Understanding Structure and Kinetics of Aβ Monomer and Fibril Ensembles Using Molecular Simulations

by

Sukanya Sasmal

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University Of California, Berkeley

Committee in charge:

Professor Teresa Head-Gordon, Chair
Professor David Wemmer
Professor Ali Mesbah

Spring 2017

Understanding Structure and Kinetics of Aβ Monomer and Fibril Ensembles Using Molecular Simulations

by

Sukanya Sasmal

Abstract

Understanding Structure and Kinetics of Aβ Monomer and Fibril Ensembles Using Molecular Simulations

by

Sukanya Sasmal

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Teresa Head-Gordon, Chair

My doctoral research involves the characterization of the structure, kinetics, and function of amyloid-β (Aβ) proteins by computational means via atomistic and coarse-grained molecular dynamics simulations. Aβ has critical clinical relevance as one of the key hallmarks of Alzheimer's disease pathology. My research has four primary foci. The first of these is studying the properties of the Aβ monomer, an intrinsically disordered protein (IDP), using all-atom simulations. Using a combination of two enhanced sampling techniques - replica-exchange molecular dynamics simulations (REMD) and temperature cool walking (TCW), I have shown that the addition of paramagnetic tags in paramagnetic relaxation enhancement (PRE) experiments of IDPs, perturbs the structural ensembles of the Aβ monomer with an increase in structural order, and the PRE experimental observables are thus not a true representation of the unmodified monomers in solution. Very few experimental techniques can provide residue-specific structural information about IDPs because of their disordered nature, and my work provides valuable insights into the usefulness of commonly used PRE experiments in the IDP field.

IDP structural ensembles are usually generated using REMD simulations with fixed charge protein models. Most computationally generated IDP ensembles using physics-based models are more ordered and compact than expected. The current focus is mostly on improving or modifying parameters in fixed charge force fields to generate more accurate conformational ensembles for IDPs. By comparing different sampling techniques and fixed charge force fields (with and without IDP-specific parameter modification) for the Aβ42 and Aβ43 peptide, I show that the sampling method used to generate the ensemble is equally important as the "correct" force field. The IDP ensembles generated using TCW have better convergence and experimental agreement than the REMD-ensemble for same amount of sampling. Thus, TCW is a better sampling alternative to REMD simulations.

Fixed charge force fields used in IDP simulations are parameterized on folded protein data, and thus predict overly structured and globular configurations for IDPs, which are usually more disordered and solvent exposed compared to folded proteins. Consequently, it is important that the molecular interactions are modeled as accurately as possible during IDP simulations. In chapter 4, using the cationic 24-residue Histatin 5 peptide as a test system, I show that the computationally generated ensemble using the polarizable AMOEBA force field is more consistent with experimental radius of gyration and secondary structure data. Thus, the many-

body polarization effect that is ignored in fixed charge force field is important for simulating IDP systems across a range of solvent-exposed to folded states, capturing the true breadth of structural biology.

The last major emphasis of this dissertation research, is investigating of kinetic elongation mechanisms of amyloid fibril (aggregates of Aβ monomer) using an in-house coarse-grained protein model. In chapter 5, I studied the mechanism of amyloid fibrils elongation via binding of monomers from solution, and demonstrated that the monomer structure only influences the kinetics, but not the overall binding mechanism. This result provides a fundamental understanding of growth of amyloid fibrils.

To my parents Dr. Ratna Sasmal Nag and Bhim Charan Sasmal

Contents

1. Introduction
1.1 Introduction 1
1.2 References 5
2. The Effect of a Paramagnetic Spin Label on the Intrinsically Disordered Peptide Ensemble of Amyloid- β
2.1 Introduction 8
2.2 Methods 10
2.3 Results 11
2.4 Discussion 18
2.5 Conclusion 19
2.6 References 20
2.7 Appendix 25
3. Comparison of enhanced sampling techniques and force fields for IDP simulations
3.1 Introduction 31
3.2 Methods 32
3.3 Results 33
3.4 Discussion and Conclusions 36
3.5 References 37
3.6 Appendix 40
4. Improved Sampling of Disordered Peptide Structural Ensembles Using the Polarizable AMOEBA Force Field
4.1 Introduction 44
4.2 Methods 45
4.3 Results 46
4.3 Discussions and Conclusions 48
4.4 References 48
5. Mechanism of Nucleation and Growth of Aβ40-Fibrils from All-Atom and Coarse-Grained Simulations
5.1 Introduction 51
5.2 Methods 54
5.3 Results 58
5.4 Discussions and Conclusions 67
5.5 References 67
5.6 Appendix 72

Acknowledgements

I would like to thank my advisor Prof. Teresa Head-Gordon for mentoring me throughout my entire graduate career at Berkeley. I would also like to thank Dr. Aurelia Ball, former graduate student in the lab who helped me to get started with my project. I had the opportunity to work closely with Dr. Nadine Schwierz-Neumann during her postdoctoral position at Berkeley and learned a lot from the scientific discussions. I also collaborated with fellow lab members James Lincoff, Eugene Yedvabny and Dr. Saurabh Belsare on some of my projects and I appreciate their help in troubleshooting the projects. I also want to extend my thanks to my committee members Prof. David Wemmer and Prof. Ali Mesbah for reviewing my dissertation draft.

I would like to acknowledge Manindra Manna and Dr. PK Banerjee from R&D, Tata Steel where I had my first research experience and who started me down the path of research. My undergraduate advisor Dr. Sunanda DasGupta has also played a crucial role in encouraging me to continue with higher studies.

I owe most of my knowledge to UC Berkeley and my alma matter, Indian Institute of Technology, Kharagpur, where I had the privilege to interact with brilliant students and learn from the best teachers. I have thoroughly enjoyed taking courses and learning from the difficult assignments. I would also like to acknowledge different free online resources I have used during my entire student life - in particular Wikipedia and Stack Overflow. I learnt a lot of new concepts and skills, and it is be hard to imagine a world without these resources.

I have been fortunate enough to have lab members, who have become my family at Berkeley - Dr. Asmit Bhowmick, Dr. Saurabh Belsare, Dr. Marielle Soniat and Alex Albaugh (soon to be Dr.!). I can't express enough gratitude to all of them - Saurabh for listening to everything I had to say; Alex for teaching with me four times and sharing office space; Marielle for being like a big sister to me and watching out for me; and Asmit for being a close friend during my undergraduate and graduate studies, and for all of the endless help and guidance. I am also thankful to my undergraduate friends (Deepashree, Varsha, Radhika, Pratiti, Satyaki, Debmalya), high-school friends (Anjali, Sanhita, Beas, Madhurima) and friends from the department (Katerina, Neelay and Diego) for their friendship and support.

Lastly, I would like to mention my parents who have been a constant support throughout my life. They have always encouraged me to pursue studies and to strive for the best. Their support and encouragement has been the most priceless thing in my life. I am also grateful for my maternal aunt and uncle for their love and support.

Chapter 1

Introduction

1.1 INTRODUCTION

Alzheimer's disease is a progressive neurodegenerative disease accounting for 60-80% cases of senile dementia[1]. 2017 estimates by the Alzheimer's Association puts the number of Americans suffering from the disease to be around 5.5 million and the figure is projected to be around 13.8 million in 2050s. More than a century has passed since the first documentation of Alzheimer's disease by Alois Alzheimer in 1907[2] and still we do not have a complete cure for the disease. A central feature of Alzheimer's disease is the presence of extracellular deposits in the brain. Almost 80 years after Alois Alzheimer's findings, researchers Glenner and Wong[3, 4] and Masters et al.[5] were able to identify the amyloid-β (Aβ) peptide as the major molecular constituent (~90%) of these extracellular deposits, also known as amyloid plaques.

The Aβ peptide is a short 39-43 residue peptide[6], with known primary sequence (Table 1) and is formed by the cleavage of the amyloid precursor protein (APP), an integral trans-membrane protein by the enzymes β- and γ- secretase[7]. Since the γ-secretase can act on multiple sites, the resulting proteolytic product (Aβ monomer) varies in length.

Table 1.1: One-letter amino acid sequence of different Aβ monomers. The differences are highlighted in red.

$A\beta_{40}$ sequence	DAEFRHDSGY EVHHQKLVFF AEDVGSNKGA IIGLMVGG
$A\beta_{42}$ sequence	DAEFRHDSGY EVHHQKLVFF AEDVGSNKGA IIGLMVGGVV
$A\beta_{43}$ sequence	DAEFRHDSGY EVHHQKLVFF AEDVGSNKGA IIGLMVGGVVT

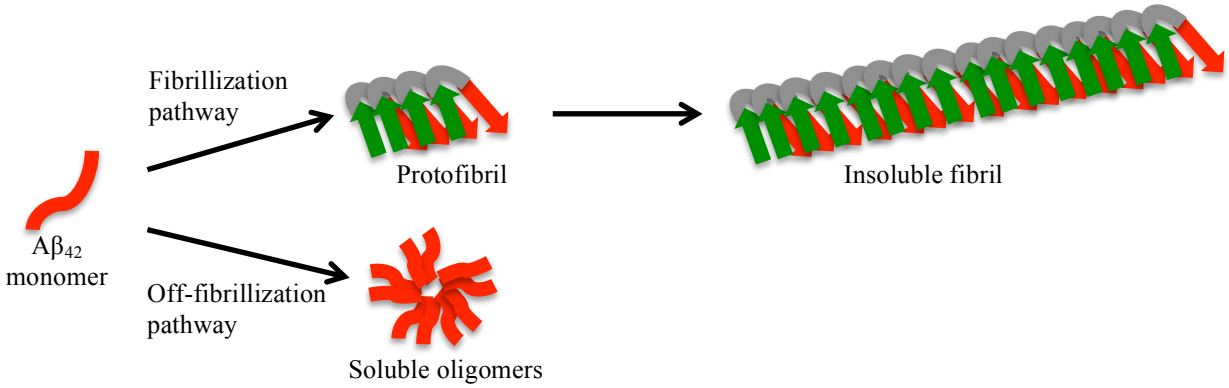


Figure 1.1: *The two main pathways for Aβ aggregate formation*[8]. 1) the fibrillization pathway forming the fibril with intermolecular cross β-sheet structure; and 2) the off-fibrillization pathway resulting in the formation of soluble oligomers, which are presumed to be less ordered and smaller in size as compared to the mature fibrils.

The peptide monomer is non-toxic and is prone to aggregation. There are two primary aggregation pathways[8] as shown in Figure 1.1: 1) the fibrillization pathway that results in the

formation of stable well-structured insoluble fibrils that are polymorphic and 2) the off-fibrillization pathway leading to the formation of smaller soluble oligomers. Overall the current consensus is that the soluble oligomers are the main contributor to the disease pathogenesis[9, 10], one reason being that Alzheimer's symptoms tend to correlate poorly with insoluble plaque burden. However the lack of disease correlation with plaque burden may arise from the fact that only certain fibril polymorphs are cytotoxic while others are benign[11]. Therefore an in-depth understanding of each of these three forms of the Aβ peptide – monomer, fibrils and oligomers is needed to have a holistic view of the molecular mechanisms at play in Alzheimer's disease, and each of them provides unique challenges in complete structural characterization of the different molecular forms. This dissertation explores the interplay between experimental and computational techniques (chapter 2-4) employed to characterize the Aβ monomer structural ensembles that seed oligomeric and fibril forms, and also different Aβ fibril elongation mechanisms by monomer addition (chapter 4). The main goal of this work is to contribute towards basic biophysical understanding of the structure and aggregation mechanisms of the Aβ peptide, that could inform therapeutic strategies to reverse the tragic effects of this brain disorder.

Intrinsically Disordered Peptides

The Aβ monomer is about 4 kDa in size and is classified as an intrinsically disordered peptide (IDP)[12]. IDPs do not have a single stable structure, and thus are very different from the standard folded proteins whose structural landscape can be described by a single funnel-shaped free energy basin[13]. IDPs generally sample an ensemble of structures, with a free energy landscape that is more flat as compared to folded proteins and contains multiple energetically competitive local minimum. Figure 1.2 shows a representative structure from the Aβ42 structural ensemble, which has minimal secondary structural features.

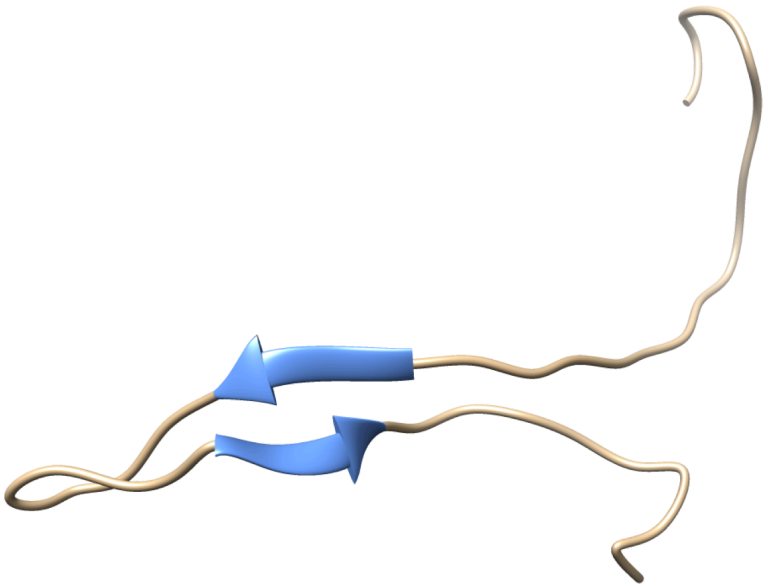


Figure 1.2: *A representative structure from the Aβ42 structural ensemble with the blue region showing the signature β-strand of Aβ42 ensemble formed by residues 18-20 and 31-33.* Molecular graphics was generated using UCSF Chimera[14].

Analysis of Paramagnetic relaxation enhancement (PRE) experiments for IDPs

Traditional protein structure determination techniques like X-ray crystallography fail in the free form of the IDP because of the absence of a single equilibrium structure[12]. On the other hand, nuclear magnetic resonance spectroscopy (NMR) has been shown to be very successful in probing the atomic level detail of the IDP structural ensemble in the native conditions, with the caveat that the experimental NMR observables report on ensemble-averaged properties[15, 16]. This is because in solution, the time-scale for inter-conversion between different transient IDP structures is usually faster than the nanoseconds-milliseconds timescale of the NMR experiments[16]. Since the structural sub-populations of the free IDP monomer ensemble are informative about oligomerization or aggregation mechanism, the characterization of the IDP structural ensemble is very relevant to the disease and requires a combination of different solution based techniques. For NMR, the data acquisition would include J-couplings that report on mostly backbone dihedral angles, chemical shifts giving information about local electronic environments, residual dipolar couplings for relative orientation of different peptide regions with respect to an external magnetic field, and nuclear Overhauser effect (NOEs) and paramagnetic relaxation enhancement (PRE) experiments that provide non-bonded long range contacts up to 5 Å and 20 Å, respectively[16].

PRE experiments have proven very powerful for study of folded protein structure in solution[17-19] because of their ability to provide residue-specific tertiary information. The PRE experiment requires the attachment of a paramagnetic spin label to the biomolecule/protein via chemical modifications, which for folded proteins has been shown to have a negligible effect on the peptide secondary and tertiary structure[20-22] and thus the structural properties remain similar to the unmodified peptide conditions. While in principle PRE experiments could be used to generate or refine IDP structural ensembles[23], equivalent control studies for the disordered peptides do not exist in the literature. Because of the structural flexibility of the IDPs, there always exist a possibility that PRE tags, which are usually hydrophobic in nature[24], will perturb the structural ensemble. In chapter 2, we try to fill this gap in the literature by investigating the effect of the addition of a PRE tag using molecular simulations of the disordered Aβ42 peptide. Our simulations show that the paramagnetic tag does affect the structure of the Aβ42 peptide, with a significant increase in the β-strand content of the peptide in the C-terminus after tag introduction.

