Normalized frequency diagrams of each amino acid residue's ϕ and ψ angles in trpzips 1 and 2 simulated using traditional AMBER force field (black) and PSBC model (red), compared with NMR structure.

An observable difference between the sequences of trpzips 1 and 2 is the position of the turn residues. This difference appears to be correlated with the difference in melting temperatures. From the histograms of phi and psi angle frequencies of the turn residues (Figure 15), the simulations are unable to produce a single peak that matches the angles of the NMR structure. Even though multiple peaks are observed, only some peaks match the NMR structure, indicating that even with the inclusion of the polarization effect, it remains inadequate in simulating the phi and psi angles. However, for every peak that matches the NMR structure, the PSBC model is able to produce a higher probability of the angle than the AMBER force field. This shows that not only is the inclusion of polarization important in folding the peptide successfully and therefore obtaining a probability of the phi and psi angles, but the PSBC model is able to accomplish the same feat without the requirement of computationally expensive quantum-mechanical calculations throughout the whole simulation.

4.4 Chapter Summary

Deciphering the protein folding process provides valuable insights to understanding the biological function of the protein at the atomic level. A single interaction may be the key to determining the success or failure of a protein to fold to its native state successfully.

While many experimental methods have been developed to understand the protein folding process, they are not without their limitations, with the greatest limitation being unable to visualize which are the interactions and intermediates that exist intermittently during the folding process. MD simulations serve the important purpose to present the dynamicity of the protein folding process as well as the ability to analyze the interactions and intermediates and observe their occurrence throughout the simulation.

The accuracy of MD simulations relies heavily on force fields, which are a set of energy functions and parameter sets to calculate the potential energies of the system simulated. However, popular force fields treat the hydrogen bond energy as a simple electrostatic equation that employs only fixed atomic charges, neglecting the inclusion of the polarization effect.

Therefore, we develop the PSBC model to address the following concerns:

1. Incorporation of the polarization effect in MD simulations
2. To update the atomic charges dynamically throughout the MD simulations

3. To maintain accuracy of the atomic charges without requiring computationally expensive quantum-mechanical calculations during the MD simulations

The successful folding of the trpzip peptides with the PSBC model emphasizes the importance of considering the polarization effect in backbone hydrogen bonds. While more rigorous benchmarking is required to improve the model, the PSBC model represents a possibility for the inclusion of the polarization effect in MD simulations to increase accuracy yet at the same time, computationally affordable.

References

1. Ouzounis, C. A.; Coulson, R. M. R.; Enright, A. J.; Kunin, V.; Pereira-Leal, J. B. Classification Schemes for Protein Structure and Function. *Nat. Rev. Genet.* **2003**, 4 (7), 508–519.
2. Lee, D.; Redfern, O.; Orengo, C. Predicting Protein Function from Sequence and Structure. *Nat. Rev. Mol. Cell Biol.* **2007**, 8 (12), 995–1005.
3. Bradford, J. R.; Siepen, J. A.; Westhead, D. R. Fundamentals of Protein Structure and Function. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*; John Wiley & Sons, Ltd: Chichester, UK, **2005**.
4. McLachlan, A. D. Protein Structure and Function. *Annu. Rev. Phys. Chem.* **1972**, 23 (1), 165–192.

5. Zwanzig, R.; Szabo, A.; Bagchi, B. Levinthal's Paradox. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (1), 20–22.
6. Martínez, L. Introducing the Levinthal's Protein Folding Paradox and Its Solution. *J. Chem. Educ.* **2014**, *91* (11), 1918–1923.
7. Karplus, M. The Levinthal Paradox: Yesterday and Today. *Fold. Des.* **1997**, *2*, S69–S75.
8. Dill, K. A.; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat. Struct. Mol. Biol.* **1997**, *4* (1), 10–19.
9. De Sancho, D.; Doshi, U.; Muñoz, V. Protein Folding Rates and Stability: How Much Is There Beyond Size? *J. Am. Chem. Soc.* **2009**, *131* (6), 2074–2075.
10. Ivankov, D. N.; Garbuzynskiy, S. O.; Alm, E.; Plaxco, K. W.; Baker, D.; Finkelstein, A. V. Contact Order Revisited: Influence of Protein Size on the Folding Rate. *Protein Sci.* **2003**, *12* (9), 2057–2062.
11. Finkelstein, A. V.; Bogatyreva, N. S.; Garbuzynskiy, S. O. Restrictions to Protein Folding Determined by the Protein Size. *FEBS Lett.* **2013**, *587* (13), 1884–1890.
12. Gregoret, L. M.; Sauer, R. T. Tolerance of a Protein Helix to Multiple Alanine and Valine Substitutions. *Fold. Des.* **1998**, *3* (2), 119–126.
13. Kim, K.; Frieden, C. Turn Scanning by Site-Directed Mutagenesis: Application to the Protein Folding Problem Using the Intestinal Fatty Acid Binding Protein. *Protein Sci.* **1998**, *7* (8), 1821–1828.
14. Baase, W. A.; Eriksson, A. E.; Zhang, X.-J.; Heinz, D. W.; Sauer, U.; Blaber, M.; Baldwin, E. P.; Wozniak, J. A.; Matthews, B. W. Dissection of

- Protein Structure and Folding by Directed Mutagenesis. *Faraday Discuss.***1992**, *93*, 173.
15. Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry***1990**, *29* (31), 7133–7155.
 16. Bowie, J. U. Membrane Protein Folding: How Important Are Hydrogen Bonds? *Curr. Opin. Struct. Biol.***2011**, *21* (1), 42–49.
 17. Rose, G. D.; Wolfenden, R. Hydrogen Bonding, Hydrophobicity, Packing, and Protein Folding. *Annu. Rev. Biophys. Biomol. Struct.***1993**, *22* (1), 381–415.
 18. Soto, P.; Mark, A. E. The Effect of the Neglect of Electronic Polarization in Peptide Folding Simulations. *J. Phys. Chem. B***2002**, *106* (49), 12830–12833.
 19. Nilar, S. H.; Pluta, T. S. Hydrogen-Bonding Capabilities Based on Polarizability Studies of Model Peptide Systems. *J. Am. Chem. Soc.***1995**, *117* (50), 12603–12607.
 20. Ji, C.; Mei, Y. Some Practical Approaches to Treating Electrostatic Polarization of Proteins. *Acc. Chem. Res.***2014**, *47* (9), 2795–2803.
 21. Duan, L. L.; Feng, G. Q.; Zhang, Q. G. Large-Scale Molecular Dynamics Simulation: Effect of Polarization on Thrombin-Ligand Binding Energy. *Sci. Rep.***2016**, *6* (1), 31488.
 22. Leontyev, I.; Stuchebrukhov, A. Accounting for Electronic Polarization in Non-Polarizable Force Fields. *Phys. Chem. Chem. Phys.***2011**, *13* (7), 2613.

23. Ji, C.; Mei, Y. Some Practical Approaches to Treating Electrostatic Polarization of Proteins. *Acc. Chem. Res.* **2014**, *47* (9), 2795–2803.
24. Gao, Y.; Li, Y.; Mou, L.; Lin, B.; Zhang, J. Z. H.; Mei, Y. Correct Folding of an α -Helix and a β -Hairpin Using a Polarized 2D Torsional Potential. *Sci. Rep.* **2015**, *5* (1), 10359.
25. Duan, L.; Zhu, T.; Ji, C.; Zhang, Q.; Zhang, J. Z. H. Direct Folding Simulation of Helical Proteins Using an Effective Polarizable Bond Force Field. *Phys. Chem. Chem. Phys.* **2017**, *19* (23), 15273–15284.
26. Duan, L. L.; Zhu, T.; Zhang, Q. G.; Tang, B.; Zhang, J. Z. H. Electronic Polarization Stabilizes Tertiary Structure Prediction of HP-36. *J. Mol. Model.* **2014**, *20* (4), 2195.
27. Duan, L. L.; Gao, Y.; Mei, Y.; Zhang, Q. G.; Tang, B.; Zhang, J. Z. H. Folding of a Helix Is Critically Stabilized by Polarization of Backbone Hydrogen Bonds: Study in Explicit Water. *J. Phys. Chem. B* **2012**, *116* (10), 3430–3435.
28. Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan Zippers: Stable, Monomeric β -Hairpins. *Proc. Natl. Acad. Sci.* **2001**, *98* (10), 5578–5583.
29. Muñoz, V.; Eaton, W. A.; Thompson, P. A.; Hofrichter, J. Folding Dynamics and Mechanism of Bold Beta-Hairpin Formation. *Nature* **1997**, *390* (6656), 196–199.
30. Muñoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. a. A Statistical Mechanical Model for Beta-Hairpin Kinetics. *Proc. Natl. Acad. Sci. USA* **1998**, *95* (11), 5872–5879.

31. Dinner, A. R.; Lazaridis, T.; Karplus, M. Understanding Beta-Hairpin Formation. *Proc. Natl. Acad. Sci.***1999**, *96* (16), 9068–9073.
32. Pande, V. S.; Rokhsar, D. S. Molecular Dynamics Simulations of Unfolding and Refolding of a Beta-Hairpin Fragment of Protein G. *Proc. Natl. Acad. Sci.***1999**, *96* (16), 9062–9067.
33. Wei, G.; Mousseau, N.; Derreumaux, P. Complex Folding Pathways in a Simple β -Hairpin. *Proteins Struct. Funct. Bioinforma.***2004**, *56* (3), 464–474.
34. Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan Zippers: Stable, Monomeric β -Hairpins. *Proc. Natl. Acad. Sci.***2001**, *98* (10), 5578–5583.
35. D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, P.A. Kollman **Amber 10** University of California, San Francisco (2008)
36. Wickstrom, L.; Okur, A.; Simmerling, C. Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling Data. *Biophys. J.***2009**, *97* (3), 853–856.
37. Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.***2007**, *3* (1), 156–169.