Developing New Computational Approaches for IDPs

Because the experimental data is still underdetermined for characterizing the IDP structural ensemble[16], computational approaches are a critical partner for driving IDP ensemble construction. The two important aspects for computational study of IDP systems are the force field and the sampling method. Since IDPs have relatively flat energy landscape as compared to folded proteins, standard molecular dynamics are not ideal since it takes excruciatingly long time to sample the entire energy landscape, a problem that is further exacerbated by use of advanced force fields that add significantly to the computational cost. Therefore the computational approaches to IDPs have relied on the most affordable enhanced sampling replica exchange molecular dynamics[25] (REMD) technique that is combined with standard two-body (fixed charge) force fields.

However, since fixed charge force fields are parameterized based on folded protein data, they are not ideal for IDP simulations and the resulting ensembles are usually more collapsed and structured than they should be. To better model the disordered nature of the IDPs, recently a

few new IDP-specific force fields have been developed[26-28], which take into account the increased protein-water interactions in an IDP by re-parameterizing the van der Waals parameters to be more favorable or by modifying the protein backbone torsional terms to make the protein backbone less structured. Even though these modified force fields can perform better than the original fixed charge force fields[29], there is still scope for more improvement in ensemble prediction by going to many-body potentials that are transferable –i.e. that work for both folded protein states as well as IDPs.

In chapter 3, we explore the interplay between sampling techniques and force fields by comparing Aβ42 and Aβ43 ensembles to the best available experimental data. The results indicate that the effect of the sampling technique is just as pronounced as the force field on the generated ensemble. In particular there is significant improvement in switching from REMD to a new sampling technique we have developed called Temperature Cool Walking[30, 31] (TCW), which we show converges faster to a more correct limiting ensemble based on experimental data. Although using IDP-specific force fields[26-28] results in some improvement, the magnitude of this improvement is relatively less than the better sampling TCW technique, and as we show are not in quantitative agreement on a well-characterized test system Histatin-5. In chapter 4, I explore the polarizable AMOEBA[32] force field as an alternative force field for simulating IDP systems, to show that AMOEBA gives better accuracy in ensemble prediction for IDPs without any parameter modifications.

Aβ fibril stability and elongation mechanisms

Aβ fibrils are very structured in nature[33], unlike the peptide monomer, which lacks a single dominant structure. The main characteristics of Aβ fibrils are intermolecular cross-β sheets, arranged in a single-stack or 2-stack or 3-stack arrangement. Side-chain hydrophobic interactions are responsible for holding the β sheets together in the fibril[34]. Aβ fibrils are polymorphic in nature with different structures arising depending on the growth conditions[35] both *in vivo* and *in vitro*. As of today, only one amyloid fibril structure has been identified which is nucleated from the brain tissues of a diseased Alzheimer's patient[11] (Figure 3).

Understanding the stability and elongation mechanism of this particular polymorph is very important in regard to the disease pathology, since there is a good possibility that this particular polymorph might be cytotoxic in humans. In chapter 5, the stability of this disease-polymorph of Aβ fibril and its aggregation mechanisms are examined using a C_α-based coarse-grained protein model[36]. The main conclusion from this part is that at least 4 cross-sections or 12 peptide chains are required to have a stable fibril for this particular polymorph; and that the primary fibril elongation mechanism by monomer addition is a two-step mechanism, consisting of a fast "docking" step, when the free monomer randomly attaches itself to the fibril, followed by a slower "locking" step, where the docked monomer goes through conformational and orientation rearrangement to form the correct native contacts present in a mature fibril[37].

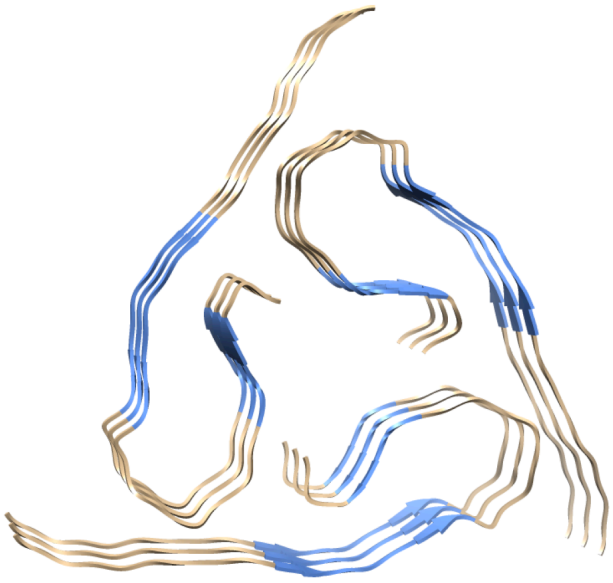


Figure 3: A structural model (PDB Id: 2M4J) for amyloid fibril with three-fold symmetry[11]. The blue regions highlight the intermolecular β-sheets present in the fibril. Molecular graphics was generated using UCSF Chimera[14].

1.2 REFERENCES

1. Alzheimer's Association, 2017 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia* **2017**.
2. Alzheimer, A., Uber Eine Eigenartige Erkrankung Der Hirnrinde. *Allgemeine Zeitschrife Psychiatrie* **1907,** *64*, 146-148.
3. 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. Comm.* **1984,** *120* (3), 885-890.
4. Glenner, G. G.; Wong, C. W., Alzheimer's Disease and Down's Syndrome: Sharing of a Unique Cerebrovascular Amyloid Fibril Protein. *Biochemical and biophysical research communications* **1984,** *122* (3), 1131-1135.
5. Masters, C. L.; Multhaup, G.; Simms, G.; Pottgiesser, J.; Martins, R.; Beyreuther, K., Neuronal Origin of a Cerebral Amyloid: Neurofibrillary Tangles of Alzheimer's Disease Contain the Same Protein as the Amyloid of Plaque Cores and Blood Vessels. *EMBO J.* **1985,** *4* (11), 2757.
6. Antzutkin, O. N., Amyloidosis of Alzheimer's A β Peptides: Solid‐State Nuclear Magnetic Resonance, Electron Paramagnetic Resonance, Transmission Electron Microscopy, Scanning Transmission Electron Microscopy and Atomic Force Microscopy Studies. *Mag. Res. Chem.* **2004,** *42* (2), 231-246.
7. Selkoe, D. J., Alzheimer's Disease: Genes, Proteins, and Therapy. *Physio. Rev.* **2001,** *81* (2), 741-766.
8. Glabe, C. G., Structural Classification of Toxic Amyloid Oligomers. *Journal of Biological Chemistry* **2008,** *283* (44), 29639-29643.

9. Walsh, D. M.; Klyubin, I.; Fadeeva, J. V.; Cullen, W. K.; Anwyl, R.; Wolfe, M. S.; Rowan, M. J.; Selkoe, D. J., Naturally Secreted Oligomers of Amyloid B Protein Potently Inhibit Hippocampal Long-Term Potentiation in Vivo. *Nature* **2002,** *416* (6880), 535-539.
10. Haass, C.; Selkoe, D. J., Soluble Protein Oligomers in Neurodegeneration: Lessons from the Alzheimer's Amyloid B-Peptide. *Nature Rev. Mol. Cell Bio.* **2007,** *8* (2), 101-112.
11. Lu, J.-X.; Qiang, W.; Yau, W.-M.; Schwieters, C. D.; Meredith, S. C.; Tycko, R., Molecular Structure of β-Amyloid Fibrils in Alzheimer's Disease Brain Tissue. *Cell* **2013,** *154* (6), 1257-1268.
12. Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T., Differences in B-Strand Populations of Monomeric Amyloid-B 40 and Amyloid-B 42. *Biophys. J.* **2013,** *104* (12), 2714-2724.
13. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G., Funnels, Pathways and the Energy Landscape of Protein Folding: A Synthesis. **1994**.
14. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., Ucsf Chimera—a Visualization System for Exploratory Research and Analysis. *Journal of computational chemistry* **2004,** *25* (13), 1605-1612.
15. Kosol, S.; Contreras-Martos, S.; Cedeño, C.; Tompa, P., Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* **2013,** *18* (9), 10802-10828.
16. Ball, K. A.; Wemmer, D. E.; Head-Gordon, T., Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-β Peptides. *J. Phys. Chem. B* **2014,** *118*, 6405-6416.
17. Qu, K.; Vaughn, J. L.; Sienkiewicz, A.; Scholes, C. P.; Fetrow, J. S., Kinetics and Motional Dynamics of Spin-Labeled Yeast Iso-1-Cytochrome C: 1. Stopped-Flow Electron Paramagnetic Resonance as a Probe for Protein Folding/Unfolding of the C-Terminal Helix Spin-Labeled at Cysteine 102. *Biochemistry* **1997,** *36* (10), 2884-97.
18. Bertini, I.; Donaire, A.; Jimenez, B.; Luchinat, C.; Parigi, G.; Piccioli, M.; Poggi, L., Paramagnetism-Based Versus Classical Constraints: An Analysis of the Solution Structure of Ca Ln Calbindin D-9k. *J. Biomol. NMR* **2001,** *21* (2), 85-98.
19. Parigi, G.; Bertini, I.; Luchinat, C.; Piccioli, M.; Piva, A., Paramagnetic Constraints in Structure-Determination Programs. *J. Inorg. Biochem.* **2001,** *86* (1), 507-507.
20. Koteiche, H. A.; McHaourab, H. S., Folding Pattern of the A-Crystallin Domain in Aa-Crystallin Determined by Site-Directed Spin Labeling. *J. Mol. Biol.* **1999,** *294* (2), 561-77.
21. Bertini, I.; Luchinat, C.; Parigi, P., *Solution NMR of Paramagnetic Molecules: Applications to Metallobiomolecules and Models*. Elsevier: Amsterdam, 2001; Vol. 2.
22. Murzyn, K.; Rog, T.; Blicharski, W.; Dutka, M.; Pyka, J.; Szytula, S.; Froncisz, W., Influence of the Disulfide Bond Configuration on the Dynamics of the Spin Label Attached to Cytochrome C. *Proteins: Struct., Func., Bioinform.* **2006,** *62* (4), 1088-100.
23. Marsh, J. A.; Forman-Kay, J. D., Ensemble Modeling of Protein Disordered States: Experimental Restraint Contributions and Validation. *Proteins: Struct., Func., Bioinform.* **2012,** *80* (2), 556-572.
24. Fawzi, N. L.; Fleissner, M. R.; Anthis, N. J.; Kálai, T.; Hideg, K.; Hubbell, W. L.; Clore, G. M., A Rigid Disulfide-Linked Nitroxide Side Chain Simplifies the Quantitative Analysis of Pre Data. *J. Biomol. NMR* **2011,** *51* (1-2), 105-114.
25. Sugita, Y.; Okamoto, Y., Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999,** *314* (1), 141-151.

26. Best, R. B.; Mittal, J., Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *J. Phys. Chem. B* **2010,** *114* (46), 14916-14923.
27. Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E., Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015,** *119* (16), 5113-5123.
28. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D., Charmm36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2016**.
29. Henriques, J.; Skepö, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: On the Accuracy of the Tip4p-D Water Model and the Representativeness of Protein Disorder Models. *J. Chem. Theory Comput.* **2016,** *12* (7), 3407-3415.
30. Brown, S.; Head‐Gordon, T., Cool Walking: A New Markov Chain Monte Carlo Sampling Method. *J. Comput. Chem.* **2003,** *24* (1), 68-76.
31. Lincoff, J.; Sasmal, S.; Head-Gordon, T., Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom. *J. Chem. Phys.* **2016,** *145* (17), 174107.
32. Ren, P.; Ponder, J. W., Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003,** *107* (24), 5933-5947.
33. Tycko, R., Solid-State Nmr Studies of Amyloid Fibril Structure. *Ann. Rev. Phys. Chem.* **2011,** *62*, 279-299.
34. Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R., A Structural Model for Alzheimer's B-Amyloid Fibrils Based on Experimental Constraints from Solid State Nmr. *Proc. Natl. Acad. Sci.* **2002,** *99* (26), 16742-7.
35. Petkova, A. T.; Leapman, R. D.; Guo, Z.; Yau, W.-M.; Mattson, M. P.; Tycko, R., Self-Propagating, Molecular-Level Polymorphism in Alzheimer's S-Amyloid Fibrils. *Science* **2005,** *307* (5707), 262-265.
36. Yap, E. H.; Fawzi, N. L.; Head‐Gordon, T., A Coarse‐Grained Α‐Carbon Protein Model with Anisotropic Hydrogen‐Bonding. *Proteins: Struct., Func., Bioinform.* **2008,** *70* (3), 626-638.
37. Sasmal, S.; Schwierz, N.; Head-Gordon, T., Mechanism of Nucleation and Growth of Aβ40 Fibrils from All-Atom and Coarse-Grained Simulations. *J. Phys. Chem. B* **2016,** *120* (47), 12088-12097.

Chapter 2

The Effect of a Paramagnetic Spin Label on the Intrinsically Disordered Peptide Ensemble of Amyloid- β

Paramagnetic relaxation enhancement (PRE) experiments is an NMR technique that has yielded important insight into the structure of folded proteins, although the perturbation introduced by the large spin probe might be thought to diminish its usefulness when applied to characterizing the structural ensembles of intrinsically disordered proteins (IDPs). We compare the computationally generated structural ensembles of the IDP amyloid-β42 (Aβ42) to an alternative sequence in which a nitroxide spin label attached to cysteine has been introduced at its N-terminus. Based on this internally consistent comparison, we find that the spin label does not perturb the signature population of the β-hairpin formed by residues 16-21 and 29-36 that is dominant in the Aβ42 reference ensemble. However the presence of the tag induces a strong population shift in a subset of the original Aβ42 structural sub-populations, including a six-fold enhancement of the β-hairpin formed by pairing between residues 27-31 and 33-38 and an increase in turn content in residues 1-3. Through back-calculation of NMR observables from the computational structural ensembles, we show that the structural differences between the labeled and unlabeled peptide would be evident in local residual dipolar couplings, and possibly differences in homonuclear $^{1}H-^{1}H$ NOEs and heteronuclear $^{1}H-^{15}N$ NOEs if the paramagnetic contribution to the longitudinal relaxation does not suppress the NOE intensities in the real experiment. This work shows that molecular dynamics simulation provides a complementary approach to resolving the potential structural perturbations introduced by reporter tags that are needed for the PRE, DEER and FRET experiments applied to IDPs.

2.1 INTRODUCTION

Intrinsically disordered proteins (IDPs) are an important class of proteins that play a significant role in cellular function as well as deleterious roles in disease[1-5]. They confound the structure-function paradigm since they do not have a single dominant tertiary structure but instead sample multiple tertiary conformations in solution. Paramagnetic relaxation enhancements (PRE) experiments are a widely used NMR technique employed to provide information about long-range order in both folded proteins[6-8] and more recently to IDP structural ensembles[9-19]. It requires the introduction of a nitroxide spin label into the peptide of interest via a covalent bond to a cysteine residue, a technique commonly known as site directed spin labeling[20]. If the protein doesn't have a cysteine residue, then cysteine is added to the sequence either as a point mutation or as an additional residue. The benefit of adding the label is because the unpaired electron spin in the PRE tag causes additional dipole-dipole interactions that result in line broadening in the NMR spectrum, from which information about distances up to 25 Å between the probe and a desired site can be extracted.

Of course the PRE tag attached to the cysteine residue is a relatively large side-group perturbation relative to the original sequence, whose structural consequences must be carefully quantified to extract useful structural information about the unlabeled system. For well-folded proteins, most available sequence sites are able to accommodate the addition of the spin label without any significant structural changes[6, 7, 21-23]. On the other hand, IDPs are remarkably

different from well-folded proteins because of their structural plasticity, such that it might be expected that point mutations, an addition of a single residue, or introduction of a spin label, might bring about extensive structural changes in the IDP ensemble. At present, however, most experiments on IDPs are analyzed under the assumption that the probe has only a localized effect on structure and dynamics, and that the corresponding long-range features of the untagged ensemble are minimally perturbed. The motivation behind this work is to have a better understanding of the effect of probe attachment to an IDP based on a comparison of its structural ensemble to the same IDP structural ensemble without a label, so that the PRE experiments on IDPs are interpreted correctly.

Here we study the effects of the PRE spin label on the structural dynamics of the amyloid-β peptide, Aβ42, a 42-residue peptide that is the major molecular player in Alzheimer's disease[24, 25]. It has been widely studied both experimentally[15-17, 24, 26-29] and computationally[30-43] and it has been shown that the monomeric form of Aβ42 is classified as an IDP, sampling an extensive set of conformations in solutions. We chose Aβ42 for our study because our previous computational work has characterized its structural ensemble extensively[41-43], and we have made thorough comparisons to previously published experimental and computational work of others[27, 32, 44-50]. In this work we have combined two different sampling techniques - Replica Exchange Molecular Dynamics[51] (REMD) and Temperature Cool Walking[52, 53] (TCW) in order to generate structural ensemble averages with error estimates that take into account sampling uncertainty. We have performed two independent simulations of the untagged monomeric Aβ42 peptide and a tagged version of the Aβ42 peptide in which we add a cysteine residue to the N-terminus to which we attach the methanethiosulfonate spin label (MTSL) (Figure 2.1), a commonly used spin label for PRE experiments. Any structural differences between the computationally generated ensembles of Aβ42 and MTSL-Cys-Aβ42 peptides that fall outside sampling uncertainty would then allow us to determine whether the spin label changes the IDP structural characteristics relative to the reference ensemble in any significant way.

From this internally consistent computational comparison we find that most of the important structural features of the untagged Aβ42 ensemble have been preserved, such as the dominant β-hairpin between the central hydrophobic cluster comprising residues 16-21 with the C-terminus residues 29-36[41-43] which is likely critical for subsequent fibril formation. However, the spin label introduces strong population shifts toward greater enhancement of what were originally lightly populated β-hairpins in the C-terminal region, including a six-fold enhancement of the β-hairpin formed by residues 27-31 and 33-38, that is largely driven by the hydrophobic character of the MTSL label. Through back-calculation of many types of NMR observables, our computational results would be supported by differences in residual dipolar couplings, and possibly differences in homonuclear $^{1}H-^{1}H$ NOEs and heteronuclear $^{1}H-^{15}N$ NOEs, if the paramagnetic contribution to the longitudinal relaxation does not suppress the NOE intensities of theses nuclei in the real experiment. We conclude that PRE experiments can be used to answer structural and mechanistic questions in regards IDP ensembles, when carefully interpreted with the aid of additional experiments combined with high quality molecular simulations.

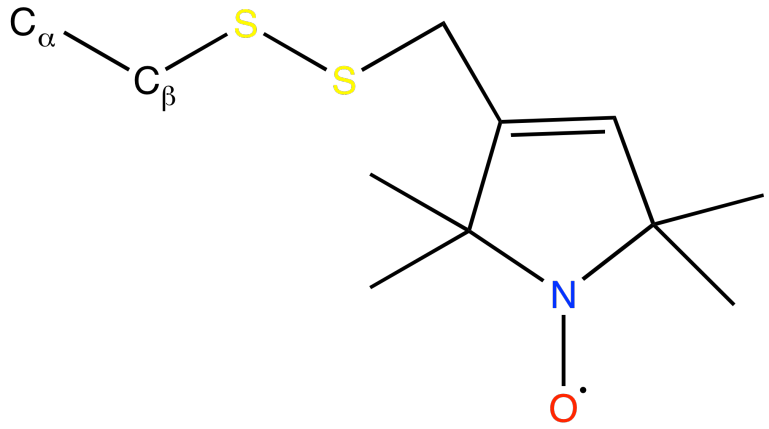


Figure 2.1: *Schematic of methanethiosulfonate spin label (MTSL)*. MTSL is a commonly used PRE tag because of its sulfhydryl-specificity and side-chain flexibility. The paramagnetic center is delocalized between the nitrogen (shown in blue) and the oxygen atom (shown in red).

2.2 METHODS

Here we report briefly on the simulation protocol and analysis methods; further details can be found in previous publications[41-43, 52].

REMD Simulation Protocol. All simulation set up and production runs utilized Amber14[54]. For addition of the PRE-tag, a Cys residue was first added to the N-terminus of the Aβ42 sequence, and the MTSL spin label was then attached through a disulfide bond at the C_β position of Cys1 of the Cys-Aβ42 peptide. The peptides were modeled using the Amberff99sb force field[55, 56] and the water molecules using the TIP4P-Ew water model[57]. The parameters for the MTSL tag were generated using the CHAMBER module of Amber by Xue et al[58], a protocol that yields a consistent force field with the standard unlabeled peptide.

We created two extended configurations for Aβ42 and the MTSL-Cys-Aβ42 peptides, and the peptides were subsequently minimized. The initial simulation boxes were ~700 nm^3 in volume, and 3 and 4 Na+ atoms were added to the simulation box to neutralize the peptide for the untagged Aβ42 and MTSL-Cys-Aβ42 respectively. The starting configurations were heated to 287 K at constant volume and equilibrated for 2 ns under constant pressure of 1 bar, to achieve the correct density. This was followed by a 2 ns high-temperature simulation at 500 K to obtain a more collapsed peptide. Two distinct collapsed states were chosen for each peptide to start two independent REMD simulations. The peptides were resolvated to obtain cubic simulation boxes of 239 and 257 nm^3 for Aβ42 and MTSL-Cys-Aβ42 peptides respectively. The final number of water molecules for the Aβ42 and MTSL-Cys-Aβ42 boxes were 7800 and 8372 respectively.

The REMD module of Amber14 molecular dynamics package was used to generate the structural ensembles. 58 replicas were used in the temperature range 287-450K with an exchange probability of 18-22% and an exchange attempt of 0.5 ps. The production runs were performed in NVT ensemble using a 1 fs time-step and constraints on the heavy atom hydrogen covalent bonds. A Langevin thermostat was used to regulate the temperature. Long- range electrostatic forces were calculated using particle mesh Ewald with a 9.0 Å cutoff for the real space electrostatics and Lennard-Jones forces. The total length of each simulation was 100 ns with the first 50 ns being discarded as equilibration; the final Boltzmann weighted ensemble consisted of 10000 structures from the final 50 ns of the 2 independent equilibrated simulations at 287 K.

TCW Simulation Protocol: TCW is a non-equilibrium alternative to REMD, using only two temperature replicas to generate a Boltzmann-weighted ensemble[53]. Structures from the high

temperature are sequentially cooled to the low temperature and detailed balance is satisfied based on corresponding heating runs that together comprise a complete Metropolis acceptance criterion. TCW has been shown in a previous publication to converge faster for small peptides such as alanine dipeptide, met-enkephalin, and the Aβ42 peptide[52]. The same equilibrated starting structures for REMD simulations were chosen for the TCW method. The high and low temperature replicas were at 456.2 K and 287 K with 50 intermediate annealing steps. Temperature was maintained using an Andersen thermostat[59]. Long-range electrostatic forces were calculated using Ewald, with a cutoff of 9.5 Å for the real space electrostatics and Lennard-Jones forces. The frequency of exchange attempt was 500 fs. For Aβ42, two independent simulations were performed for 100 ns with the first 10 ns being discarded as equilibration and for MTSL-Cys-Aβ42, two simulations for 60 ns each was performed with the first 10 ns being discarded again.

Trajectory Analysis. Both the cpptraj module of Amber and our own in-house codes for evaluating NMR observables and ensemble properties were used for analyzing the structural ensemble properties of Aβ42 and MTSL-Cys-Aβ42. For the analysis, the 1st cysteine residue of the modified MTSL-Cys-Aβ42 peptide will be referred as Residue 0. The radius of gyration (R_g) was calculated using backbone heavy atoms for residues Asp1-Ala42; secondary structures were assigned using the DSSP criterion; hydrogen bonds and salt-bridges were calculated using a distance cutoff of 3.5 Å and 4.0 Å between heavy atoms respectively and a 60° angle cutoff for both.

Calculation of NMR Observables. Full details for back calculation of NMR observables have been reported in previous publications, including chemical shifts from ShiftX2[60], J-coupling constants[61], Residual Dipolar Couplings (RDCs) based on local[62] and global alignments[63], and ^{1}H-^{1}H and ^{1}H-^{15}N NOEs[41, 42] from the 287 K structural ensembles and MD correlation[41, 42] times for both the tagged and untagged peptide.

2.3 RESULTS

The characterization of an IDP structural ensemble is a highly underdetermined problem when compared to that of the folded protein class, and thus it is paramount to quantify the errors in computational reproducibility through independent ensemble calculations of the same IDP sequence. This is necessary to ensure that a fair comparison is made as to whether there are meaningful structural ensemble differences between the IDP Aβ42 and MTSL-Cys-Aβ42 peptides that arise from perturbations due to the spin label vs. what arises from intrinsic uncertainties due to sampling limitations of the IDP ensemble. Although errors in force fields may contribute to direct agreement with experiment, use of the same protein and water force fields (Amberff99sb force field[55, 56] and TIP4P-Ew water model[57]) provides an internally consistent comparison among the computationally generated ensembles.

We therefore have performed two independent *de novo* MD simulations of the Aβ42 and MTSL-Cys-Aβ42 peptides using standard REMD[64-66] and also two independent TCW simulations of the two peptides in this work. Together these independent ensembles provide a measure of reproducibility between computationally generated IDP ensembles, such that the standard deviation among them defines the uncertainty due to the sampling protocol. Having thus defined the calculated model uncertainties as a standard deviation in all properties of the Aβ42 reference ensemble, any structural ensemble differences that are larger than this intrinsic sampling error when comparing against the MTSL-Cys-Aβ42 ensemble would thus define a predicted perturbation to the structural sub-populations due to the spin label.

Figure 2.2 shows the normalized probability density distributions of the radius of gyration (R_g) in the Aβ42 and MTSL-Cys-Aβ42 ensemble. The mean and STD values in the R_g for the Aβ42 averaged ensembles and the MTSL-Cys-Aβ42 ensembles are 12.6 ± 1.7 Å and 11.6 ± 1.5 Å respectively. All untagged Aβ42 ensembles are qualitatively the same in the sense that each exhibits a longer tail in the R_g distribution compared to the tagged peptide. The MTSL-Cys-Aβ42 ensemble has slightly more compact distribution than the Aβ42 ensemble, but there is enough uncertainty in the R_g estimates that we need to consider other types of analysis of their structural characteristics to quantify the effect of the spin label on the observed sub-populations of structure.

Figure 2.3 shows the calculated tendencies by residue for each peptide to form helical structures (α-helix, 3-10 helix and π-helix), β-structure (intramolecular β-bridges, β-hairpins or β-sheets), as well as localized turns (with and without hydrogen-bonds), as averages over their conformational ensembles. Table 1 provides the subset of secondary structure classifications in Figure 2.3 that are supported by internal hydrogen-bonds (according to DSSP[67]), which would likely comprise secondary structure sub-populations that are typically more populated due to greater Boltzmann weighting. It is important to note we have combined populations of turns and helices at the same sequence positions since any helical structure is typically localized to only one turn.

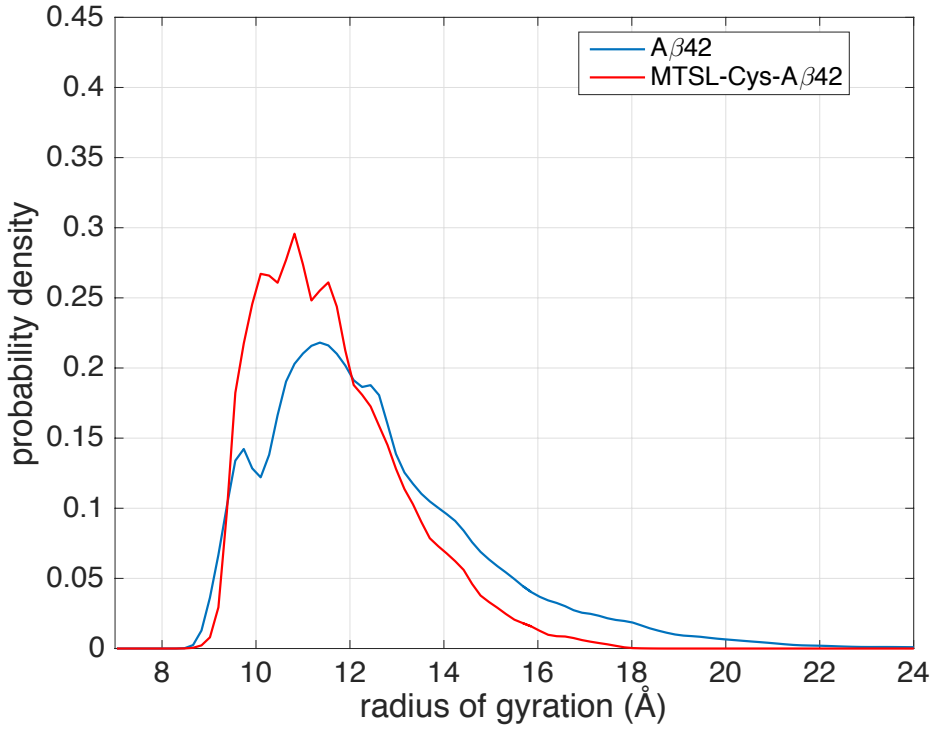


Figure 2.2: *Probability distribution of radius of gyration of the simulated Aβ42 (blue) and MTSL-Cys-Aβ42 (red) ensembles.*

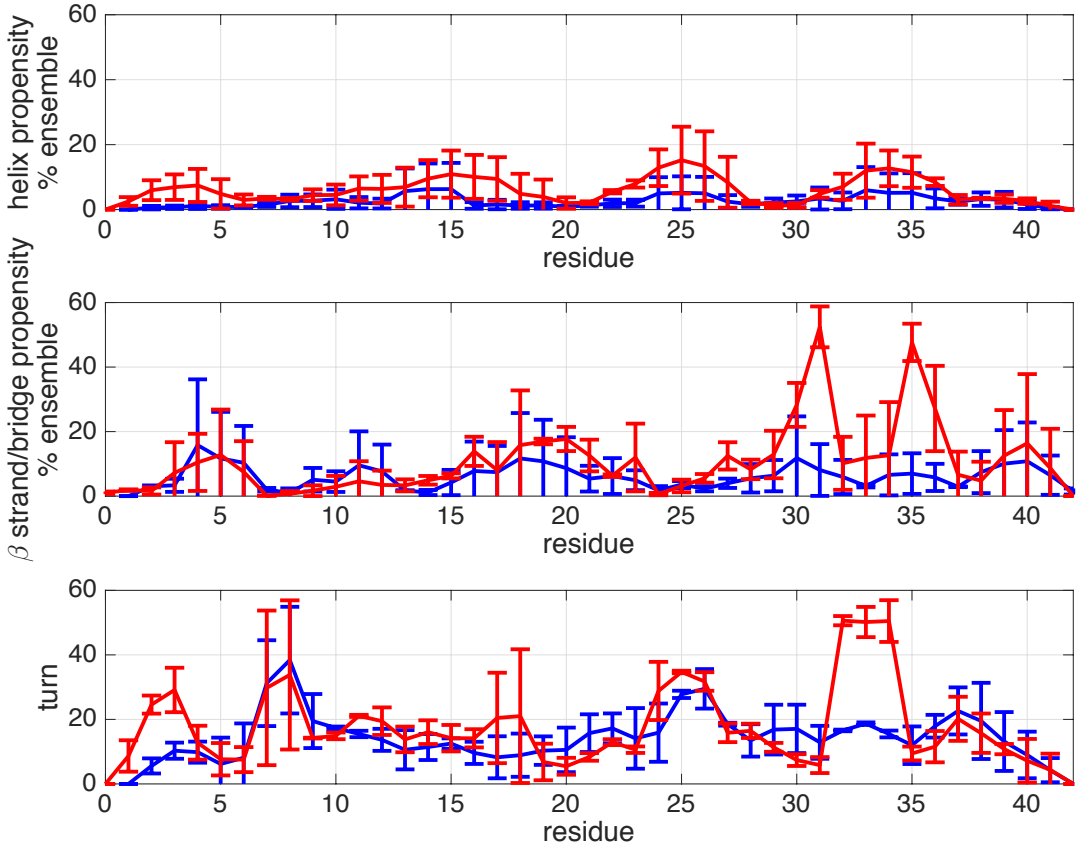


Figure 2.3: *Percentage of the simulated Aβ42 (blue) and MTSL-Cys-Aβ42 (red) ensemble involved in different types of secondary structure.* α-helix (top), β−bridges or β−strands (middle), and for turns (bottom). Standard deviations are based on the differences between the two ensembles generated from REMD and TCW sampling.