38. Pastor, R. W.; Brooks, B. R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.***1988**, *65* (6), 1409–1419.
39. Uberuaga, B. P.; Anghel, M.; Voter, A. F. Synchronization of Trajectories in Canonical Molecular-Dynamics Simulations: Observation, Explanation, and Exploitation. *J. Chem. Phys.***2004**, *120* (14), 6363–6374.
40. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.***1993**, *98* (12), 10089–10092.
41. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. . Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.***1977**, *23* (3), 327–341.
42. Kolinski, A; Ilkowski, B; Skolnick, J. Dynamics and thermodynamics of beta-hairpin assembly: Insights from various simulation techniques. *Biophys. J* **1999**, *77*, 2942–2952.
43. Du, D; Zhu, Y; Huang, CY; Gai, F. Understanding the key factors that control the rate of beta-hairpin folding. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15915–15920.
44. Garcia, AE; Sanbonmatsu, KY. Exploring the energy landscape of a beta hairpin in explicit solvent. *Proteins* **2001**, *42*, 345–354.

45. Zhou, R; Berne, BJ; Germain, R. The free energy landscape for beta hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14931–14936.
46. Lee, J; Shin, S. Understanding beta-hairpin formation by molecular dynamics simulations of unfolding. *Biophys. J* **2001**, *81*, 2507–2516.
47. Zagrovic, B; Sorin, EJ; Pande, V. Beta-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol* **2001**, *313*, 151–169.
48. Zhou, Y; Linhananta, A. Role of hydrophilic and hydrophobic contacts in folding of the second beta-hairpin fragment of protein G: molecular dynamics simulation studies of an all-atom model. *Proteins* **2002**, *47*, 154–162.
49. Bolhuis, PG. Transition-path sampling of beta-hairpin folding. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12129–12134.
50. Wei, G; Mousseau, N; Derreumaux, P. Complex folding pathways in a simple beta-hairpin. *Proteins* **2004**, *56*, 464–474.

Chapter 5

5. Application of the Polarized Structure-Specific Backbone Charge Scheme on the Folding of Beta3s

5.1 Introduction

Figure 16 Schematic representation of beta3s. The hydrogen bonds are indicated in green dotted lines, while the purple arrows indicate CαH protons that give rise to NOEs.

Proteins are the biological machinery that execute specific vital functions in every cell of the human body by folding into their 3D structures.¹⁻³ However, *in vitro* experiments are unable to decipher the folding process dynamically at the atomic level and requires *in silico* methods such as molecular dynamics (MD) simulations to do so.⁴ MD simulations, which are based on force fields (FFs), are used to extract dynamic atomic information such as time evolution of secondary structures⁵, binding energies between receptors and ligands⁶, examining the actual dynamics of various biological processes like protein folding⁷⁻⁹ and movement of molecules through ion channels¹⁰⁻¹². By improving our understanding of the biological processes via molecular dynamic (MD) simulations, various applications such as peptide

design¹³⁻¹⁵, protein engineering¹⁶ and computer-aided drug design¹⁷ are made possible.

While MD simulations are very useful in providing the atomic details of biological processes, the accuracy of the simulations is heavily reliant on the accuracy of FFs and the sampling methods.¹⁸⁻²⁰ For example, the folding of a peptide may require hydrogen bonds to be formed between certain specific residues. If the force field is inaccurate in calculating these hydrogen bond energies and concludes that they are weaker than other non-covalent interactions, then these hydrogen bonds will not be formed and the peptide will be directed towards a misfolded structure that is different from the experimentally determined structure. Non-covalent interactions are fragile on their own, but there is strength in numbers. Therefore, the smallest inaccuracy in FFs can cause errors when evaluating non-covalent interactions, which will result in significant conformational errors during MD simulations.

One of the strongest non-covalent interactions is the hydrogen bond, of which many studies have found that commonly used FFs such as AMBER, CHARMM, etc. are unable to account for the polarization effect during the formation of hydrogen bonds in MD simulations.²¹⁻²³ Particularly for atomic charges, electrostatic potential (ESP) of individual amino acids were first obtained from quantum-mechanical calculations of di-residue systems in different secondary structures. Then, the ESP was fitted to the di-residue atoms to obtain atomic charges. Finally, the atomic charges were averaged between the different secondary structure to generate a final set of atomic charges that are to remain constant during the MD simulation.

Many studies showed that the consideration of polarization effect greatly increases the accuracy of MD simulations and result in the development of various models such as the fluctuating charge model²⁴⁻²⁶, the Drude oscillators²⁷⁻²⁹, inducible dipoles³⁰ and multipole electrostatics^{31,32}. In addition to these models, the newer models include the polarized protein-specific charge (PPC) model^{33,34}, the effective polarizable bond (EPB) method³⁵ and the polarizable hydrogen bond (PHB) model³⁶. All these models attempted the explicit inclusion of the polarization effect by dynamically updating the atomic charges throughout the MD simulation.

In this work, we introduce the polarized structure-specific backbone charge (PSBC) model, where it accounts for the polarization effect in MD simulation by updating the atomic charges of the backbone hydrogen bond atoms according to charge transfer equations derived from quantum-mechanical calculations. Compared to other polarizable models, the PSBC model does not require quantum-mechanical calculations of the peptide simulated at every time-step of the simulation and maintains the dynamic update of atomic charges, thereby reducing the computational cost and time while accounting for the polarization effect dynamically at the same time.

The PSBC model is applied to the Beta3s peptide, a *de novo* designed three-stranded beta sheet whose structure is folded *in vitro* and studied by NMR. The nuclear overhauser enhancements (NOEs) indicated the population of a monomeric structure in aqueous solution at 10°C, which is a three-stranded anti-parallel beta structure with turns at Gly-6-Gly-7 and Gly-14-Gly-15.³⁷ Although many studies managed to fold the Beta3s peptide successfully

with the CHARMM force field³⁸⁻⁴⁰, there are no studies of AMBER force field being able to do so currently. Therefore, the present work focuses on the stability of backbone hydrogen bonds and anti-parallel beta sheet folding for Beta3s as a benchmark protein using both PSBC and AMBER charges.

5.2 Simulation Details

The initial linear structure of beta3s was generated by LEaP using the AMBER96 force field, followed by steepest descent and conjugate gradient minimizations of 10,000 steps each under the GB^{OBC} implicit solvent model. The minimization is stopped when the energy gradient converges below 1.0×10^{-2} kcal/mole-Å. No periodicity and restraints were present during the minimization. The non-bonded cutoff distance is set at 999 Å. After minimization, the beta3s peptide is heated to 283 K over 100 ps before subjected to productive equilibration for 100 ns. The collision frequency applied is 4 ps^{-1} . A total of 3 runs each are conducted under the AMBER force field and the PSBC model, amounting to 600 ns of simulation time.

5.3 Results and Discussion

Figure 17 Most populated structures from the 3 runs in AMBER and PSBC.

Cluster analysis reveals the most populated structure in an MD simulation. Our analysis shows that in general, the three-stranded beta structure populates the MD simulation under the PSBC model. However, despite the standard AMBER96 force field favoring beta sheet formation, it is unable to stabilize and elongate the beta strands. Specifically, in runs 2 and 3, the AMBER force field is able to position two of the beta strands at the N-terminus in an anti-parallel manner, but the third strand is placed in parallel to the N-terminus strand instead of anti-parallel to the middle strand. In this study, the simulations in run 3 for both standard AMBER force field and the PSBC model will be used for further analysis.

Figure 18 DSSP analysis of run 3 in AMBER and PSBC. It is clearly observed that the AMBER force field is unable to fold a stable three-stranded anti-parallel β -sheet in the absence of the PSBC model.

DSSP was used to monitor the secondary structures formed throughout the simulation. As determined by the DSSP algorithm, Figure 18 shows the evolution of the secondary structures during the simulation of the third run. Under the PSBC model, the two anti-parallel beta strands at the N-terminus are formed within the first 5 ns, followed by the formation of the last beta

strand at the C-terminus at about 12 ns. The three anti-parallel beta strands, which are in agreement with the NMR structure proposed by the NOEs given by Alba et al., are then stabilized throughout the simulation. On the other hand, AMBER is only able to fold and stabilized traces of the two very short anti-parallel beta strands at the N-terminus, while misfolding the third beta strand at the C-terminus to be parallel to the first beta strand at the N-terminus.

Figure 19 NOE distances of the 4 residue pairs proposed experimentally to confirm the successful folding of beta3s under AMBER (left) and PSBC (right).

The shortest distance between 2 protons give rise to the most intense NOEs. The folded NMR structure of beta3s suggested four pairs of C α protons (listed in Figure 19) that provided the most intense NOEs since the folding into a three-stranded anti-parallel beta sheet will place these pairs of protons in very close proximity. Therefore, in order to ascertain whether our simulation has correctly folded beta3s, the distances between the four pairs of C α

protons were analyzed. The PSBC model clearly shows the four NOE distances to be at a significantly low value, as compared to the standard AMBER force field. In addition, the NOE distances remain stabilized for most of the time during the simulation. This indicates that the PSBC model not only is able to fold beta3s correctly, but it is also able to maintain the stability of the secondary structures to give rise to the stability of the NOE distances.