When Table 2.1 and Figure 2.3 are considered together, it is evident that many of the secondary structure sub-populations are quantitatively preserved between the Aβ42 and MTSL-Cys-Aβ42 ensemble. The reference peptide has a large number of highly populated turns and helices throughout the sequence (Figure 2.3) whose population percentages are almost completely explained by their stabilization due to internal hydrogen-bonds (Table 2.1). The most prominent β-hairpin structure in the original Aβ42 ensemble is formed by β-strand pairings between the central hydrophobic cluster and residues 29-36, sometimes stabilized by turns in the residue region Ala21-Gly25, such that the overall percentage of this β-hairpin averaged across all untagged Aβ42 ensembles is ~11%[41-43]. We find that the MTSL-Cys-Aβ42 structural ensemble stabilizes this β-hairpin at close to 18%, and this population of β-hairpin is largely preserved within statistical uncertainty compared to the reference ensemble. Even so, this β-hairpin is not present in the simulated Aβ40 ensemble, and so the fact that it is retained in the labeled Aβ42 peptide is encouraging for preserving one of the primary structural differences between the two amyloid-β peptides.

Table 2.1: *Comparison of important secondary structures present in the Aβ42 and the MTSL-Cys-Aβ42 ensemble.* Shaded boxes indicate the structural features whose populations differ in the untagged and tagged ensembles by a larger standard deviation than observed for REMD and TCW ensembles of the Aβ42 ensemble.

Sequence Region	Secondary structure %	Aβ42	MTSL-Cys-Aβ42
Localized secondary structure (helices and turns) for N-terminal residues 1-15	Turn 1-2/Turn 2-3	4.7 ± 1.8	23.3 ± 2.3
	Turn 3-4/Turn 5-6	13.0 ± 12.3	10.0 ± 5.0
	Turn 7-8/Turn 8-9	31.2 ± 24.3	32.6 ± 25.0
	Turn 9-10/Turn 10-11	16.8 ± 1.2	15.6 ± 0.2
	Turn 11-12/Turn 13-14	17.1 ± 3.0	13.4 ± 1.3
	Turn/Helix 14-15	12.6 ± 5.2	9.3 ± 3.4
Localized secondary structure (helices and turns) for central residues 16-30	Turn 21-22/Turn 22-23	12.2 ± 4.0	11.6 ± 2.5
	Turn 23-24/Turn 24-25	21.4 ± 5.1	22.8 ± 9.5
	Turn/Helix 25-26	22.6 ± 5.0	21.2 ± 4.4
	Turn/Helix 26-27	11.6 ± 6.2	12.8 ± 4.4
	Turn 29-30	12.0 ± 4.2	4.6 ± 2.4
Localized secondary structure (helices and turns) for C-terminal residues 31-42	Turn 30-31/ Turn 31-32	11.3 ± 4.7	3.8 ± 0.2
	Turn 32-33	9.1 ± 1.3	45.6 ± 1.6
	Turn 33-34	11.8 ± 6.1	3.2 ± 2.2
	Helix 34-35/Helix 35-36	5.1 ± 7.2	13.5 ± 1.2
	Turn 36-37	9.1 ± 4.1	5.2 ± 0.1
	Turn 37-38	14.6 ± 13.3	9.3 ± 6.7
Residues involves in β-strand pairings	β-strands 4-7 and 10-12	5.0 ± 3.7	3.1 ± 1.3
	β-strands 3-6 and 31-41	0.5 ± 0.7	6.6 ± 8.6
	β-strands 16-21 and 29-36	11.2 ± 3.1	17.8 ± 13.8
	β-strands 27-31 and 33-38	6.5 ± 2.1	38.0 ± 6.9
	β-strands 34-36 and 39-40	1.2 ± 0.5	5.4 ± 6.8

There are however some substantial shifts from the sub-population percentages seen for the unlabeled peptide in the C-terminal regions of the sequence due to the spin label. The reference Aβ42 ensemble exhibits two additional small sub-populations: (1) an anti-parallel β-strand pairing between residues 27-31 and 33-38 in ~6.5% of the ensemble, (2) and an N-terminal turn 1-2/2-3 present in ~5% of the ensemble. The percentages within these sub-populations shift dramatically in the presence of the Cys-MTSL addition to the N-terminus. While the perturbation in the turn population in the N-terminal might be expected, the most striking effect of the spin label is the greater enhancement in the population of the β-hairpin population (~38%), stabilized by β-turn 32-33. (Figure S2.1). This increase in secondary structure content is largely driven by the hydrophobic interactions of the spin label with many residues along the chain (Figure 2.4), which introduces new long-range interaction between the N- and C-terminus of the MTSL-Cys-Aβ42 peptide that contributes to a more collapsed ensemble, corroborated by the Rg data. Figure 2.5 shows a heavy atom contact map that averages over the Aβ42 and MTSL-Cys-Aβ42 structural ensembles; although the overall pattern of residue-residue contacts are similar in the two sequences, the MTSL-Cys residue is involved in

long-range heavy-atom contacts with residues 9-42 in >90% of the ensemble that explain these population shifts (Figure S2.2).

Considering the hydrophobic nature of the tag, a basic intermolecular interaction that is captured more than adequately by the force fields used here, the differences between the Aβ42 and MTSL-Cys-Aβ42 cannot be attributed wholesale to non-optimized force-field parameters but instead would be expected to give rise to the changes observed here. Thus, our observations of strong population shifts arising in the MTSL-Cys-Aβ42 ensemble can be tested experimentally. However we require a set of NMR observables that might reveal the structural differences. We have previously found that chemical shifts, J-couplings, and global residual dipolar couplings (RDC) are not sufficiently discriminating NMR observables to differentiate between IDP ensembles for Aβ40 vs. Aβ42, and Figures S2.3-S2.5 in the supplementary information confirm that these observables are not particularly helpful for discerning differences between Aβ42 and MTSL-Cys-Aβ42. However, we have previously found that local RDCs introduced by the Forman-Kay group[68], as well as homonuclear $^{1}H-^{1}H$ NOEs and heteronuclear $^{1}H-^{15}N$ NOEs, were found to better aid in the discrimination between different ensemble for Aβ40 vs. Aβ42[42].

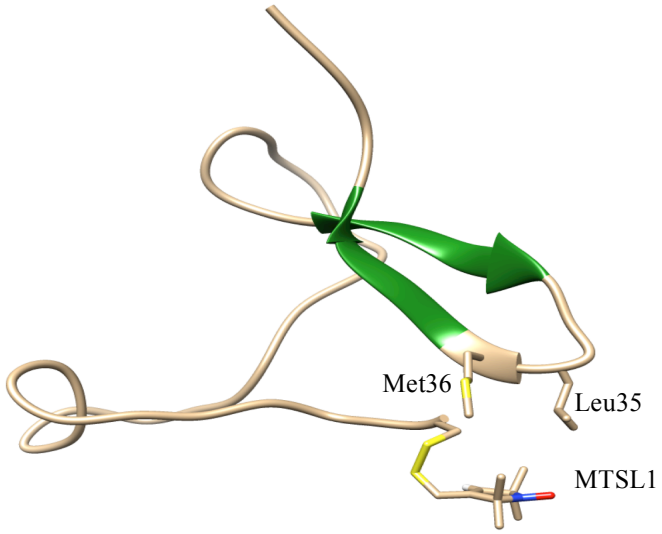


Figure 2.4: *Representative configuration of MTSL-Cys-Aβ42 ensemble* showing increased side-chain interactions between MTSL-Cys and Leu35 and Met36, which contributes to a more collapsed structural ensemble and β-strand pairings (green) between residues 29-32 and 36-38. This new β structure is present in negligible amounts in the untagged peptide. UCSF Chimera was used for the molecular graphics[69].

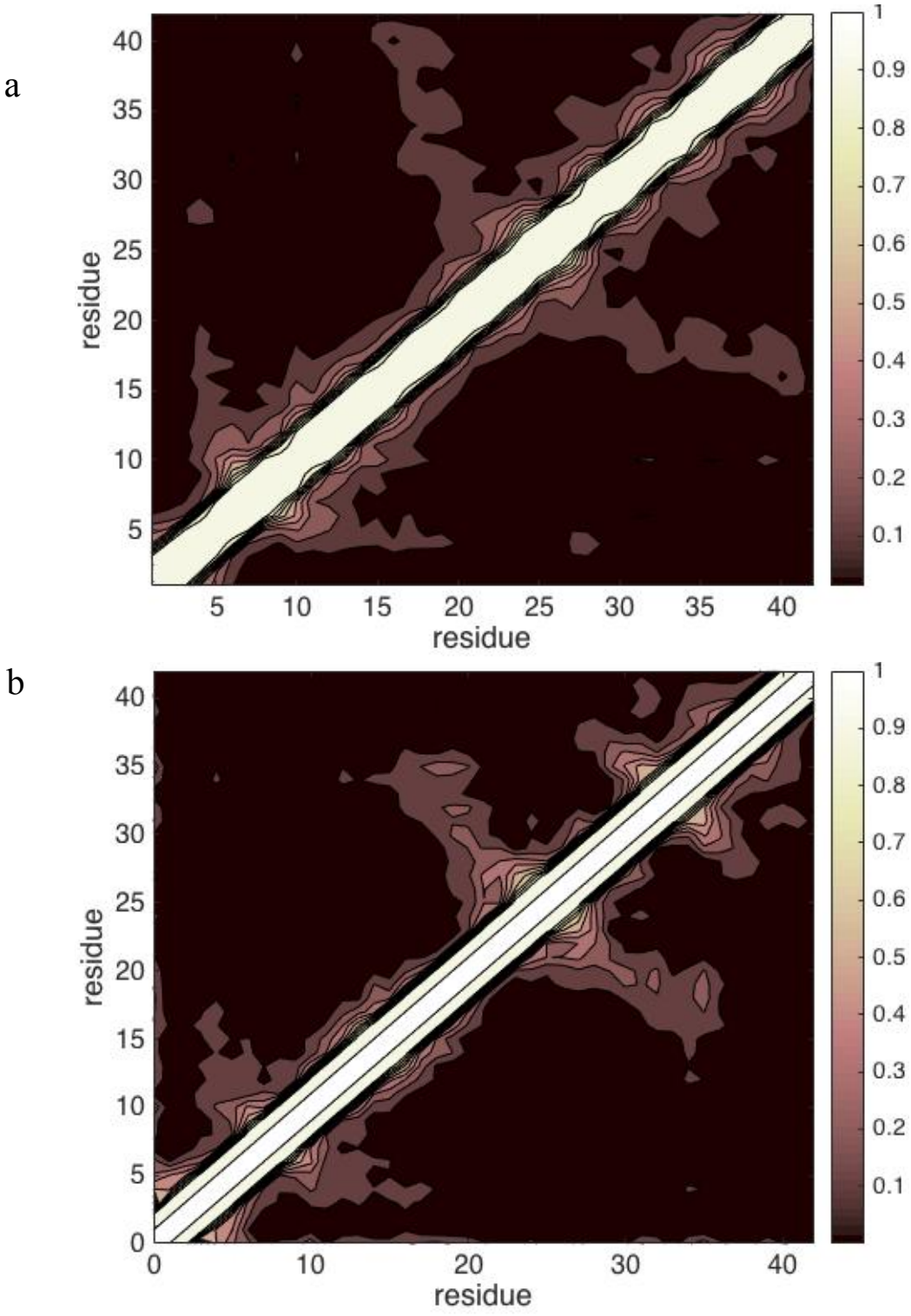


Figure 2.5: *Heavy-atom contact map of the simulated ensembles of (a) Aβ42 and (b) MTSL-Cys-Aβ42.* The contact maps portray the probability of interactions between each pair of residues in the simulated monomer ensembles. A contact is represented between two residues if any of their heavy atoms are less than 5 Å apart, and averaged over the ensemble to represent it with a probability ranging from 0 (black) to 1 (white).

Figure 2.6 shows the local $^{1}D_{NH}$ RDCs back-calculated from the untagged and tagged ensembles. It is evident that the local $^{1}D_{NH}$ values are larger on average in the presence of the spin label, and with clear signatures of the most populated β-strand in the MTSL-Cys-Aβ42 ensemble between 27-31 and 33-38 with a turn at position 30-32. Thus we would predict that the $^{1}D_{NH}$ RDCs would have enough signal above the highly averaged conformational background to detect these β-hairpins. Figure 2.7 compares the ^{15}N NOE enhancement factor for REMD simulations of the Aβ42 and MTSL-Cys-Aβ42 ensembles. While the trend is similar in both the ensembles, the MTSL-Cys-Aβ42 ensemble has higher NOE intensity values indicating that the backbone is more ordered, essentially over the entire peptide. In our simulated homonuclear NOE peaks for the Aβ42 and MTSL-Cys-Aβ42 ensembles, we find that the additional structure in the C-terminal region manifests as additional NOEs in the spin-labeled peptide (Figure S2.6). That said, it should be noted that the unpaired electron in the PRE tag present in the MTSL-Cys-Aβ42 ensemble contributes to the longitudinal relaxation of the nuclei of interest, and typically dominates the relaxation process such that the underlying NOEs are strongly diminished in the real experiment.[7] Since we do not consider the paramagnetic relaxation component for the NOE calculations for the MTSL-Cys-Aβ42 ensemble, further experimental data is required in regards to whether the ^{1}H and ^{15}N NOE features would have sufficient strength to be observed.

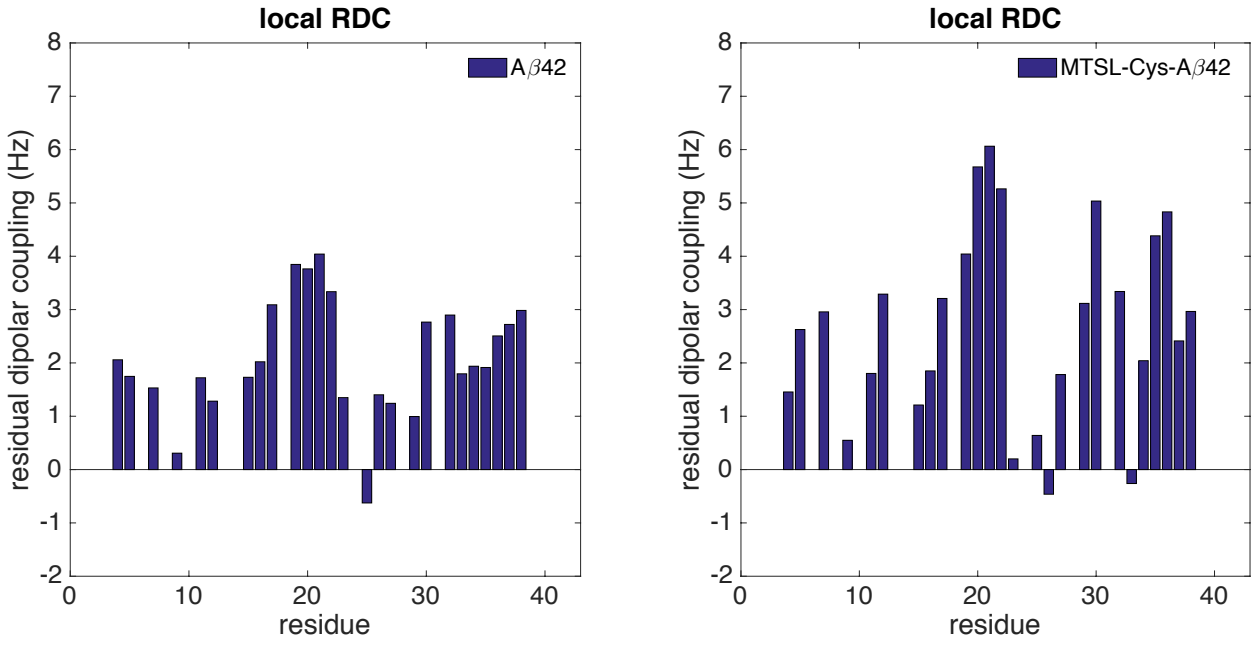


Figure 2.6: *Calculated local residual dipolar couplings for Aβ42 and MTSL-Cys-Aβ42.* Results generated based on local alignments using the ENSEMBLE software package[68].