Figure 20 Hydrogen bond distances of various residue pairs in beta3s. The PSBC model (shown in red) is able to stabilize the hydrogen bonds to a greater extent than the AMBER force field (shown in black).

The purpose of the PSBC model is to allow the dynamic description of backbone hydrogen bonds during a MD simulation. Therefore, we conducted hydrogen bond analysis on hydrogen bonds found just next to every NOE pair. As seen in Figure 20, although the standard AMBER force field is able to stabilize some hydrogen bonds to a similar extent as the PSBC model. The greater fluctuation seen from AMBER indicates that fixed atomic charges are able to identify possible backbone hydrogen bonds but unable to derive a high

stability to those hydrogen bonds once they are formed. In agreement to the DSSP and cluster analyses where the beta strand at the C-terminus is observed to have misfolded at the wrong position, the hydrogen bonds that are found between the middle strand and the strand at the C-terminus are much further in AMBER than in the PSBC model.

Figure 21 Free energy landscapes of the beta3s folding process using the radius of gyration and the Ca- Ca distances of the residue pairs that give rise to the NOEs. The lowest energy basin is indicated in red while the highest energy basin is indicated in blue.

Besides being able to fold the 3D structure of the peptide successfully, an adequate force field should also be able to show that the folded structure is the lowest energy structure since protein folding has been determined to be a process that locates the native structure to be the lowest energy structure. Free energy landscapes (FELs) using each of the 4 NOE distances and the radius of gyration as the axes were constructed for AMBER and PSBC (Figure 21). The correctly folded beta3s is expected to occupy the lowest energy basin

with the smallest radius of gyration and the shortest NOE distance. While AMBER is able to indicate that its folded structure is found as an isolated small area in some of its FELs, other FELs show that its folded structure does not possess the shortest NOE distances. Instead, the lowest energy basin is spread out over a range of NOE distances, indicating that the AMBER force field is unable to locate a unique lowest energy folded structure. The PSBC model, on the other hand, clearly shows that although many structures are sampled during the MD simulation, most of these structures are high in energy and therefore are not the native structure. Only a lowest energy basin is found in each of its FEL with a small radius of gyration and a low NOE distance. This shows that the PSBC model is able to not only locate the lowest energy structure, but this lowest energy structure possesses physical qualities that indicate it is correctly folded.

5.4 Chapter Summary

The MD simulation is a useful tool in not only reproducing the 3D structure of the folded protein or peptide, but more importantly, it is able to allow the observation of the folding process. While current force fields are able to describe the folding process most of the time, there is still certain inadequacy that is hindering the successful folding of proteins in MD simulations. Therefore, the PSBC model is developed to increase the accuracy of current force fields by basing on the fundamental knowledge that the successful folding of secondary structure heavily relies on the formation of strong and stable backbone hydrogen bonds. Although current force fields describe the hydrogen bond as an electrostatic term where its energy relies on

the bond distance, it is inadequate in describing the polarization effect between the hydrogen bond atoms. Therefore, the PSBC model supplements this inadequacy by considering the relationship between atomic charges and bond distance when polarization takes place during hydrogen bond formation. With the polarization effect considered in the electrostatic term, we have shown that it is effective in rescuing misfolding even if the inaccuracies in the force field used does not favor a particular secondary structure.

References

1. Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, 426 (6968), 884–890.
2. Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, 6 (3), 197–208.
3. Jahn, T. R.; Radford, S. E. The Yin and Yang of Protein Folding. *FEBS J.* **2005**, 272 (23), 5962–5970.
4. Daggett, V. Protein Folding–Simulation. *Chem. Rev.* **2006**, 106 (5), 1898–1916.
5. Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, 22 (12), 2577–2637.
6. Gilson, M. K.; Zhou, H.-X. Calculation of Protein-Ligand Binding Affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, 36 (1), 21–42.

7. Scheraga, H. A.; Khalili, M.; Liwo, A. Protein-Folding Dynamics: Overview of Molecular Simulation Techniques. *Annu. Rev. Phys. Chem.* **2007**, *58* (1), 57–83.
8. Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9* (9), 646–652.
9. Shea, J.-E.; Brooks III, C. L. FROM FOLDING THEORIES TO FOLDING PROTEINS: A Review and Assessment of Simulation Studies of Protein Folding and Unfolding. *Annu. Rev. Phys. Chem.* **2001**, *52* (1), 499–535.
10. Shrivastava, I. H.; Sansom, M. S. P. Simulations of Ion Permeation Through a Potassium Channel: Molecular Dynamics of KcsA in a Phospholipid Bilayer. *Biophys. J.* **2000**, *78* (2), 557–570.
11. Shi, Q.; Izvekov, S.; Voth, G. A. Mixed Atomistic and Coarse-Grained Molecular Dynamics: Simulation of a Membrane-Bound Ion Channel. *J. Phys. Chem. B* **2006**, *110* (31), 15045–15048.
12. Roux, B.; Karplus, M. Molecular Dynamics Simulations of the Gramicidin Channel. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23* (1), 731–761.
13. Yang, W.; Lai, L. Computational Design of Ligand-Binding Proteins. *Curr. Opin. Struct. Biol.* **2017**, *45*, 67–73.
14. Childers, M. C.; Daggett, V. Insights from Molecular Dynamics Simulations for Computational Protein Design. *Mol. Syst. Des. Eng.* **2017**, *2* (1), 9–33.
15. Poorebrahim, M.; Sadeghi, S.; Rahimi, H.; Karimipoor, M.; Azadmanesh, K.; Mazlomi, M. A.; Teimoori-Toolabi, L. Rational Design of DKK3

- Structure-Based Small Peptides as Antagonists of Wnt Signaling Pathway and in Silico Evaluation of Their Efficiency. *PLoS One* **2017**, *12* (2), e0172217.
16. Ebert, M. C.; Pelletier, J. N. Computational Tools for Enzyme Improvement: Why Everyone Can – and Should – Use Them. *Curr. Opin. Chem. Biol.* **2017**, *37*, 89–96.
17. Ganesan, A.; Coote, M. L.; Barakat, K. Molecular Dynamics-Driven Drug Discovery: Leaping Forward with Confidence. *Drug Discov. Today* **2017**, *22* (2), 249–269.
18. González, M. A. Force Fields and Molecular Dynamics Simulations. *École thématique la Société Française la Neutron*. **2011**, *12*, 169–200.
19. Baker, C. M. Polarizable Force Fields for Molecular Dynamics Simulations of Biomolecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5* (2), 241–254.
20. Best, R. B.; Buchete, N.-V.; Hummer, G. Are Current Molecular Dynamics Force Fields Too Helical? *Biophys. J.* **2008**, *95* (1), L07–L09.
21. Ji, C.; Mei, Y. Some Practical Approaches to Treating Electrostatic Polarization of Proteins. *Acc. Chem. Res.* **2014**, *47* (9), 2795–2803.
22. Duan, L. L.; Feng, G. Q.; Zhang, Q. G. Large-Scale Molecular Dynamics Simulation: Effect of Polarization on Thrombin-Ligand Binding Energy. *Sci. Rep.* **2016**, *6* (1), 31488.
23. Leontyev, I.; Stuchebrukhov, A. Accounting for Electronic Polarization in Non-Polarizable Force Fields. *Phys. Chem. Chem. Phys.* **2011**, *13* (7), 2613.

24. Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization Effects in Molecular Mechanical Force Fields. *J. Phys. Condens. Matter* **2009**, *21* (21), 333102–333121.
25. Banks, J. L.; Kaminski, G. A.; Zhou, R. H.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. Parametrizing a Polarizable Force Field from Ab Initio Data. I. the Fluctuating Point Charge Model. *J. Chem. Phys.* **1999**, *110* (2), 741–754.
26. Shi, Y.; Ren, P.; Schnieders, M.; Piquemalc, J. P. Polarizable Force Fields for Biomolecular Modeling. *Rev. Comput. Chem.* **2015**, *28*, 51–86.
27. Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013**, *9* (12), 5430–5449.
28. Lemkul, J. A.; Huang, J.; Roux, B.; Mackerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016**, *116* (9), 4983–5013.
29. Stanley, G. G. *Force Fields and Minimization*.
30. Elking, D.; Darden, T.; Woods, R. J. Gaussian Induced Dipole Polarization Model. *J. Comput. Chem.* **2007**, *28* (7), 1261–1274.
31. Murray, J. S.; Politzer, P. Molecular Electrostatic Potentials and Noncovalent Interactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, e1326.

32. Zhang, C.; Bell, D.; Harger, M.; Ren, P. Polarizable Multipole-Based Force Field for Aromatic Molecules and Nucleobases. *J. Chem. Theory Comput.* **2017**, *13* (2), 666–678.
33. Zhang, D. W.; Zhang, J. Z. H. Molecular Fractionation with Conjugate Caps for Full Quantum Mechanical Calculation of Protein-Molecule Interaction Energy. *J. Chem. Phys.* **2003**, *119* (7), 3599–3605.
34. Ji, C.; Mei, Y.; Zhang, J. Z. H. Developing Polarized Protein-Specific Charges for Protein Dynamics: MD Free Energy Calculation of pKa Shifts for Asp26/Asp20 in Thioredoxin. *Biophys. J.* **2008**, *95* (3), 1080–1088.
35. Xiao, X.; Zhu, T.; Ji, C. G.; Zhang, J. Z. H. Development of an Effective Polarizable Bond Method for Biomolecular Simulation. *J. Phys. Chem. B* **2013**, *117* (48), 14885–14893.
36. Gao, Y.; Lu, X.; Duan, L. L.; Zhang, J. Z. H.; Mei, Y. Polarization of Intraprotein Hydrogen Bond Is Critical to Thermal Stability of Short Helix. *J. Phys. Chem. B* **2012**, *116* (1), 549–554.
37. de Alba, E.; Santoro, J.; Rico, M.; Jiménez, M. A. De Novo Design of a Monomeric Three-Stranded Antiparallel Beta-Sheet. *Protein Sci.* **1999**, *8* (4), 854–865.
38. Ferrara, P.; Caflisch, A. Folding Simulations of a Three-Stranded Antiparallel Beta -Sheet Peptide. *Proc. Natl. Acad. Sci.* **2000**, *97* (20), 10780–10785.
39. Cavalli, A.; Haberthür, U.; Paci, E.; Caflisch, A. Fast Protein Folding on Downhill Energy Landscape. *Protein Sci.* **2003**, *12* (8), 1801–1803.

40. Rao, F.; Settanni, G.; Caflisch, A. Estimation of Folding Probabilities and Φ Values From Molecular Dynamics Simulations of Reversible Peptide Folding. In *Protein Folding Protocols*; Humana Press: New Jersey; pp 225–250.

Chapter 6

6. Application of the PSBC model onto the folding of α -helices with varying helical propensities

6.1 Introduction

Protein folding has continually been a riddle to researchers both experimentally and theoretically.¹ The “protein folding problem” that was initiated around half a century ago summarizes the importance of apprehending protein folding and also highlighted the main research topics that are continuously being conducted in the protein community up till today.² Some of the principal studies conducted to comprehend protein folding include the understanding of the connection between primary and tertiary structures of proteins, exploring the mechanism accounting for the rapidity of protein folding^{3,4} and formulating computer algorithms capable of predicting the unique three-dimensional structures of proteins in relation to its amino acid sequence. Ongoing work placed in gaining more knowledge on proteins and how they fold is a vital effort in all fields of the sciences as it enabled researchers to gain useful insights on protein-related diseases such as Alzheimer’s disease and Creutzfeldt-Jakob disease that are often associated with protein misfolding.⁵

Folded proteins assumed its lowest energy conformation by establishing a combination of non-covalent interactions involving amino acid-amino acid interactions and amino acid-solvent interactions.³³ Modelling of the exact nature of these interactions makes theoretical prediction of native

structures a complex task as the correctness of the prediction depends heavily on the quality of the force field used.⁶ The use of classical molecular dynamics (MD) simulation in protein research has become a norm in recent years. Existing traditional force fields such as AMBER, CHARMM and GROMOS has enabled the universal application of these force fields in the modelling of a variety of protein types, making the exploration of the dynamics, structures and functions of proteins effortless. However, this advantage comes with a deficiency that is brought about by the mean-field treatment of amino acids by these force fields. Atomic charges assigned to atoms are constant throughout the simulation performed despite changes in the electrostatic environment of the protein. This aspect of the force field may compromise the accurate description of proteins by force fields since the inherent variation in the atomic charges of atoms in the presence of neighbouring atoms or an external field was not accounted.

Distortion in the electron density of an atom when in contact with atoms in the vicinity or in the presence of an external field works in the favor of lowering the overall electrostatic interaction energy of the system.⁷ Hence, it is imperative for force fields to consider the polarizability of atoms by introducing parameters accounting for polarization effects into simulations. This is especially so during the modelling of protein folding whereby continual changes in conformation occur prompting constant changes in the surrounding environment of the atoms.⁸ Integration of polarization effects into force fields has been accomplished using popular methods such as the induced dipole model⁹⁻¹¹, a fluctuating charge model¹², Drude oscillator model¹³ and quantum

mechanics (QM) method. While different protocols were followed by these polarizable models, all of them shared a common purpose that is initiating a molecular feedback towards the fluctuation in the electrostatic environment of atoms during simulations. Multitudes of force field development efforts to improve existing force fields has been carried out making headway in effectively empowering researcher to describe proteins and interactions governing their tertiary configuration more precisely than before.

Encouraging results have been observed through studies examining the effectiveness of incorporating polarization effect into protein modelling.¹⁴ Stabilization of proteins in its native conformation and better stability of hydrogen bonds that are often underestimated in non-polarizable force fields are some of the positive outcomes of these studies.¹⁵ Hydrogen bonding plays an essential role in maintaining the structural integrity of proteins' secondary structures specifically α -helices and β -sheets. For this reason, accurate electrostatic representation of hydrogen bonds by force fields may serve as a possible solution to improve the precision of native structure prediction and folding dynamics studies. Strategies involving on-the-fly updating of charges of atoms involved in backbone hydrogen bonding such as adaptive hydrogen bond-specific charge (AHBC) scheme¹⁶ and polarizable hydrogen bond (PHB) model¹⁷ have been effective in the *ab initio* folding of α - helices of various lengths. These studies reflected the feasibility of improving the predictive accuracy of MD simulations by building polarization effects into force fields through charge updating schemes.

Here, we will be introducing a new charge model named the polarized structure-specific backbone charge (PSBC) model. Similar to the AHBC scheme and the PHB model, the PSBC model accounts for the polarizability of proteins by periodically updating the charges of atoms involved in backbone hydrogen bonding during the simulation. While the AHBC scheme subjects residues participating in backbone hydrogen bonding to QM calculations during the charge updating step, the PSBC and the PHB models conduct on-the-fly charge updates of atoms participating in backbone hydrogen bond (C, O, N and H) using a mathematical function that defines the relationship between hydrogen bond length and charge transferred between hydrogen bond donor and acceptor during formation. This reduces the computational cost associated with repeated QM calculations entailed when the AHBC scheme is utilized.

Derivation of the PSBC and the PHB models involved the initial QM calculations of a pair of dialanine residues with separation between hydrogen bond donor (N) and acceptor (O) ranging between 2.5 Å and 6.5 Å. This step is essential to obtain distance-dependent charges of O and N atoms that are fitted using the restrained electrostatic potential (RESP) method. Despite sharing a similar concept, there are several factors which differentiate these two formula-based polarized models. Firstly, the differences in the geometry of backbone hydrogen bonds in α -helix and anti-parallel β -sheet were considered during formula derivation resulting in two separate functions for α -helix and anti-parallel β -sheet. This factor rendered PSBC a structure-specific polarization model which may be useful in improving the efficiency of folding simulations for proteins with multiple secondary structures. Secondly, the

electrostatic potentials (ESP) of O and N atoms are computed in a single QM calculation without considering background effects in the PSBC model. This is contrast to the PHB model which subject one of the alanine residues to QM calculation with the other residue present as background charges positioned at its corresponding atomic coordinates.

Positive outcomes arising from the incorporation of electrostatic polarization in the folding of peptides with high helicity has been heavily proven through the application of AHBC and PHB models. However, folding of peptides with low helicity is still a challenge as most existing force fields have high propensities for helical structures.²⁹ To examine the capability of the PSBC model to fold proteins of different helicity, we have folded three peptides namely ER, E5 and ES peptides from their linear conformation using conventional MD simulation. The helicity of these peptides as determined through circular dichroism (CD) spectroscopy is 0.720, 0.359 and 0.217 respectively. For comparison sake, we have also conducted parallel folding simulations of these three peptides using traditional AMBER force fields. This study serves as a benchmark test to examine the feasibility of the PSBC model to fold helices. Application of the PSBC model in the folding of β -sheets and mixed-fold proteins will be explored in future studies.

6.2 Simulation Details

To examine the capability of the PSBC model to fold peptides of varying helicity, we have conducted conventional MD simulations to fold three peptides namely ER, E5 and ES from their linear conformations. The linear structures of these peptides were first constructed using LEaP module in

AmberTools 1.2. Folding simulations of ER and ES were conducted in implicit solvation which was described using the generalized Born solvation model (igb5). E5, on the other hand, was folded in a water box. AMBER ff03 force field was employed in the running of all simulations in this study. For simulations performed using the PSBC model, all parameters used are consistent with that of AMBER ff03 force field except for the partial charges of hydrogen bond donors and acceptors that are subjected to periodic on-the-fly updates.

Prior to the production run, minimization of the linear structures of the three peptides was performed. For ER and ES, peptides were minimized in 10000 steps using the steepest descent method and thereafter using the conjugate gradient method until the energy gradient of the system is less than 0.01 kcal/mol-Å. For E5 peptide which was solvated in a TIP3P water box, minimization was conducted in two steps. Firstly, minimization of water molecules and sodium ions that were added to neutralize the system was performed in 10000 steps using the steepest descent method and another 10000 steps through the conjugate gradient method. The whole E5-water system was subsequently minimized using the steepest descent method for 5000 steps and later minimized using the conjugate gradient method until the energy gradient of the system converges to less than 0.01 kcal/mol-Å. A minimum distance of 10 Å was set between peptide and the periphery of the TIP3P water box.

After minimization, ER, E5 and ES systems were heated to their experimental temperatures of 298 K, 277 K and 273 K respectively in 100 ps.¹⁻⁶

Langevin thermostat with collision frequency of 4 ps⁻¹ was used to regulate the temperature during the simulation. Production runs were performed for 100 ns for ER and 20 ns each for E5 and ES. An integration time step of 1 fs was used. All non-bonded interactions in E5 are calculated with the Particle Mesh Ewald method using a cutoff distance of 12 Å. All bonds involving hydrogen atoms were constrained using the SHAKE algorithm. For each peptide, four MD simulations each were performed using the PSBC model and conventional AMBER force field. During MD simulations performed using the PSBC model, charge update was performed every 0.5 ps for ER, every 0.1 ps for E5 and every 0.05 ps for ES.