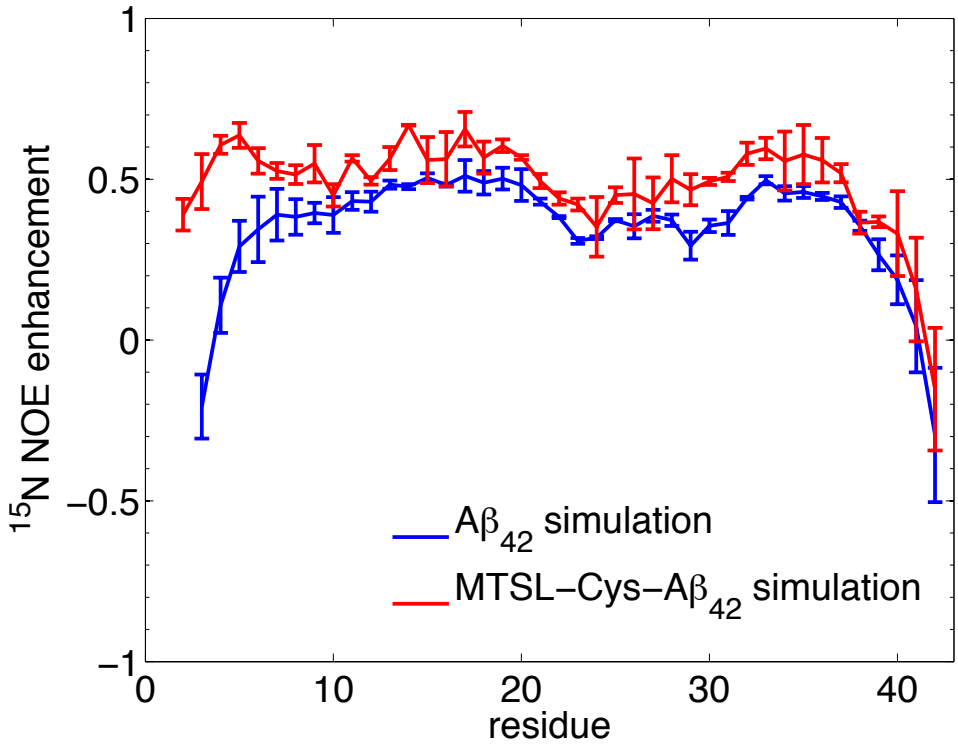


Figure 2.7: *Comparison of the simulated ^{1}H-^{15}N NOE enhancement for Aβ42 (blue) and MTSL-Cys-Aβ42 (red) from REMD simulations.* Higher values in the tagged peptide support a more ordered backbone.

2.4 DISCUSSION

Our results that structural perturbations are inevitable due to the addition of a spin label are supported by a number of previous PRE studies on folded as well as intrinsically disordered proteins. Although for well-folded proteins most sites are able to accommodate the addition of a invasive tag like MTSL without any significant structural changes[21, 70, 71], both computational[23] and experimental studies[6] on the folded state of cytochrome c did show a slight decrease in protein helicity after probe attachment. In the study by Kjaergaard *et al*. on the disordered ACTR peptide, they observed 3-7% increase in the populations of existing helical segments after introduction of the MTSL spin label, although they considered the perturbations not strong enough to influence their conclusions about long-range interactions between the targeted set of helical domains[19]. MTSL and TEMPOL, a different PRE tag, have been shown to have high affinity towards hydrophobic surfaces[72, 73], and the increased surface interaction in turn induces more aggregation in the labeled Aβ peptide[16]. Finally, another way of conducting PRE experiments is to introduce small sequence motifs which bind to paramagnetic metal ions like Cu^{2+}, Ni^{2+} or lanthanide ions, but these methods are not popular for IDP experiments because they require larger sequence modification if the motif is not present in the original protein and would certainly perturb the underlying structural ensemble significantly[74].

However, the question is what can be learned from PRE experiments when applied to IDPs in solution, in spite of the inevitable perturbations due to the tag. In the light of current simulation results, PRE experiments aimed to look into the kinetics of fibril or oligomer growth mechanisms of Aβ42 (or Aβ40 peptide[15]), or interactions with membranes[75], should be analyzed

with caution. Fibrils usually elongate by a faster 'dock' and slower 'lock' mechanism[76], and the tagged peptide might take longer to properly lock because of the introduction of free energy barriers associated with the new/enhanced β-structure assemblies. The enhanced structure formation appears to arise from excessive hydrophobicity of the MTSL label could be reduced by using a modified MTSL tags designed to be more hydrophilic[77, 78]. Such labels are a welcome addition given the importance of experiments such as PRE that can probe long-range order in IDP structures in solution, although a similar study to this one would be needed to help quantify the structural ensemble perturbations introduced by the more hydrophilic label.

2.5 CONCLUSION

The most widely used method for gaining detailed structural insights into IDPs is to combine experimental techniques such as NMR, infrared spectroscopy, and small angle X-ray scattering with computational techniques like molecular dynamics (MD), Monte Carlo simulations, or knowledge-based methods[43, 79, 80]. The experimental observables are typically used to either validate IDP structural ensembles that are generated from Boltzmann weighted molecular dynamics simulations[32, 36, 41-43], to select for conformations from computationally generated IDP (typically random coil) ensembles[68, 79, 80], or to provide structural restraints during guided MD simulations[9, 11, 79, 81]. Here we have instead utilized molecular simulation to predict the perturbations caused by addition of the commonly used MTSL tag in PRE experiments when applied to the disordered Aβ42 peptide, and to provide suggestions as to what experiments might be used to confirm those predictions.

We have employed several independent calculations of the structural ensemble of the native Aβ42 ensemble to provide uncertainty estimates due to sampling protocol, in order to distinguish real structural perturbations measured from the average of two independently generated ensembles of the labeled MTSL-Cys-Aβ42 peptide. To sum up the effects of the spin label, a majority of the main structural features of the peptide, including the dominant β-hairpin formed by residues 16-21 and 29-36[41, 43], are not affected by the perturbations introduced by the tag. Nevertheless, the tag causes population enhancements in the C-terminus involving β-hairpin formed by residues 27-31 and 33-38. The MTSL tag, being hydrophobic in nature, has side-chain interactions with other residues in most of the ensemble, resulting in a more collapsed ensemble, which in turn promotes the formation of β-strands in the C-terminus. If ^{1}H and ^{15}N NOE experiments can be performed without the paramagnetic spin label dominating the relaxation mechanisms, experiments should point towards a more ordered backbone and new NOE peaks arising from the enhanced/new β strand contacts. Our results could also be supported by RDC experiments in which local alignments are considered[68], which will show increased $^{1}D_{NH}$ values resulting from higher β- structure propensities, especially in the C-terminus.

In conclusion, PRE experiments, fluorescence resonance energy transfer (FRET), and double electron-electron resonance (DEER) experiments can probe long-range order in polypeptide structures in solution, although the requirement of a reporter tag will likely perturb the underlying structural ensemble to some degree[9,11,74]. Furthermore, the nature of the structural perturbation will vary with peptide size, sequence, the tag type and its position, and whether it is being used to probe a folded or disordered structural ensemble. Since quantifying the exact extent of the perturbations caused by the spin label can be difficult experimentally, we believe that molecular dynamics simulation is a powerful and complementary approach to resolving the potential structural perturbations, and to provide suggested experiments which can confirm the theoretical predictions.

2.6 REFERENCES

1. Uversky, V.; Gillespie, J.; Fink, A., Why Are "Natively Unfolded" Proteins Unstructured under Physiologic Conditions? *Proteins: Struct., Func., Bioinform.* **2000,** *41* (3), 415-427.
2. Tompa, P., Intrinsically Unstructured Proteins. *Trends in Biochem. Sci.* **2002,** *27* (10), 527-533.
3. Dyson, H. J.; Wright, P. E., Intrinsically Unstructured Proteins and Their Functions. *Nature Reviews: Mol. Cell Biol.* **2005,** *6* (3), 197-208.
4. Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L., Function and Structure of Inherently Disordered Proteins. *Curr. Opin. Struct. Bio.* **2008,** *18* (6), 756-64.
5. Higo, J.; Nishimura, Y.; Nakamura, H., A Free-Energy Landscape for Coupled Folding and Binding of an Intrinsically Disordered Protein in Explicit Solvent from Detailed All-Atom Computations. *J. Am. Chem. Soc.* **2011,** *133* (27), 10448-10458.
6. Qu, K.; Vaughn, J. L.; Sienkiewicz, A.; Scholes, C. P.; Fetrow, J. S., Kinetics and Motional Dynamics of Spin-Labeled Yeast Iso-1-Cytochrome C: 1. Stopped-Flow Electron Paramagnetic Resonance as a Probe for Protein Folding/Unfolding of the C-Terminal Helix Spin-Labeled at Cysteine 102. *Biochemistry* **1997,** *36* (10), 2884-97.
7. Bertini, I.; Luchinat, C.; Parigi, P., *Solution NMR of Paramagnetic Molecules: Applications to Metallobiomolecules and Models*. Elsevier: Amsterdam, 2001; Vol. 2.
8. Bertini, I.; Donaire, A.; Jimenez, B.; Luchinat, C.; Parigi, G.; Piccioli, M.; Poggi, L., Paramagnetism-Based Versus Classical Constraints: An Analysis of the Solution Structure of Ca Ln Calbindin D-9k. *J. Biomol. NMR* **2001,** *21* (2), 85-98.
9. Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M., Mapping Long-Range Interactions in α-Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005,** *127* (2), 476-477.
10. Vise, P.; Baral, B.; Stancik, A.; Lowry, D. F.; Daughdrill, G. W., Identifying Long-Range Structure in the Intrinsically Unstructured Transactivation Domain of P53. *Proteins: Struct., Func., Bioinform.* **2007,** *67* (3), 526-530.
11. Allison, J. R.; Várnai, P.; Dobson, C. M.; Vendruscolo, M., Determination of the Free Energy Landscape of α-Synuclein Using Spin Label Nuclear Magnetic Resonance Measurements. *J. Am. Chem. Soc.* **2009,** *131* (51), 18314-18326.
12. Ganguly, D.; Chen, J., Structural Interpretation of Paramagnetic Relaxation Enhancement-Derived Distances for Disordered Protein States. *J. Mol. Biol.* **2009,** *390* (3), 467-477.
13. Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M., NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **2010,** *132* (24), 8407-8418.
14. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A., Current Status of the Amoeba Polarizable Force Field. *J. Phys. Chem. B* **2010,** *114* (8), 2549.
15. Fawzi, N.; Ying, J.; Torchia, D., A.; Clore, G. M., Kinetics of Amyloid-β Monomer to Oligomer Exchange by NMR Relaxation. *J. Am. Chem. Soc.* **2010,** *132*, 9948-9951.
16. Fawzi, N. L.; Ying, J.; Ghirlando, R.; Torchia, D. A.; Clore, G. M., Atomic-Resolution Dynamics on the Surface of Amyloid-β Protofibrils Probed by Solution NMR. *Nature* **2011,** *480*, 268-272.
17. Gu, L.; Liu, C.; Guo, Z., Structural Insights into A β 42 Oligomers Using Site-Directed Spin Labeling. *J Biol. Chem.* **2013,** *288* (26), 18673-83.

18. Yamaguchi, T.; Matsuzaki, K.; Hoshino, M., Interaction between Soluble A β -(1-40) Monomer and A β -(1-42) Fibrils Probed by Paramagnetic Relaxation Enhancement. *FEBS Lett.* **2013,** *587* (6), 620-4.
19. Iešmantavičius, V.; Jensen, M. R.; Ozenne, V. R.; Blackledge, M.; Poulsen, F. M.; Kjaergaard, M., Modulation of the Intrinsic Helix Propensity of an Intrinsically Disordered Protein Reveals Long-Range Helix–Helix Interactions. *J. Am. Chem. Soc.* **2013,** *135* (27), 10155-10163.
20. Hubbell, W. L.; Gross, A.; Langen, R.; Lietzow, M. A., Recent Advances in Site-Directed Spin Labeling of Proteins. *Curr. Opin Struct. Biol.* **1998,** *8* (5), 649-56.
21. Koteiche, H. A.; McHaourab, H. S., Folding Pattern of the Alpha-Crystallin Domain in alphaA-Crystallin Determined by Site-Directed Spin Labeling. *J. Mol. Biol.* **1999,** *294* (2), 561-77.
22. Parigi, G.; Bertini, I.; Luchinat, C.; Piccioli, M.; Piva, A., Paramagnetic Constraints in Structure-Determination Programs. *J. Inorg. Biochem.* **2001,** *86* (1), 507-507.
23. Murzyn, K.; Rog, T.; Blicharski, W.; Dutka, M.; Pyka, J.; Szytula, S.; Froncisz, W., Influence of the Disulfide Bond Configuration on the Dynamics of the Spin Label Attached to Cytochrome C. *Proteins: Struct., Func., Bioinform.* **2006,** *62* (4), 1088-100.
24. Barrow, C. J.; Yasuda, A.; Kenny, P. T.; M.G., Z., Solution Conformations and Aggregational Properties of Synthetic Amyloid β-Peptides of Alzheimer's Disease. Analysis of Circular Dichroism Spectra. *J. Mol. Biol.* **1992,** *225*, 1075-1093.
25. Goedert, M.; Spillantini, M. G., A Century of Alzheimer's Disease. *Science* **2006,** *314* (5800), 777-781.
26. Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R., A Structural Model for Alzheimer's β-Amyloid Fibrils Based on Experimental Constraints from Solid State NMR. *Proc. Natl. Acad. Sci.* **2002,** *99* (26), 16742-7.
27. Yan, Y.; Wang, C., Aβ 42 Is More Rigid Than Aβ 40 at the C Terminus: Implications for Aβ Aggregation and Toxicity. *J. Mol. Biol.* **2006,** *364* (5), 853-62.
28. Yan, Y.; Liu, J.; McCallum, S.; Yang, D.; Wang, C., Methyl Dynamics of the Amyloid-β Peptides Aβ40 and Aβ42. *Biochem. Biophys. Res. Comm.* **2007,** *362* (2), 410-414.
29. Roche, J.; Shen, Y.; Lee, J. H.; Ying, J.; Bax, A., Monomeric Aβ1–40 and Aβ1–42 Peptides in Solution Adopt Very Similar Ramachandran Map Distributions That Closely Resemble Random Coil. *Biochemistry* **2016,** *55* (5), 762.
30. Massi, F.; Peng, J. W.; Lee, J. P.; Straub, J. E., Simulation Study of the Structure and Dynamics of the Alzheimer's Amyloid Peptide Congener in Solution. *Biophys. J.* **2001,** *80* (1), 31-44.
31. Ma, B.; Nussinov, R., Stabilities and Conformations of Alzheimer's β-Amyloid Peptide Oligomers (Aβ16–22, Aβ16–35, and Aβ10–35): Sequence Effects. *Proc. Natl. Acad. Sci.* **2002,** *99* (22), 14126-14131.
32. Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E., The Alzheimer's Peptides A β 40 and 42 Adopt Distinct Conformations in Water: A Combined Md / NMR Study. *J. Mol. Biol.* **2007,** *368* (5), 1448-57.
33. Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T., Structure and Dynamics of the Aβ (21-30) Peptide from the Interplay of NMR Experiments and Molecular Simulations. *J. Am. Chem. Soc.* **2008,** *130* (19), 6145-58.