6.3 Results and Discussion

Figure 22 (A) Helical content of ER, E5 and ES peptide determined through simulations conducted as a function of the experimental helicity of the three peptides. (B) Summary of the helical content predicted through simulations

and the experimental helicity of ER, E5 and ES peptide. (C) Amino acid sequences of ER, E5 and ES peptide.

In this work, the *ab initio* foldings of three helical peptides namely ER, E5 and ES from their linear states were conducted to assess the capability of the PSBC model to fold proteins of different helicity.¹⁶⁻²¹ The amino acid sequences of these peptides are given in Figure 22. Parallel folding simulations of the three peptides using conventional AMBER ff03 force field were also carried out for comparison purposes. Four MD simulations were performed for each peptide using both polarized (PSBC) and non-polarized (AMBER) force fields.

The experimental helicity of ER, E5 and ES as estimated from the ellipticity of these peptides acquired at 222 nm using circular dichroism (CD) spectroscopy by others¹⁶⁻²¹ are 0.720 (298 K), 0.359 (273 K) and 0.217 (277 K) respectively. The high helicity of the ER peptide is attributed to the predominantly alanine makeup and strategic placement of Glu and Arg at *i*+4 positions, promoting the formation of salt bridges that stabilize the helix. (Figure 23A)¹⁶ On the contrary, the placement of Glu at an unfavorable position in AQ host peptide (Ac-(AAQAA)3Y- NH₂) has led to the decrease in helicity from 0.483 to 0.359.^{18,21} ES peptide understandably have lower helicity compared to the polyalanine variants (ER and E5) as the primary sequence of this peptide comprised of several residues with low helical propensities such as Ile (0.41 kcal/mol), Ser (0.50 kcal/mol) and Thr (0.66 kcal/mol)²².

The mean helical contents calculated by averaging over four simulations conducted using PSBC and AMBER charge together with

experimental helicity are shown in Figure 22A. In this study, amino acids are assumed to be helical only when their dihedral angles are within the range of $(\phi, \psi) = (-60 \pm 20, -40 \pm 20)$.²³ The mean helical content obtained from each simulation was computed by averaging the helical content of equilibrium structures defined by the convergence of the backbone root-mean-square deviation (RMSD) to a value less than 3 Å when referenced to the lowest energy structure acquired from each trajectory. Utilizing the PSBC model, the mean helical contents of the equilibrium structures of ER, E5 and ES are 0.745, 0.318 and 0.224 respectively. On the other hand, the mean helical contents calculated based on simulations conducted under AMBER ff03 force field for ER, E5 and ES peptides are 0.661, 0.512 and 0.378 respectively. While both simulation types successfully reproduce the helicity order of ER, E5 and ES peptides, the PSBC model provided helical content predictions that were closer to the experimentally determined helicity. Additionally, all simulations conducted under mean-field tend to overestimate the helical contents of peptides with low helicity namely E5 and ES (Figure 22A).

The only factor that differentiates the two simulation types, PSBC and AMBER, is the atomic charges of the backbone atoms, N, H, C and O. The better helical content prediction by the PSBC model suggests the importance of considering electrostatic polarization of backbone hydrogen bonds during MD simulation and charge updating may be a feasible method of accounting for protein polarizability. Furthermore, inclusion of polarization effect of atoms involved in backbone hydrogen bonding may possibly prevent the overestimation of helical structures in proteins and peptides of low helicity. In

the sections that follow, we have conducted in-depth exploration of the folding dynamics of the peptides with and without backbone polarizability factor. We will also highlight the factors that resulted in the better agreement between predicted helical content and experimental helicity when backbone polarizability was considered.

6.3.1 Peptide with high helicity

Figure 23 (A) Schematic representation of (i+4)-ER peptide with salt bridges (red dotted lines) formed between Glu and Arg residues illustrated. (B) Secondary structure assignment for ER peptides simulated using PSBC (*top*) and AMBER (*bottom*) charges. (C) RMSF of ER peptide during the last 50 ns of the simulations conducted using PSBC (black) and AMBER (red) charges.

Based on helical contents predicted for ER (Figure 22B), both simulation types seem to have successfully fold the ER peptide into a complete α -helix (native structure). However, the slight difference in helical content estimated by the two simulations may imply a concomitant difference in the conformation of the folded peptide. DSSP (Define Secondary Structure of Proteins) algorithm²⁴ was used to observe the changes in the secondary structure of the peptide during the folding simulations. The plots of secondary structure type as a function of simulation time are presented in Figure 23B. Both simulations were able to fold the extended structure of ER into a full helix as indicated by the blue lines in the DSSP plots. This observation is expected as AMBER ff03 force field possesses high helical propensity and is capable of effectively folding helical peptides and proteins even without the consideration of polarization effects.^{25,26} However, closer scrutinization of the trajectories acquired using both PSBC and AMBER charges revealed differences that accentuated the importance of polarization effect of backbone hydrogen bonds in the modeling of protein folding.

Figure 24 Secondary structure variation of the ER peptide during the folding simulation performed using PSBC and AMBER charges. DSSP algorithm was used to monitor the changes in secondary structures.

A factor that differentiates the folding simulations conducted with and without the consideration of the polarization effect is the better stabilization of the folded helix upon formation when PSBC was used. From the DSSP plots in Figure 23B, the utilization of the PSBC model resulted in lesser transitions between α -helix and 3_{10} -helix upon achieving the native state. This is in contrary to the simulation conducted without polarization effect. Transitions between α -helix and 3_{10} -helix during the AMBER-based simulation were more frequent even after the folding of ER into a complete α -helix. Frequent switches between these two configurational states were noted in all simulation conducted using traditional, non-polarizable force field. (Figure 24) To further support this observation, root-mean-square fluctuations (RMSF) of the later half of the simulations was calculated to monitor the stability of the

folded peptide during the simulation (Figure 23C). The lower RMSF values acquired for simulations under polarized force field corroborated the likelihood of polarization effect of backbone hydrogen bonds influencing the stability of ER upon folding. Additionally, the rapid folding of the peptide into its native configuration in the PSBC-based simulation was apparent based on the DSSP plots in Figure 23B which showed the development of the α -helix earlier into the simulation (around 20 ns) compared to the AMBER-based simulation. This observation was noted in all simulations performed for ER. (Figure 24)

Figure 25 (A) Free energy landscape of ER folding simulations conducted using PSBC with structures representing populations on the free energy landscape included. Overlap of the theoretical structure of ER (yellow) and folded ER (cyan) acquired through simulations conducted using PSBC. (B) Free energy

landscape of ER folding simulations conducted using AMBER charges with structures representing populations on the free energy landscape included. Overlap of the theoretical structure of ER (yellow) and folded ER (cyan) acquired through simulations conducted using AMBER charges. (C) Lowest energy conformations (grey) acquired through the four simulations conducted using PSBC and AMBER charges overlapped with the theoretical structure of ER (black) for comparison.

Distributions of protein structures attained during both simulations were also examined by constructing two-dimensional free energy landscapes using weighted histogram analysis method.^{27,28} Backbone RMSD and radius of gyration (Rg) are selected as the reaction coordinates of the free energy landscapes produced in Figure 25. The backbone RMSD of the peptide was calculated relative to a theoretically constructed ER (Figure 23A) since the experimental structure was not available. The presence of a prominent population of backbone RMSD around 2 Å and Rg between 8 to 9 Å in Figure 3A and 3B further evinces the successful folding of ER peptide from its linear state by both PSBC- and AMBER-based simulations. The folded structures acquired by both simulations were comparable to the theoretical structure of ER as showed in Figure 25 (right).

Although the folding of ER was achieved by both simulations, free energy landscapes in Figure 25A and 25B portrayed differences that offer insights on the role of polarization effect in the stabilization of α -helix. From the free energy landscape constructed based on simulation conducted using AMBER charges, a large region with relatively gradual energy slopes were

observed at backbone RMSD under 5 Å and Rg between 8 to 10 Å. This relatively smooth area composed of structures in both folded and partially folded states. The small energy barrier between these two states may suggest the equal preference of the ER peptide to assume folded and partially folded conformations when mean-field was utilized (Figure 25B, inset). In contrast to the observations seen for the simulation performed under AMBER force field, simulations conducted using the PSBC model gave rise to definitive populations corresponding to folded and unfolded (RMSD $\approx$ 4.5 Å) states with energy barrier of more than 1.0 kcal/mol, indicating the preference for the folded state. The presence of two distinctive populations in the free energy landscape in Figure 25A and the rapid folding of the protein observed from the DSSP plot in Figure 23B effectively illustrates the two-state folding behavior of α -helices upon factoring the polarization effect of backbone hydrogen bonds during MD simulation.¹⁶

Additionally, we have extracted the lowest energy structures sampled during all simulations performed and compared their conformations with the theoretical structure of ER. Based on Figure 25C, the lowest energy structure sampled for all four MD simulations performed using the PSBC model gave rise to perfectly helical structures that showed good agreement with the theoretical configuration of ER. This is in contrast to the simulations conducted using traditional AMBER force field that showed varying lowest energy conformations acquired during the four simulation runs. A slightly bent helix was attained when ER was folded using AMBER charges. Fixed charges implemented by traditional force field may underestimate the strength of

backbone hydrogen bonds hence favoring the bent helix compared to the perfectly straight helix proposed for ER.¹⁶

Figure 26 (A) Schematic representation of 13 possible backbone hydrogen bonds in ER peptide. (B) (top) Average occupancy (bottom) and average length of 13 backbone hydrogen bonds formed in folded ER during the four simulations performed each using PSBC and AMBER charges. (C) Probability distribution of backbone hydrogen bonds formed during all simulations conducted using PSBC (blue) and AMBER (red) charges.

The efficient folding and stabilization of secondary structures such as α -helix is significantly dependent on the stability of backbone hydrogen bonds. Factoring the inhomogeneity of the protein environment during folding simulations through the integration of electrostatic polarization effect into

force fields have been shown in numerous studies to be beneficial towards the stabilization of backbone hydrogen bonds. Hence, to investigate the capacity of the PSBC model to stabilize backbone hydrogen bonds, we have calculated the average occupancy and length of 13 possible backbone hydrogen bonds formed in ER during the four simulations each performed using PSBC and AMBER charges (Figure 26). Utilization of the PSBC model saw the higher average occupancy and shorter average length of 13 backbone hydrogen bonds possibly formed in ER peptide compared to the simulations conducted under mean-field. (Figure 26B) These data suggested that the backbone hydrogen bonds formed in the folding simulation using PSBC are stronger than the backbone hydrogen bonds formed through simulations performed under mean-field, justifying the ability of the PSBC model to enhance the stability of backbone hydrogen bonds. Subsequently, enhancement of the backbone hydrogen bonds strength through the consideration of backbone polarization effect resulted in better folding accuracy and stabilization of the perfectly helical ER peptide. (Figure 25A and 25C)

The helical character of the ER peptide folded under polarized and non-polarized force field was further scrutinized by tallying the number of backbone hydrogen bonds formed during the 100 ns simulations conducted. Backbone hydrogen bond is assumed to be formed when the length between the carbonyl oxygen atom and amide hydrogen atom that are four residues apart is less than 3.0 Å and the angle between the acceptor O, hydrogen and donor N atoms is between 120° and 180°. Figure 4C showed the distribution of the number of hydrogen bonds formed by ER peptide sampled during the

simulation. The simulation conducted using PSBC displayed a narrow distribution in hydrogen bond count with most of the protein structures sampled forming 11 to 12 backbone hydrogen bonds. On the other hand, simulation conducted using AMBER charges showed a broader distribution in hydrogen bond count with structures having three to eight backbone hydrogen bonds constituting more than 60% of the trajectory in comparable probability of occurrences. This analysis further emphasized the vital role of backbone polarization effect in the development of backbone hydrogen bonds. As such, improving the stability of backbone hydrogen bonds is a key factor in the folding and stabilization of perfectly α -helical peptides and the PSBC model has managed to accomplish this task.

6.3.2 Peptides with low helicity

While the folding of proteins with high helicity are often achievable due to the high helical propensities of existing force fields, the folding of peptides with low helicity is still a challenge that needs to be addressed. To verify the ability of the PSBC model to fold proteins of varying helicity, foldings of two peptides with low helicity namely E5 and ES have also been carried out. Differences in the folding dynamics of E5 and ES with and without the consideration of the polarization effect of backbone hydrogen bonds will be scrutinized here. In this work, only the folding of E5 was attempted using explicit solvation model as the utilization of an implicit solvation model has resulted in the misfolding of E5 into a structure with high helical content regardless of whether the polarization effect was accounted for or not. (data not shown) This is understandable as the combination of AMBER ff03 force

field and GB solvation models has been reported to boost α -helical propensity.^{25,26,30} The ease of α -helix formation in implicit solvation simulation may be further aggravated in the case of E5 owing to its dominantly alanine composition. Hence, to minimize the preference of the force field towards α -helical structures, explicit solvation simulation was applied to fold E5.

Figure 27 Secondary structure assignment for (A) E5 and (B) ES peptides simulated using PSBC (*top*) and AMBER (*bottom*) charges. Representative structures of the three most populated clusters acquired through cluster analyses performed for MD trajectories acquired for the folding of (C) E5 and (D) ES peptides under PSBC (*top*) and AMBER ff03 force field (*bottom*). Red represents 310-helix, blue represents α -helix and cyan represents turn.

Since no experimental or proposed structure is available for ES and E5, the feasibility of the PSBC model to precisely determine the helical content of

peptides with low helicity will be substantiated by comparing two structural features of the folded protein attained through simulations performed under AMBER force field and the PSBC model. These two features are the secondary structures assumed by the folded protein during the simulation and backbone ($i+4$)-hydrogen bonds formed during the simulation. Overestimation of helical structures is a common characteristic of AMBER-based simulations of peptides with low helicity as seen through helical contents calculated in Figure 22. This observation was substantiated by the DSSP plots constructed for a representative simulation of both E5 and ES in Figure 5A and 5B respectively which showcased the folding of more helical structures (3_{10} -helix and α -helix) followed by their stabilization when these peptides were folded under mean-field. On the contrary, implementation of the PSBC model in the folding of E5 and ES saw lesser helical structure folded compared to when fixed charges were implemented (Figure 27B, top). To further support the structural variation observed through the DSSP plots, we have conducted cluster analyses via K-means algorithm³¹ to extract structural information from the MD trajectories.

Representative structures of three of the most populated clusters are presented in Figure 27C and 27D for E5 and ES peptides respectively. E5 and ES peptides folded using mean-field are inclined to fold into α -helix and this conformation constitutes more than half of the 20 ns simulation. (Figure 27C and 27D) Simultaneously, the folding of E5 and ES under the PSBC model did not favor the folding of α -helix. Under the PSBC model, cluster analysis of the trajectory acquired for E5 folding showed the dominant presence of structures

assuming a 3_{10} -helical conformation (>80% occurrence) while the folding of ES favored the turn conformation (>90% occurrence). From these observations, the importance of backbone hydrogen bond polarizability in accurately predicting the helical propensity of peptides is portrayed. The inclusion of polarization effect during MD simulation prevented the overestimation of α -helix seen in AMBER-based simulation hence providing us with a closer to experimental prediction of the helical contents of E5 and ES peptides. In the following two sections, we will discuss in greater detail the difference in the folding dynamics of both E5 and ES peptides under polarized (PSBC) and non-polarized (AMBER ff03) force field. The analyses presented below are based on a single trajectory from each simulation type for ease of comparison.

6.3.3 Folding of E5 peptide

Figure 28 (A) Schematic representation of four backbone hydrogen bonds formed when E5 peptide was folded using standard AMBER charges. (B) Occupancy of the four possible backbone hydrogen bonds formed in E5 during the simulation performed using PSBC (blue) and AMBER (red) charges. (C) Variation in the length of backbone hydrogen bonds possibly formed between hydrogen bond pairs (7@O-11@H, 8@O-12@H, 9@O-13@H and 10@O-14@H) in E5 during simulations performed under AMBER force field and PSBC.

To better comprehend the difference in the folding dynamics of E5 under polarized and non- polarized force field, the activity of backbone hydrogen bonds formed between atoms four residues apart were analyzed. The occupancies and lengths of four backbone hydrogen bonds namely Ala7@O—Ala11@H, Ala8@O—Ala12@H, Gln9@O—Ala13@H and Ala10@O—Gln14@H (Figure 28A) were calculated from simulations performed under AMBER force field and the PSBC model and presented in Figure 28B and 28C respectively. The favored folding of α -helix observed from the DSSP plot in Figure 27A (bottom) for AMBER-based simulation, was supported by the high occupancies of the four aforementioned backbone hydrogen bonds. (Figure 28B) The quick stabilization of these hydrogen bonds to lengths lower than 3.5 Å after only 2 ns further elucidated the stability of the α -helix during the AMBER-based simulation. (Figure 28C) On the contrary, only backbone hydrogen bond between Ala8 and Ala12 was formed in the PSBC-based simulation as indicated by the significant occurrence of this hydrogen bond in Figure 28B and the decrease in the distance between Ala8@O and Ala12@H to less than 3.5 Å in Figure 28C. The formations of other backbone hydrogen

bonds were not detected or are significantly weaker compared to its counterpart in the AMBER simulation. This was demonstrated by the lower occupancies and greater fluctuations of these hydrogen bonds at lengths longer than 4 Å.

Dynamics of hydrogen bonds during protein folding is not only governed by hydrogen bonds formed among amino acids. Hydrogen bonds formed between protein and aqueous solvent also play a significant role in determining the native conformation of the protein.³³ Since the folding of E5 was conducted in explicit water solvent, difference in protein-solvent interaction with and without the consideration of backbone polarization effect was explored. Of particular interest are the protein-water interactions around backbone ($i+4 \rightarrow i$) hydrogen bond pairs highlighted in Figure 28 which were responsible for the high helicity predicted in the AMBER-based simulation. Hence, we have counted the number of water molecules within 3.4 Å radii of the atoms in the four hydrogen bond pairs illustrated in Figure 28 namely 7O-11H, 8O-12H, 9O-13H and 10O-14H. Data was acquired from the last 5 ns of the folding simulation. By comparing the protein-water contacts in this region, we hope to shed some light on how the integration of backbone polarization effect may prevent the overestimation of helical motifs in E5 peptide.

Figure 29 (A) Number of water molecules within 3.4 Å radii of the atoms in the four hydrogen bond pairs namely 7O-11H, 8O-12H, 9O-13H and 10O-14H. (B) Schematic representation of backbone hydrogen bonds formed between 8O and 11H ($i+3 \rightarrow i$ hydrogen bonding), 8O and 12 H ($i+4 \rightarrow i$ hydrogen bonding) and 8O and 13H ($i+5 \rightarrow i$ hydrogen bonding) for the representative structure of the most populated cluster achieved through PSBC-based simulation. (C) Variation in the length of backbone hydrogen bonds possibly formed between Ala8 and Ala11 and Ala8 and Ala13 in E5 folded using the PSBC model.