34. Lam, A.; Teplow, D.; Stanley, H.; Urbanc, B., Effects of the Arctic (E22-->G) Mutation on Amyloid β-Protein Folding: Discrete Molecular Dynamics Study. *J. Am. Chem. Soc.* **2008,** *130* (51), 17413-17422.
35. Melquiond, A.; Dong, X.; Mousseau, N.; Derreumaux, P., Role of the Region 23-28 in Aβ Fibril Formation: Insights from Simulations of the Monomers and Dimers of Alzheimer's Peptides Aβ 40 and Aβ 42. *Curr. Alz. Res.* **2008,** *5* (3), 244-250.
36. Sgourakis, N. G.; Merced-Serrano, M.; Boutsidis, C.; Drineas, P.; Du, Z.; Wang, C.; Garcia, A. E., Atomic-Level Characterization of the Ensemble of the Aβ (1-42) Monomer in Water Using Unbiased Molecular Dynamics Simulations and Spectral Algorithms. *J. Mol. Biol.* **2011,** *405* (2), 570-83.
37. Wu, C.; Shea, J.-E., The Structure of Intrinsically Disordered Peptides Implicated in Amyloid Diseases: Insights from Fully Atomistic Simulations. In *Computational Modeling of Biological Systems: From Molecules to Pathways*, Dokholyan, N., Ed. Springer: Berlin, 2012; pp 215-227.
38. Lin, Y. S.; Bowman, G. R.; Beauchamp, K. A.; Pande, V. S., Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. *Biophys. J.* **2012,** *102* (2), 315-24.
39. Lin, Y. S.; Pande, V. S., Effects of Familial Mutations on the Monomer Structure of A β 42. *Biophys. J.* **2012,** *103* (12), L47-9.
40. Rosenman, D. J.; Connors, C. R.; Chen, W.; Wang, C.; Garcia, A. E., A β Monomers Transiently Sample Oligomer and Fibril-Like Configurations: Ensemble Characterization Using a Combined Md/NMR Approach. *J. Mol. Biol.* **2013,** *425* (18), 3338-59.
41. Ball, K. A.; Phillips, A. H.; Nerenberg, P. S.; Fawzi, N. L.; Wemmer, D. E.; Head-Gordon, T., Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-β Peptides. *Biochemistry* **2011,** *50* (35), 7612-28.
42. Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T., Differences in β-Strand Populations of Monomeric Amyloid-β 40 and Amyloid-β 42. *Biophys. J.* **2013,** *104* (12), 2714-2724.
43. Ball, K. A.; Wemmer, D. E.; Head-Gordon, T., Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-β Peptides. *J. Phys. Chem. B* **2014,** *118*, 6405-6416.
44. Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B., Amyloid β-Protein (Aβ) Assembly: Aβ 40 and Aβ 42 Oligomerize through Distinct Pathways. *Proc. Natl. Acad. Sci.* **2003,** *100* (1), 330-5.
45. Kim, W.; Hecht, M. H., Sequence Determinants of Enhanced Amyloidogenicity of Alzheimer A β 42 Peptide Relative to A β 40. *J. Biol. Chem.* **2005,** *280* (41), 35069-35076.
46. Weinreb, P. H.; Jarrett, J. T.; Lansbury, P. T., Peptide Models of a Hydrophobic Cluster at the C-Terminus of the β-Amyloid Protein. *J. Am. Chem. Soc.* **1994,** *116* (23), 10835-10836.
47. Morimoto, A.; Irie, K.; Murakami, K.; Masuda, Y.; Ohigashi, H.; Nagao, M.; Fukuda, H.; Shimizu, T.; Shirasawa, T., Analysis of the Secondary Structure of β-Amyloid (A β 42) Fibrils by Systematic Proline Replacement. *J. Biol. Chem.* **2004,** *279* (50), 52781-52788.
48. Irie, K.; Murakami, K.; Masuda, Y.; Morimoto, A.; Ohigashi, H.; Ohashi, R.; Takegoshi, K.; Nagao, M.; Shimizu, T.; Shirasawa, T., Structure of β-Amyloid Fibrils and Its Relevance to Their Neurotoxicity: Implications for the Pathogenesis of Alzheimer's Disease. *J. Biosci. Bioeng.* **2005,** *99* (5), 437-447.

49. Lim, K. H.; Collver, H. H.; Le, Y. T.; Nagchowdhuri, P.; Kenney, J. M., Characterizations of Distinct Amyloidogenic Conformations of the A β (1–40) and (1–42) Peptides. *Biochem. Biophys. Res. Comm.* **2007,** *353* (2), 443-449.
50. Maji, S. K.; Loo, R. R. O.; Inayathullah, M.; Spring, S. M.; Vollers, S. S.; Condron, M. M.; Bitan, G.; Loo, J. A.; Teplow, D. B., Amino Acid Position-Specific Contributions to Amyloid β-Protein Oligomerization. *J. Biol. Chem.* **2009,** *284* (35), 23580-23591.
51. Sugita, Y.; Okamoto, Y., Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999,** *314* (1), 141-151.
52. Lincoff, J.; Sasmal, S.; Head-Gordon, T., Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom. *J. Chem. Phys.* **2016,** *145* (17), 174107.
53. Brown, S.; Head‐Gordon, T., Cool Walking: A New Markov Chain Monte Carlo Sampling Method. *J. Comput. Chem.* **2003,** *24* (1), 68-76.
54. Case, D. A. B., V.;Berryman, J.T.; Betz, R.M.;Cai, Q.; Cerutti, D.S.; Cheatham, III, T.E.; Darden, T.A.; Duke, R.E.; Gohlke, H. et al. *Amber 14*, University of California, San Francisco, 2014.
55. 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., Func., Bioinform.* **2006,** *65* (3), 712-25.
56. Wickstrom, L.; Okur, A.; Simmerling, C., Evaluating the Performance of the Ff99sb Force Field Based on NMR Scalar and Coupling Data. *Biophys. J.* **2009,** *97* (3), 853-856.
57. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an Improved Four-Site Water Model for Biomolecular Simulations: Tip4p-Ew. *J. Chem. Phys.* **2004,** *120* (20), 9665-78.
58. Xue, Y.; Skrynnikov, N. R., Motion of a Disordered Polypeptide Chain as Studied by Paramagnetic Relaxation Enhancements, ^{15}N Relaxation, and Molecular Dynamics Simulations: How Fast Is Segmental Diffusion in Denatured Ubiquitin? *J. Am. Chem. Soc.* **2011,** *133* (37), 14614-28.
59. Andersen, H. C., Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980,** *72* (4), 2384-2393.
60. Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S., Rapid and Accurate Calculation of Protein 1h, 13c and 15n Chemical Shifts. *J. Biomol. NMR* **2003,** *26* (3), 215-40.
61. Karplus, M., Contact Electron‐Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.* **1959,** *30* (1), 11-15.
62. Marsh, J. A.; Baker, J. M. R.; Tollinger, M.; Forman-Kay, J. D., Calculation of Residual Dipolar Couplings from Disordered State Ensembles Using Local Alignment. *J. Am. Chem. Soc.* **2008,** *130* (25), 7804-7805.
63. Zweckstetter, M.; Bax, A., Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *J. Am. Chem. Soc.* **2000,** *122* (15), 3791-3792.
64. Hansmann, U. H., Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997,** *281* (1), 140-150.
65. Marinari, E.; Parisi, G., Simulated Tempering: A New Monte Carlo Scheme. *Europhys. Lett.* **1992,** *19* (6), 451.
66. Hukushima, K.; Nemoto, K., Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Japan* **1996,** *65* (6), 1604-1608.

67. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen‐Bonded and Geometrical Features. *Biopolymers* **1983,** *22* (12), 2577-2637.
68. Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W.-Y.; Forman-Kay, J. D., Characterization of Disordered Proteins with Ensemble. *Bioinform.* **2013,** *29* (3), 398-399.
69. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., Ucsf Chimera—a Visualization System for Exploratory Research and Analysis. *Journal of computational chemistry* **2004,** *25* (13), 1605-1612.
70. Matthews, B. W., Studies on Protein Stability with T4 Lysozyme. *Adv. Protein Chem.* **1994,** *46*, 249-278.
71. Mchaourab, H. S.; Lietzow, M. A.; Hideg, K.; Hubbell, W. L., Motion of Spin-Labeled Side Chains in T4 Lysozyme. Correlation with Protein Structure and Dynamics. *Biochemistry* **1996,** *35* (24), 7692-7704.
72. Deschamps, M. L.; Pilka, E. S.; Potts, J. R.; Campbell, I. D.; Boyd, J., Probing Protein-Peptide Binding Surfaces Using Charged Stable Free Radicals and Transverse Paramagnetic Relaxation Enhancement (Pre). *J. Biomol. NMR* **2005,** *31* (2), 155-160.
73. Showalter, S. A.; Bruschweiler-Li, L.; Johnson, E.; Zhang, F.; Brüschweiler, R., Quantitative Lid Dynamics of Mdm2 Reveals Differential Ligand Binding Modes of the P53-Binding Cleft. *J. Am. Chem. Soc.* **2008,** *130* (20), 6472-6478.
74. Clore, G. M.; Iwahara, J., Theory, Practice and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem. Rev.* **2009,** *109* (9), 4108-4139.
75. Yagi-Utsumi, M.; Kameda, T.; Yamaguchi, Y.; Kato, K., NMR Characterization of the Interactions between Lyso-Gm1 Aqueous Micelles and Amyloid Β. *FEBS Lett.* **2010,** *584* (4), 831-6.
76. Sasmal, S.; Schwierz, N.; Head-Gordon, T., Mechanism of Nucleation and Growth of A β 40 Fibrils from All-Atom and Coarse-Grained Simulations. *J. Phys. Chem. B* **2016,** *120* (47), 12088-12097.
77. Fawzi, N. L.; Fleissner, M. R.; Anthis, N. J.; Kálai, T.; Hideg, K.; Hubbell, W. L.; Clore, G. M., A Rigid Disulfide-Linked Nitroxide Side Chain Simplifies the Quantitative Analysis of Pre Data. *J. Biomol. NMR* **2011,** *51* (1-2), 105-114.
78. Card, P. B.; Erbel, P. J.; Gardner, K. H., Structural Basis of Arnt Pas-B Dimerization: Use of a Common Β-Sheet Interface for Hetero-and Homodimerization. *J. Mol. Biol.* **2005,** *353* (3), 664-677.
79. Marsh, J. A.; Forman-Kay, J. D., Ensemble Modeling of Protein Disordered States: Experimental Restraint Contributions and Validation. *Proteins: Struct., Func., Bioinform.* **2012,** *80* (2), 556-572.
80. Schneider, R.; Huang, J.-R.; Yao, M.; Communie, G.; Ozenne, V.; Mollica, L.; Salmon, L.; Jensen, M. R.; Blackledge, M., Towards a Robust Description of Intrinsic Protein Disorder Using Nuclear Magnetic Resonance Spectroscopy. *Mol. BioSys.* **2012,** *8* (1), 56-68.
81. López-Méndez, B.; Güntert, P., Automated Protein Structure Determination from NMR Spectra. *J. Am. Chem. Soc.* **2006,** *128*, 13112-13222.

2.6 APPENDIX

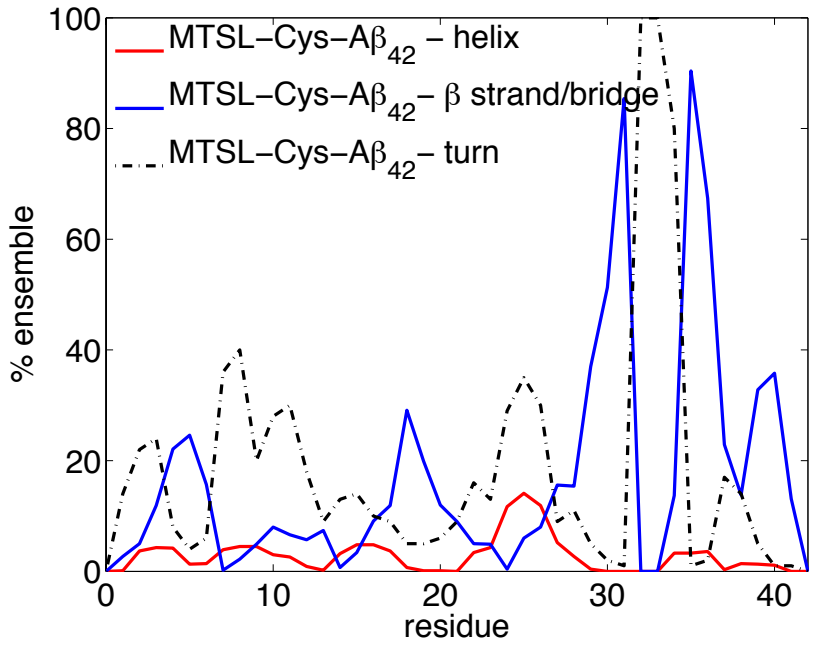


Figure S2.1: *Percentages of MTSL-Cys-Aβ42 sub-ensembles containing turn 32-33 for REMD simulations.*

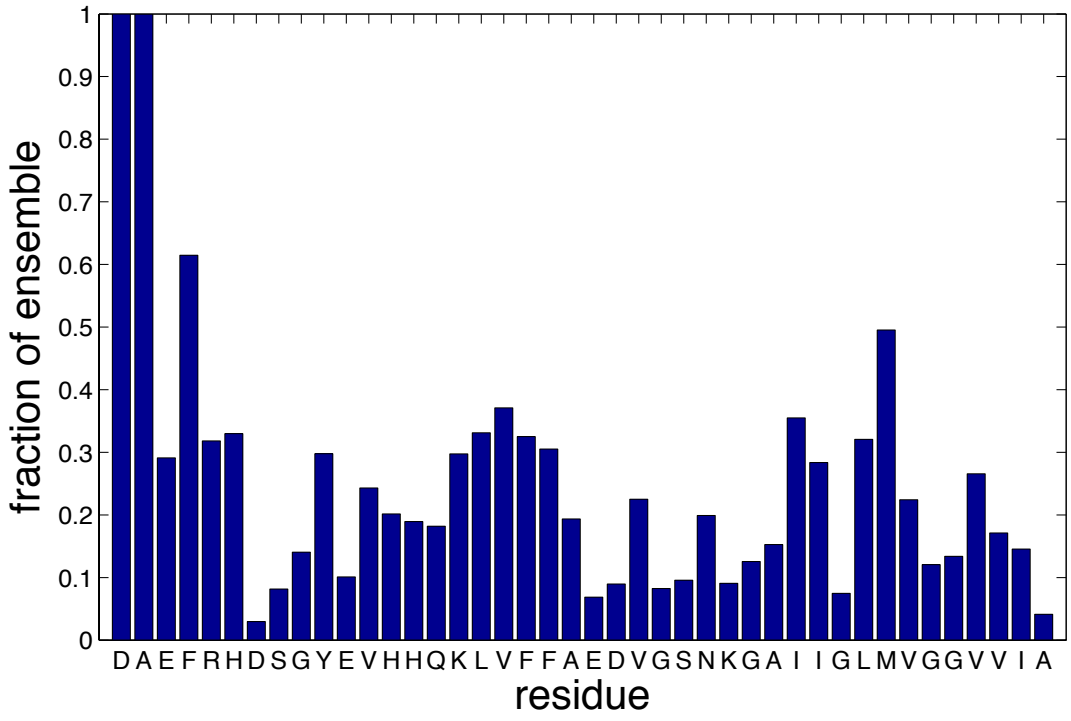


Figure S2.2: *Plot of the probability of a heavy atom contact between the MTSL-Cys residue and other residues in the Aβ42 sequence for REMD simulations.*