From Figure 29A, simulation performed using the polarizable force field saw more water molecules being situated closer to the aforementioned hydrogen bond pairs compared to that in AMBER simulation with the exception of the second pair, 8O-12H. The proximity of more water molecules to atoms in the hydrogen bond pairings of 7O-11H, 9O-13H and 10O-14H

hindered the formation of ($i+4 \rightarrow i$) hydrogen bonding thus preventing the folding of E5 into an α -helical configuration in the PSBC simulation. This phenomenon emphasized the importance of assimilating backbone polarization effect into force fields as this action may provide us with a more reasonable description of backbone-water interactions which may possibly prevent the overestimation of helical structures observed in the AMBER simulation.

The solvation around atoms in the 8O-12H hydrogen bond pair was comparable in both PSBC and AMBER simulation indicating the aversion of these atoms to form hydrogen bonds with surrounding water molecules. This is understandable in the case of AMBER simulations since α - helix was the preferred conformation and the 8O-12H hydrogen bonding occurred in 63% of the acquired trajectory. (Figure 28B) For simulation conducted under the PSBC model, scrutinization of the most populated structure of the trajectory discerned the formation of three types of hydrogen bonds centering at the backbone oxygen atom of Ala8. Backbone hydrogen bonds formed are between 8O and 11H ($i+3 \rightarrow i$ hydrogen bonding), 8O and 12 H ($i+4 \rightarrow i$ hydrogen bonding) and 8O and 13H ($i+5 \rightarrow i$ hydrogen bonding). (Figure 29B) These backbone hydrogen bonds are considerably stable given that the distances between the donor and acceptor atoms is shorter than 3 Å and these hydrogen bonds persisted upon formation as showed in Figure 28C and 29C. Development of these three hydrogen bonds may explain the preference of E5 to fold into a 3_{10} -helix in the PSBC simulation.

Figure 30 (A) Schematic representation of the representative structure of the most populated cluster acquired through PSBC-based simulation of E5 with possible backbone-sidechain hydrogen bond pairings between Gln 4 and Gln 9 (D1 and D2) and Glu 6 and Gln 9 (D3) labelled accordingly. (B) Schematic representation of the representative structure of the most populated cluster acquired through AMBER-based simulation of E5 with Glu4, Glu6 and Glu9 labelled accordingly. (C) Variation in the distances, D1, D2 and D3 as the simulation progresses for E5 folded using PSBC (blue) and AMBER (red) charges.

Application of the PSBC model in the folding of E5, on the other hand, displayed preference towards 3_{10} -helix instead of the α -helical motif folded in AMBER-based simulation. More than 80% of the trajectory possesses a 3_{10} -helical conformation implying the stable presence of 3_{10} -helix upon folding when the polarizability of the backbone hydrogen bonds are accounted.

Stabilization of the 3_{10} -helix may be attributed to the side-chain-backbone interactions between Gln 4 and Gln 9 and between Glu6 and Gln 9 in the form of hydrogen bonding as suggested by the representative structure of the most populated cluster shown in Figure 28A. Side-chain-backbone interactions has been reported by Spassov et al. to likely possess a stabilizing influence on the native conformation of proteins.³² In fact, side-chain-backbone interactions were proposed to play a more vital role in distinguishing a protein's native state from their misfolded state compared to side-chain-side-chain interactions.³⁴ Hence, to monitor these side-chain-backbone interactions in E5, we have calculated (1) distance between C α atom of Gln 4 and C δ atom of Gln 9 (D1), (2) the distance between C α atom of Gln 9 and C δ atom of Gln 4 (D2) and (3) the distance between C α atom of Glu 6 and C δ atom of Gln 9 (D3). (Figure 30)

Based on Figure 30C, the aforementioned side-chain-backbone interactions are established early during the PSBC-based MD simulations as portrayed by the stabilization of D1, D2 and D3 at less than 5 Å in less than 5 ns. Even though disruption of side-chain-backbone interaction between the side-chain of Gln 4 and the backbone of Gln 9 (D2) was observed during the final 2 ns of the PSBC-based simulation, the side-chain-backbone interactions among Gln 4, Glu 6 and Gln 9 (D1 and D3) were maintained. On the contrary, none of the highlighted side-chain-backbone interactions in Figure AA was formed in AMBER-based simulation (Figure 30C) as the side chains of Gln 4 and Gln 9 preferred to be fully exposed to the surrounding solvent as shown in Figure 30B. The preservation of the side-chain-backbone interactions among

Gln 4, Glu 6 and Gln 9 could be another possible explanation for the favored folding of 310-helix over α -helix in PSBC-based simulations, corroborating the low helicity of E5 in water.

6.3.4 Folding of ES peptide

Figure 31 (A) Schematic representation of three backbone hydrogen bonds formed when ES peptide was folded using standard AMBER charges. (B) Occupancy of the three possible backbone hydrogen bonds formed in ES during the simulation performed using AMBER charges. Zero occupancy of these backbone hydrogen bonds were recorded for PSBC-based simulation.

Since the folding of E5 was conducted in explicit water, we would like to confirm the ability of the PSBC model to fold low helicity peptides in implicit solvation. Hence, we have conducted an additional folding simulation using ES peptide. Likewise, we have examined the activity of backbone hydrogen bonds formed in a $i+4 \rightarrow i$ fashion to compare the preference of the force field used to favor a folded state with higher helicity than that determined experimentally. From the most populated structure achieved through cluster analysis of ES peptide folded under AMBER force field, three backbone hydrogen bonds formed between Glu1 and Glu5, Ser2 and Arg6 and Leu3 and Ile7 were discerned as illustrated in Figure 31A. The formation of these backbone hydrogen bonds were further verified through the high percentages of occurrence observed for the AMBER simulation in Figure 31B. Similar calculation of the percentage occurrence of the aforementioned hydrogen bonds yielded none for the PSBC simulation, corroborating the results achieved through the DSSP plot in Figure 27B and the cluster analyses in Figure 27D that showcased the preference of the PSBC model to assume the turn conformation rather than an α -helix.

Figure 32 (A) Variation in the length of backbone hydrogen bonds possibly formed between 1O-5H, 2O-6H and 3O-7H when ES was folded using AMBER force field. The lengths between the aforementioned atom pairs was also calculated for the simulation performed under PSBC model for comparison. (B) Average hydrogen bond energy calculated for hydrogen bonds formed between 1O-5H, 2O-6H and 3O-7H in ES folded under AMBER force field (red) and PSBC model (blue).

Preference of the AMBER simulation to favor the folding of α -helix was further corroborated in Figure 32A and 32B whereby the lengths of the aforementioned hydrogen bonds were tracked and the average hydrogen bond energy were calculated respectively. Figure 32A showed the stabilization

of the hydrogen bond length of 1O-5H, 2O-6H and 3O-7H to around 3 Å after 7 ns for the first two hydrogen bond pair and after just 2ns for the last pair. (Figure 31C) Additionally, the development of hydrogen bonds between Glu 1 and Glu 5 and Ser 2 and Arg6 coincides with the folding of the α -helix at around 7 ns (DSSP plot in Figure 27B for AMBER simulation). The results achieved for PSBC contrasted that of AMBER. Hydrogen bonds 1O-5H, 2O-6H and 3O-7H were not formed during the simulation (Figure 31B and 31C) hence substantiating the absence of α -helical features in the DSSP plot of the ES peptide folded using the PSBC scheme. These observations may suggest the likelihood of overestimating hydrogen bond strengths when polarization effect was not considered leading to higher helicity than determined experimentally. This was substantiated in Figure 32B whereby stronger hydrogen bonds were observed to have been established in the AMBER simulation while formation of the three backbone hydrogen bonds was not favored in the PSBC simulation.

6.4 Chapter Summary

Ever since the discovery of the relationship between protein structure and its biological function, protein folding became an increasingly popular research field as studies seek to understand the interactions formed or broken during the folding process at the atomic level and how these occurrences affect the protein's biological function. In recent years, there has been an increase in the number of diseases that are related to protein misfolding, such as Alzheimer's disease and Creutzfeldt-Jakob disease. This has prompted a

greater importance to understand protein folding as effective therapeutics are able to be derived.

Protein folding assumes that the folded protein is the lowest energy conformation, which is also known as the native state. The native state is attained via a combination of non-covalent interactions between the amino acids, as well as the interactions with the solvent. In order to model these interactions, MD simulations are employed. Traditional force fields such as AMBER, CHARMM and GROMOS have enabled the application of MD simulations to study protein folding dynamically at the atomic level. However, these force fields currently treat the atomic charges as constants during the course of the simulation. The fundamental understanding of polarization due to the difference in electronegativities has long been established. Therefore, the absence of the polarization effect in the force fields has limited the accuracy of MD simulations significantly.

The PSBC model accounts for the polarizability of proteins by conducting on-the-fly charge updates of atoms participating in backbone hydrogen bond (C, O, N and H) using a mathematical function that relates the hydrogen bond length and charge transferred between hydrogen bond donor and acceptor during formation.