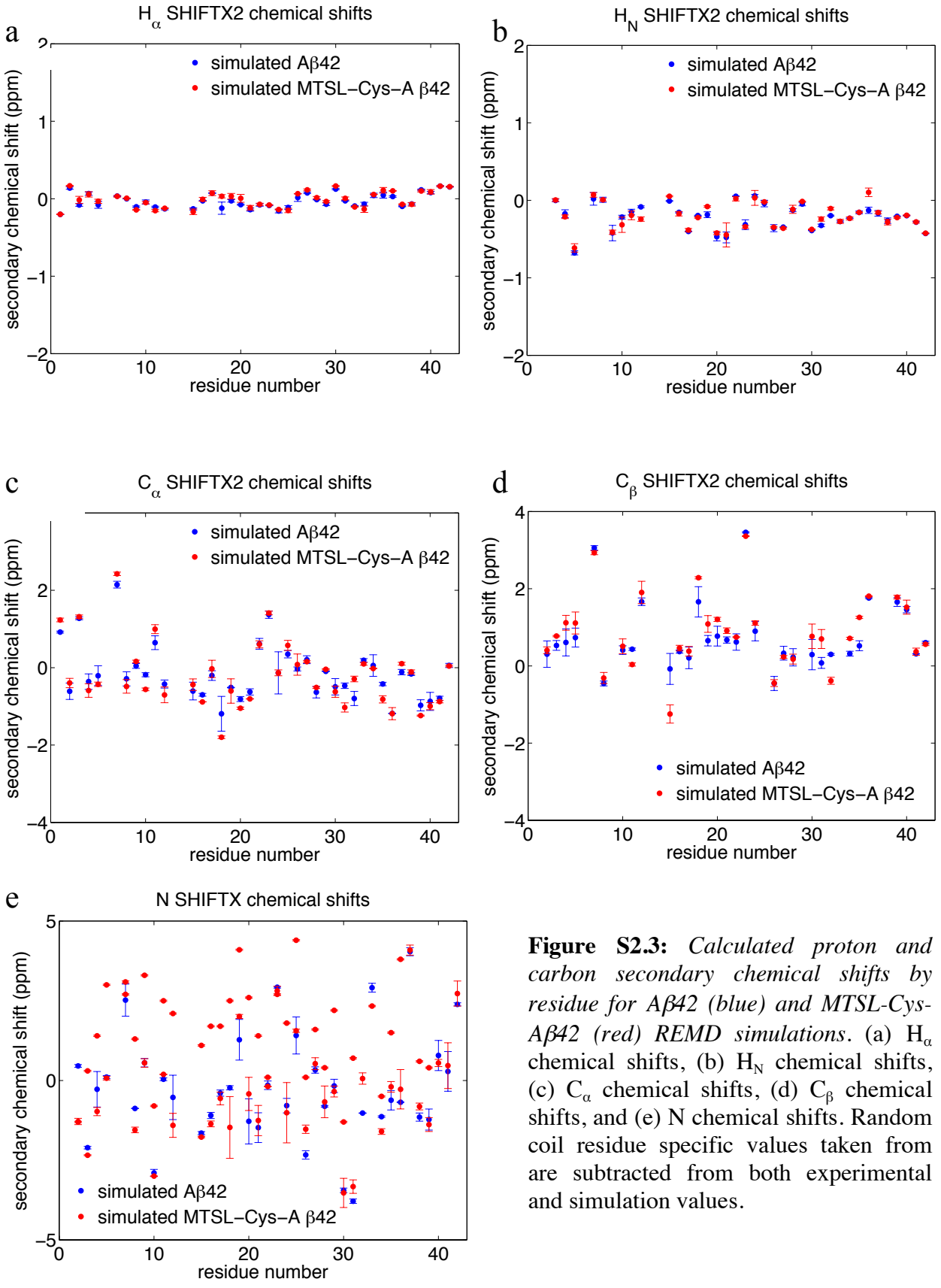


Figure S2.3: *Calculated proton and carbon secondary chemical shifts by residue for Aβ42 (blue) and MTSL-Cys-Aβ42 (red) REMD simulations*. (a) H_α chemical shifts, (b) H_N chemical shifts, (c) C_α chemical shifts, (d) C_β chemical shifts, and (e) N chemical shifts. Random coil residue specific values taken from are subtracted from both experimental and simulation values.

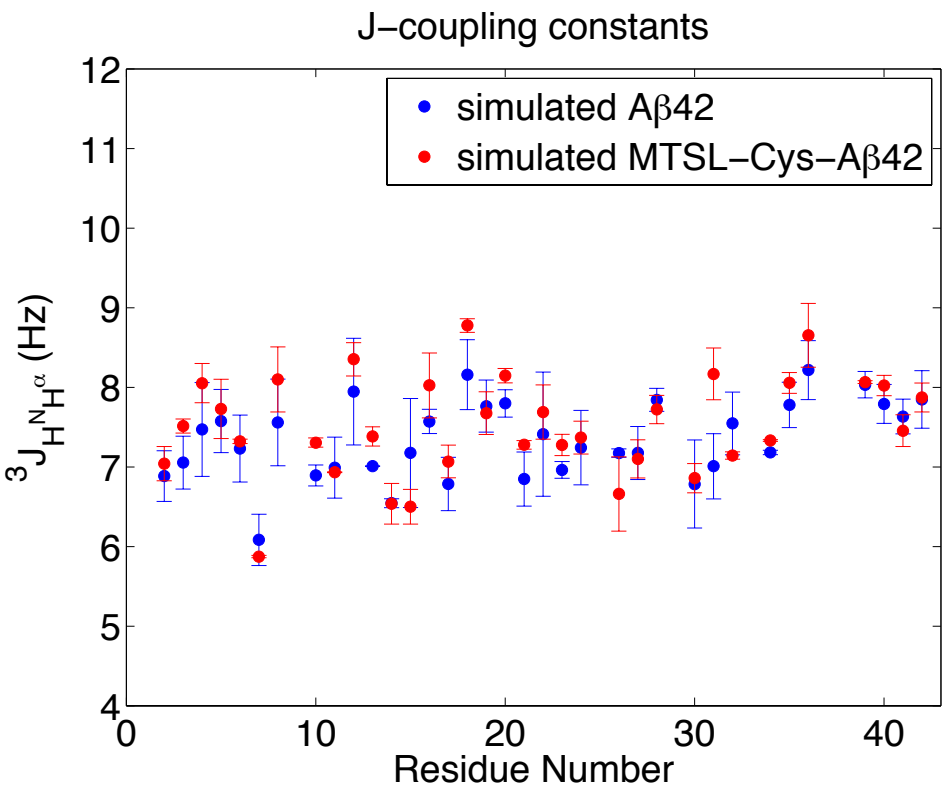
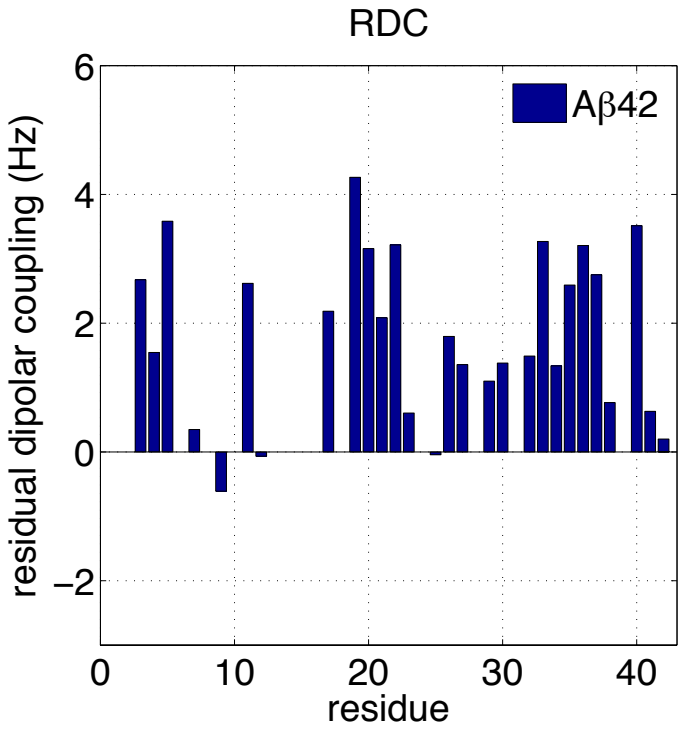


Figure S2.4: *J-coupling constants for backbone amides for Aβ42 (blue) and MTSL-Cys-Aβ42 (red) for REMD simulations.* Simulation uncertainty bars represent rms difference between two independent simulations and the average.

a



b

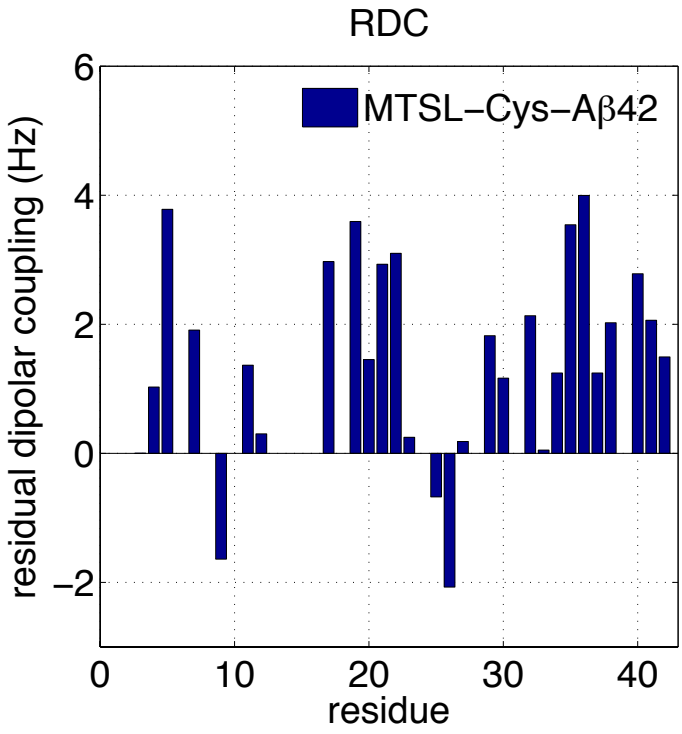


Figure S2.5: *Calculated RDCs for (a) Aβ42 and (b) MTSL-Cys-Aβ42 structural ensembles generated using REMD simulations.* RDCs based on global alignments calculated from PALES.

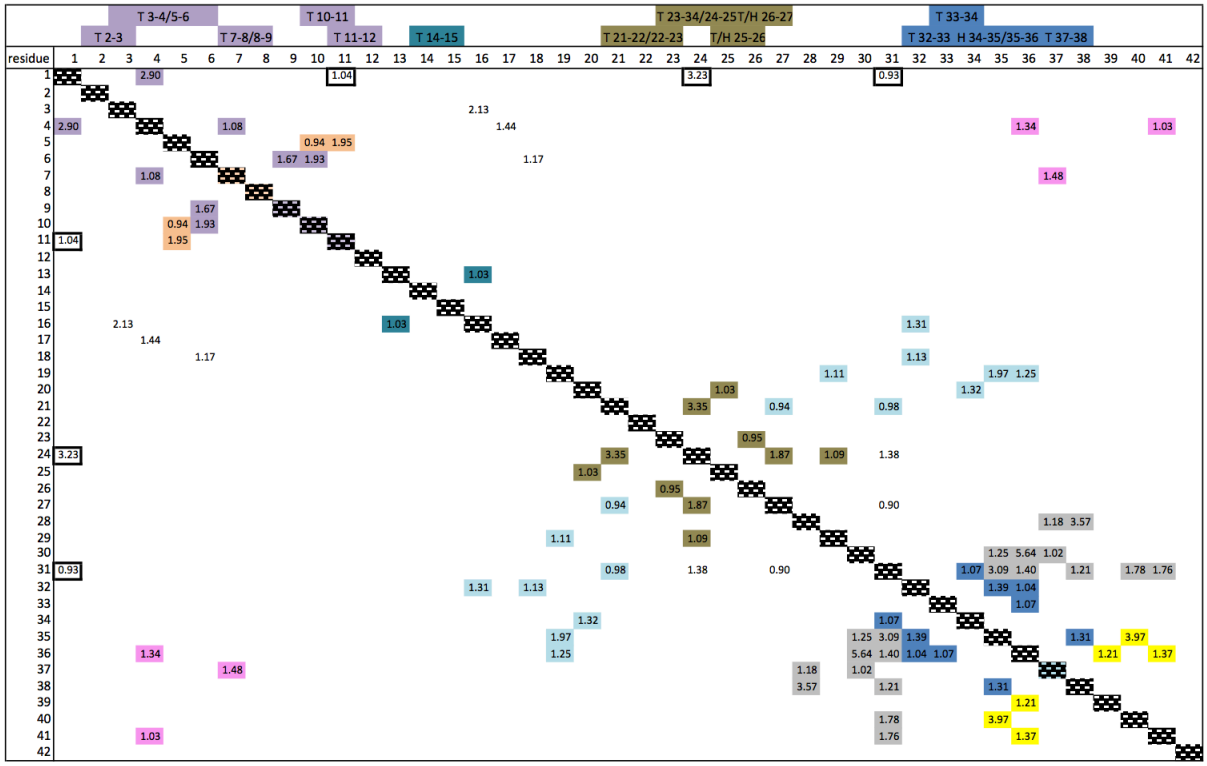


Figure S2.6: *Contact map of strong simulated ^{1}H-^{1}H NOE intensities for MTSL-Cys-Aβ42 from REMD simulations that are dominated by a single contact*. Strong experimental intensities that define β-strands 4-7 and 10-12 (orange), β-strands 3-6 and 31-41 (pink), β-strands 16-21 and 29-36 (cyan), β-strands 27-31 and 33-38 (grey), β-strands 34-36 and 39-40 (yellow) and hydrophobic contacts (outlined black) indicated in the map. The color-coding for helices and turns are shown in the figure.

Chapter 3

Comparison of enhanced sampling techniques and force fields for IDP simulations

Complete characterization of structural ensembles of intrinsically disordered peptides (IDPs) has been quite challenging both experimentally and computationally because of the dynamic nature of IDP structural ensembles. The most common way to generate disordered ensembles via computational means is to use replica exchange molecular dynamics (REMD), an enhanced sampling technique, combined with an off the shelf pairwise-additive protein and water force field. However, this popular approach has been shown to predict overly collapsed and ordered structure for IDPs when compared to experimental solution results. In order to obtain better agreement with experimental results, a number of research groups have made parameter modifications to standard force fields to make them more suitable for IDP simulations (although they are then incompatible for use with folded proteins). In this work, we have compared the computationally generated ensembles of the Amyloid-β42 (Aβ42) and Amyloid-β43 (Aβ43) using two enhanced sampling techniques, the standard REMD and our recent new enhanced sampling temperature cool walking (TCW) technique, and simulated using both standard and IDP-specific force fields. Our simulations show that that the TCW technique generates disordered structural ensembles in much better agreement with experiments than the REMD technique. We also see substantial improvement in ensemble characterization when using IDP-tailored force fields, but the relative magnitude of this improvement is less when compared to the TCW technique.

3.1 INTRODUCTION

Intrinsically disordered proteins (IDPs) are a class of proteins that do not adopt a single equilibrium structure and sample an ensemble of fully/partially disordered structures. IDPs have been found to play an important role in the signaling, transcription and regulation within the cell, and also have been implicated in various neurodegenerative diseases and cancer pathways [1]. Hence, it is extremely crucial to have accurate structural characterization tools for IDPs, however experimental techniques like nuclear magnetic resonance spectroscopy (NMR) can only report on the ensemble-averaged properties for IDPs since the time scales for NMR experiments are slower than the time scales for inter-conversion between IDP substates. In order to have a detailed understanding of the substates present in an IDP ensemble, usually a combination of both experimental tools and molecular simulation techniques is employed[2-5].

One of the most common ways to generate disordered ensembles via computational means is to use replica exchange molecular dynamics[6] (REMD), an enhanced sampling technique, combined with an off the shelf pairwise-additive protein and water force field. The Amber[7], GROMOS[8], OPLS-AA[9] and Charmm[10, 11] force fields are the commonly used protein models for IDPs in combination with TIP3P or TIP4P/TIP4P-Ew water models (and will be referred to here on out as standard force fields). Since pairwise additive force fields are parameterized using mostly folded protein crystallographic data[12], they are found to be not ideal for disordered force field simulations since they introduce a bias toward more collapsed and ordered IDP structural ensembles when compared to native conditions[13-15]. Furthermore, multiple

groups have shown that the IDP structural ensembles generated using different standard force fields vary considerably in terms of secondary structure content[16-18]; using $A\beta_{16-22}$ peptide, Nguyen *et al.* [16] have demonstrated that Amber99 predicts more helical structures, GROMOS96 favors more β-strand structures while OPLS doesn't have a particular secondary structure preference.