The inclusion of the polarization effect has seen positive outcomes in the folding of α -helices with high helicity. However, a good polarization model should possess the versatility of folding peptides with interactions of varying strength. Therefore, the PSBC model was applied to the ER, E5 and ES peptides as they possess different propensities for helicity. The various analyses

demonstrated clearly that the inclusion of polarization effect is important to fold peptides to their native states. In addition, dissimilar to other polarizable charge models, the PSBC model has shown its versatility to folding α -helices of varying helicity, which remains as a very challenging task for many charge models.²⁹

In conclusion, the PSBC model is able to account for electrostatic polarization in force fields and increase the accuracy in folding peptides to their native states. In addition, the simplicity of the PSBC model resolves the problem of computationally expensive on-the-fly quantum-mechanical calculations. While more rigorous benchmarking is required to validate the PSBC model critically, the model is a good start to more accurate MD simulations, which leads to deeper insights about protein folding.

References

1. Fetrow, J. S.; Giammona, A.; Kolinski, A.; Skolnick, J. The Protein Folding Problem: A Biophysical Enigma. *Curr. Pharm. Biotechnol.* **2002**, 3 (4), 329–347.
2. Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The Protein Folding Problem. *Annu. Rev. Biophys.* **2008**, 37 (1), 289–316.
3. Eaton, W. A.; Muñoz, V.; Hagen, S. J.; Jas, G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J. Fast Kinetics and Mechanisms in Protein Folding. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, 29 (1), 327–359.
4. Daggett, V.; Fersht, A. Opinion: The Present View of the Mechanism of Protein Folding. *Nat. Rev. Mol. Cell Biol.* **2003**, 4 (6), 497–502.

5. Soto, C. Unfolding the Role of Protein Misfolding in Neurodegenerative Diseases. *Nat. Rev. Neurosci.* **2003**, 4 (1), 49–60.
6. Guvench, O.; MacKerell, A. D. Comparison of Protein Force Fields for Molecular Dynamics Simulations; 2008; pp 63–88.
7. Qiang Cui, Markus Meuwly, P. R. *Many-Body Effects and Electrostatics in Biomolecules*; 2016.
8. Sinha, N.; Smith-Gill, S. J. Electrostatics in Protein Binding and Function. *Curr. Protein Pept. Sci.* **2002**, 3 (6), 601–614.
9. Applequist, J.; Carl, J. R.; Fung, K.-K. Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities. *J. Am. Chem. Soc.* **1972**, 94 (9), 2952–2960.
10. van Duijnen, P. T.; Swart, M. Molecular and Atomic Polarizabilities: Thole's Model Revisited. *J. Phys. Chem. A* **1998**, 102 (14), 2399–2407.
11. Elking, D.; Darden, T.; Woods, R. J. Gaussian Induced Dipole Polarization Model. *J. Comput. Chem.* **2007**, 28 (7), 1261–1274.
12. Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization Effects in Molecular Mechanical Force Fields. *J. Phys. Condens. Matter* **2009**, 21 (33), 333102.
13. Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016**, 116 (9), 4983–5013.

14. Friesner, R. A. Modeling Polarization in Proteins and Protein–ligand Complexes: Methods and Preliminary Results; 2005; pp 79–104.
15. Duan, L. L.; Feng, G. Q.; Zhang, Q. G. Large-Scale Molecular Dynamics Simulation: Effect of Polarization on Thrombin-Ligand Binding Energy. *Sci. Rep.* **2016**, 6 (1), 31488.
16. Meuzelaar H, Tros M, Huerta-Viga A, van Dijk CN, Vreede J, Woutersen S. Solvent-Exposed Salt Bridges Influence the Kinetics of α -Helix Folding and Unfolding. *J Phys Chem Lett* **2014**; 5: 900-904.
17. Panman MR, van Dijk CN, Meuzelaar H, Woutersen S. Communication: Nanosecond folding dynamics of an alpha helix: Time-dependent 2D-IR cross peaks observed using polarization- sensitive dispersed pump-probe spectroscopy. *J Chem Phys* **2015**; 142: 041103
18. Rohl CA, Chakrabartty A, Baldwin RL. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Prot Sci* **1996**; 5: 2623-2637.
19. Munoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. *Nat Struc Biol* **1994**; 1: 399-409.
20. Cooke B, Schmidler SC. Statistical prediction and molecular dynamics simulation. *Biophys J* **2008**; 95: 4497-4511.
21. Scholtz JM, Qian H, Robbins VH, Baldwin RL. The energetics of ion-pair and hydrogen- bonding interactions in a helical peptide. *Biochem* **1993**; 32: 9668-9676.

22. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* **1998**; 75: 422-427.
23. Baker, E. N.; Hubbard, R. E. Hydrogen Bonding in Globular Proteins. *Prog. Biophys. Mol. Biol.* **1984**, 44 (2), 97–179.
24. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**;22:2577-2637.
25. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Prot Struc Func Bioinform* **2006**; 65: 712.
26. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS ONE* **2012**; 7: e32131.
27. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* **1992**;13:1011– 1021.
28. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. Multidimensional free- energy calculations using the weighted histogram analysis method. *J Comput Chem* **1995**;16:1339–1350.
29. Best RB, Buchete NV, Hummer G. Are Current Molecular Dynamics Force Field too Helical? *Biophys J* **2008**; 95: L07-L09.

30. Okur A, Wickstrom L, Layten M, Geney R, Song K, Hornak V, Simmerling C. Improved Efficiency of Replica Exchange Simulations Through Use of a Hybrid Explicit/Implicit Solvation Model. *J Chem Theory Comput* **2006**; 2: 420–433.
31. Shao J, Tanner SW, Thompson N, Cheatham TE. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithm. *J Chem Theory Comput* **2007**; 3: 2312-2334.
32. Spassov VZ, Yan L, Flook PK. The dominant role of side-chain backbone interactions in structural realization of amino acid code. ChiRotor: A side-chain prediction algorithm based on side-chain backbone interactions. *Prot Sci* **2007**; 16: 494-506.
33. Vila J, Williams RL, Grants JA, Wojcik J, Scheraga HA. The intrinsic helix-forming tendency of L-alanine. *Proc Natl Acad Sci USA* **1992**; 89: 7821–7825.

Chapter 7

7. Conclusion

MD simulations provided a whole new level in understanding the protein folding process at the atomic level dynamically. However, the use of inaccurate force fields due to the absence of polarizability has limited this understanding significantly. While numerous efforts to include polarizability has been made over the years, most of the models or methods require computationally expensive quantum-mechanical calculations that are conducted on-the-fly. This feature limited the models to only small peptides. The PSBC model is therefore developed, not only to include polarizability in MD simulations, but to also significantly reduce the computational cost while maintaining the accuracy of the MD simulations.

The PHB model has demonstrated its ability to fold α -helices with high helical propensity. Chapter 2 explores the application of PHB model on the mixed-fold protein, FSD-EY. The results have shown that while the PHB model is able to reduce the computational cost by applying mathematical functions that were predefined by quantum-mechanical calculations, it was unable to fold the β -sheet region effectively. This observation indicated the inadequacy of the model and signalled a possible biasness towards α -helices, which could be attributed to the strength of backbone hydrogen bonds being stronger in α -helices with high helical propensity as compared to β -sheets. This study paved a positive direction for the PSBC model to reduce computational cost.

With the PSBC model developed, it was benchmarked against well-known simple β -sheets – Tryptophan zippers (trpzip1 and 2) to find out if the PSBC model is able to fold β -sheets in Chapter 3. The successful folding of the trpzip1 and the various structural and energetic analyses demonstrated that the PSBC model, unlike its predecessors, is able to fold β -sheets. This marks a significant milestone for the PSBC model because the backbone hydrogen bond strength and the cooperativity of β -sheets are weaker than that of α -helices. If the PSBC model is unable to capture weak hydrogen bonds and solely over-strengthens the hydrogen bond strength, then it should have folded the trpzip1 to α -helices. This study therefore concludes that not only is the PSBC model able to reduce computational cost of on-the-fly quantum-mechanical calculations, but the model does not possess biasness towards α -helices and is able to effectively fold β -sheets.

In Chapter 4, the benchmark standard for the PSBC model was raised by applying the model to fold a 3-stranded β -sheet, Beta3s. A 3-stranded β -sheet is harder to fold than a 2-stranded β -sheet such as the trpzip1 because the structure is softer and any significant inaccuracy in the force field will result in incomplete folding of the β -sheet. In addition, if the polarizable charge model possesses biasness towards α -helices, then Beta-3s would have a higher tendency to fold into a α -helix than the trpzip1. The PSBC model is able to fold Beta-3s effectively, reassuring that the model is free of biasness towards α -helices.

The previous chapters have clearly demonstrated various features of the PSBC model: 1. It is able to reduce computational cost significantly by

applying mathematical functions that were predefined by quantum-mechanical calculations instead of conducting these calculations on-the-fly during the MD simulation, 2. Unlike its predecessors, there was no evident biasness towards α -helices, 3. β -sheets has a weaker hydrogen bond network than α -helices, but the PSBC model is able to capture both weak and stronger hydrogen bonds during the simulations. Therefore, after verifying the capability of the PSBC model in folding β -structures, Chapter 5 puts the PSBC model to the test by determining its ability to fold α -helices of various lengths. There have been many simulation studies that reported the successful folding of long α -helices, but none were able to adequately fold weak or short α -helices. The ability of the PSBC model to fold α -helices with varying helical propensities eliminates the possibility of biasness towards α -helices with high helical propensity, as well as reaffirming that the model is versatile in capturing weak backbone hydrogen bonds both in α -helices and β -sheets.

The PSBC model is an exciting model that learns the successes of its predecessors, and resolves the problems that its predecessors faced when they were developed. In addition, the PSBC model is able to fold α -helices and β -sheets with varying propensities, strongly demonstrates that the inclusion of polarizability in force fields is important to the accurate folding of secondary structures and capturing of weak and strong backbone hydrogen bonds effectively.