In order to make better and more uniform predictions across different IDPs, recently a few research labs have modified parameters in standard force fields to tailor them to IDP applications. The major difference between a folded protein and an IDP is that IDP backbones are more extended and disordered with more solvent exposure, which these IDP-specific force fields try to take into account either by adjusting the water-protein London dispersion interactions to be more favorable[13, 14], refining the peptide backbone parameters to produce more expanded structures[13, 15], or changing the salt-bridge interactions[13]. Henriques *et al.*[19, 20] have shown that these modified force fields do give better agreement with experimental data for the disordered Histatin 5 peptide.

The other important and often not enough emphasized aspect of generating IDP structural ensembles is the enhanced sampling technique itself. Since IDPs have a relatively flat energy landscape with multiple local minima, it takes a substantial amount of time to sample all possible configurations with the correct weighting of sub-populations using traditional molecular dynamics simulations. The most common enhanced sampling technique used in the IDP field at present is the replica exchange molecular dynamics[6] (REMD) method. Recently, we have developed the temperature cool walking (TCW) technique[21, 22], an alternative to REMD for peptide simulations, which can generate equilibrium structural ensembles and has been shown to converge faster for small systems such as alanine dipeptide and met-encephalin peptides where well-defined and quantitative metrics of convergence are available.

In this chapter, we compare different combinations of enhanced sampling techniques (REMD and TCW) and protein force fields (standard and IDP-specific) and show that for accurate sampling of disordered protein structural ensembles, the enhanced sampling method is as equally important as the protein and water force field. We have simulated the disordered structural ensemble of the Alzheimer's Amyloid-β42 (Aβ42) and Amyloid-β43 (Aβ43) peptides, which differ only by a threonine residue in the C-terminus. Fawzi and co-workers have shown that the two Aβ peptides have very similar structural features throughout the sequence with only minor differences in the C-terminus. Our results show that the biggest improvement in ensemble prediction is obtained by switching from REMD to TCW for a fixed simulation length, and different force field combinations.

3.2 METHODS

Peptide Simulations. The starting Aβ42 and Aβ43 configurations were created using the *tleap*[23] module, and the peptides were subsequently minimized and equilibrated in NPT ensemble at 1 bar to obtain the correct density. The box sizes were approximately 60Å on each side. Four different simulations were carried out for each of the two peptides: (1) REMD simulations using the standard Amberff99sb[7] protein force field and the TIP4P-Ew[24] water model; (2) TCW simulations using the standard Amberff99sb protein force field and the TIP4P-Ew water model; (3) TCW simulations using the Amberff99sb-ildn[25] protein force field and the IDP-specific TIP4P-D[14] water model; and (4) TCW simulations using the IDP-specific Charmm36m[13] protein force field and the Charmm-TIP3P[11] water model. For each of the four simulations for Aβ42 and Aβ43, we ran two independent trajectories with different starting configurations and velocities

for 100 ns. The first 50 ns was considered as the equilibration time, with only the last 50 ns being analyzed for the results. Thus, the final Boltzmann weighted ensembles contained 10000 structures at 287 K.

REMD Simulation Protocol. The Amber14[23] molecular dynamics package was used to perform the REMD simulations with 58 temperature replicas in the temperature range 287-450K. The temperature schedule was chosen such that the exchange probability is maintained in the range 18-22%. Exchanges between neighboring replicas were attempted every 0.5 ps. The REMD simulations were performed in the NVT ensemble with a time-step of 1 fs and with constraints on the heavy atom hydrogen bonds. A Langevin thermostat was used to maintain constant temperature. Particle-mesh Ewald was used to calculated long-range electrostatics with a 9.0 Å cutoff for the real space electrostatics and Lennard-Jones forces.

TCW Simulation Protocol: TCW is a faster sampling alternative to REMD and uses only two temperature replicas to generate a equilibrium ensemble at the target temperature[21, 22]. Starting with expanded structures at the high temperature, sequential cooling is performed to obtain structures at the low temperature such that detailed balance is satisfied. The high and low temperature replicas were at 456.2 K and 287 K, with 50 intermediate temperatures for the cooling schedule. Temperature was regulated using an Andersen thermostat[26]. Ewald summation was used for the long-range, with a cutoff of 9.5 Å for the real space electrostatics and Lennard-Jones forces. Exchanges were attempted every 8 ps. TCW simulations were performed in the OpenMM software package[27].

Trajectory Analysis. The structural ensembles were analyzed using both the cpptraj[28] module of Amber and in-house codes. The DSSP[29] criterion was used to assign secondary structures. Details about the back calculation of NMR observables have been reported in previous publications[30, 31] by the group, including chemical shifts from ShiftX2[32] and J-coupling constants[30, 31].

3.3 RESULTS

In this section, results pertaining to the four different computationally generated Aβ42 and Aβ43 structural ensembles are presented. Aβ42 and Aβ43 have been shown to have similar structural order in the monomeric form using NMR experiments with slight differences in the C-terminus[33]. Thus, one would expect the computationally generated structural ensembles to be very similar for Aβ42 and Aβ43 for a given enhanced sampling technique and force field. Figure 3.1 shows the radius of gyration distribution for the Aβ42 and Aβ43 ensembles, with Table 3.1 reporting the mean and standard deviation for the ensembles. The average secondary structure percentage for the residue in the Aβ42 and Aβ43 sequence is plotted Figure 3.2. Figure 3.1, 3.2 and Table 3.1 together with heavy atom contact maps in Figure S3.1-S3.2 demonstrate the performance of the different simulation techniques and force fields.

Using the standard protocol, the REMD+AMberff99sb simulations predict structurally dissimilar ensembles for the two peptides based on the radius of gyration data and secondary structure plots. The most striking difference is in the enhancement of the Turn 6-9 region in the Aβ43 ensemble with a simultaneous increase in the β strand formed by residues 3-5 and 10-12 with very little statistical difference between the two independent Aβ43 trajectories. However, this result is in direct contradiction with various NMR data ($^{1}H_N$ and ^{15}N chemical shift data, ^{15}N R_2, ^{15}N R_1, and heteronuclear $^{15}N-^{1}H$ nuclear Overhauser effect values) reported by Conicella *et al.*[33], which suggests no major structural differences in the 3-12 residue region for the two peptide monomers. The disagreement with experimental values is likely a sign that both the

independent simulations were stuck in the local minima for the particular N-terminal substate and this inconsistency with the experiments is resolved as soon as the simulation method is changed to TCW, with the same protein and water force fields. For the same amount of sampling time (100 ns), TCW is able to improve on the sampling and predicts ensembles with similar structural feature for the Aβ42 and Aβ43 peptides. It is difficult to distinguish between computationally generated IDP structural ensembles since the error associated with the calculation of NMR observables from computer simulations is often larger than the experimental errors[5]. However, with our current approach, we are able to make conclusions about the efficiency of the two enhanced sampling techniques since we are comparing only the similarity between the Aβ42 and Aβ43 peptides as a measure of convergences of the two simulations.

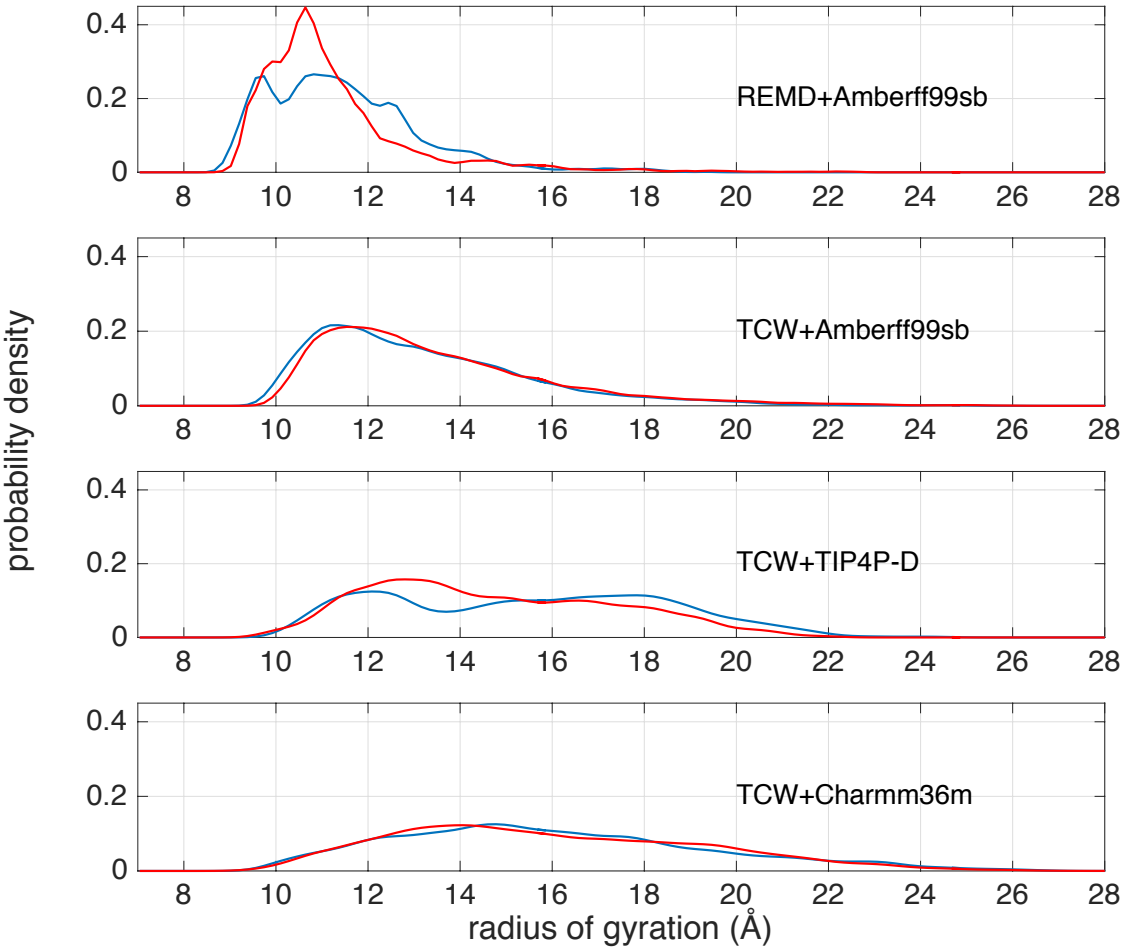


Figure 3.1: *Radius of gyration distribution for the Aβ42(blue) and Aβ43(red) ensemble for different force fields and sampling techniques.* TCW predicts similar distribution for the Aβ42 and Aβ43 ensembles unlike the REMD simulations, demonstrating it converges faster. Also, IDP-tailored force fields results in more extended structures with broader distribution.

Table 3.1: Mean and standard deviation of radius of gyration (in Å) distribution for the disordered Aβ42 and Aβ43 structural ensembles.

	REMD + Amberff99sb	TCW + Amberff99sb	TCW + TIP4P-D	TCW + Charmm36m
Aβ42	11.5 ± 1.7	13.2 ± 2.4	15.6 ± 3.2	16.0 ± 3.4
Aβ43	11.3 ± 1.8	13.5 ± 2.6	14.7 ± 2.6	16.0 ± 3.3

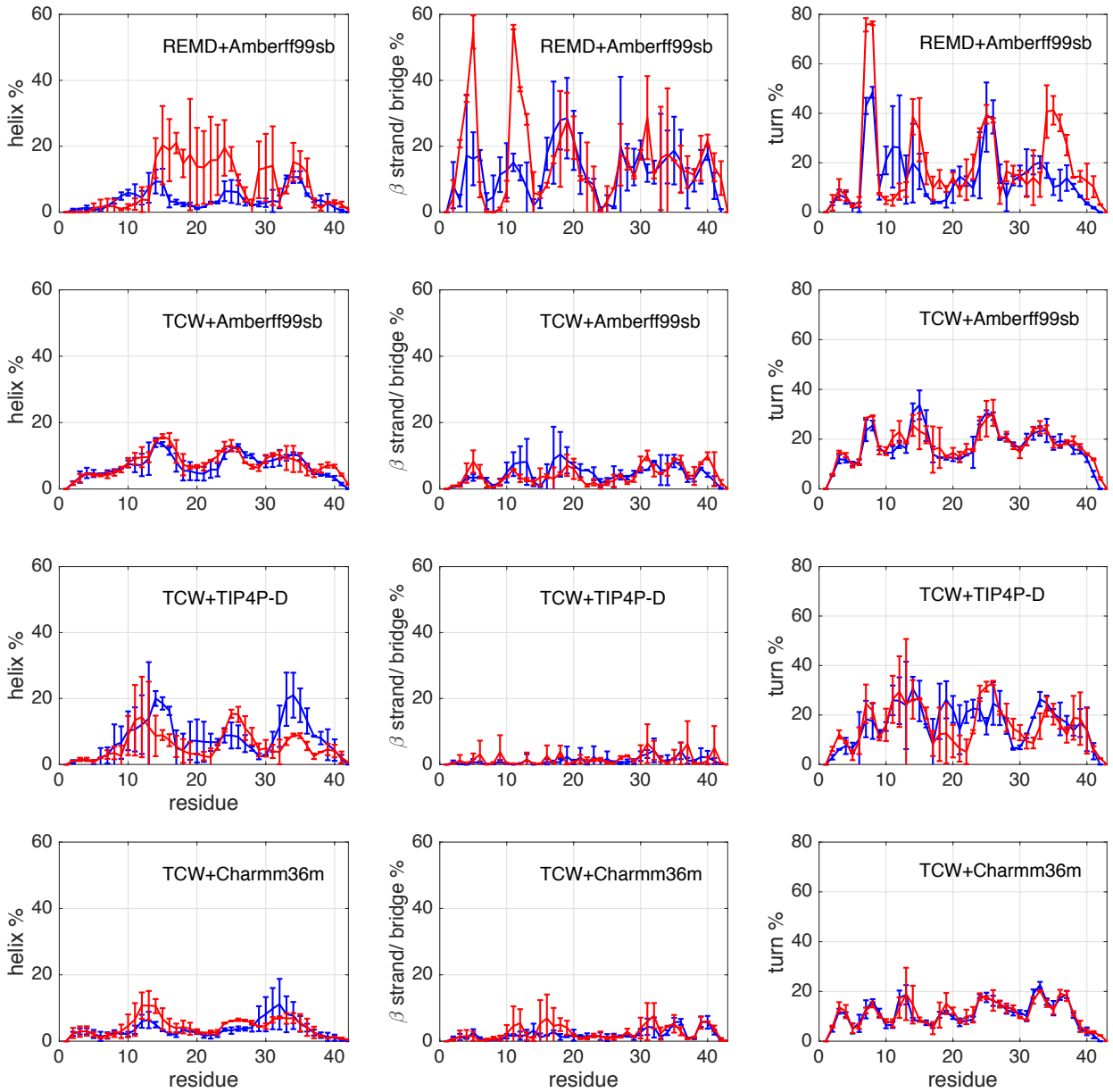


Figure 3.2: *Comparison of helix, β-strand/bridge and turn propensity for the Aβ42(blue) and Aβ43(red) sequence for different force fields and sampling techniques.* Standard deviations are based on the two independent trajectories for each simulation. TCW results in better convergence between the disordered, yet similar Aβ42 and Aβ43 ensembles.

Table 3.2: χ^2 for simulated J-coupling constants for the disordered Aβ42 and Aβ43 structural ensembles.

	REMD + Amberff99sb	TCW + Amberff99sb	TCW + TIP4P-D	TCW + Charmm36m
Aβ42	2.73	2.11	2.54	1.30
Aβ43	2.78	2.55	2.69	1.09

Next we studied the effect of standard (Amberff99sb) versus IDP-specific force fields for the TCW simulations. The IDP-specific force fields (TIP4P-D and Charmm36m) predicts more expanded structures with loss in structure order, broader radius of gyration distribution and fewer long-range non-bonded contacts as demonstrated through the radius of gyration, secondary structure and heavy atom contact map plots. This loss in structural order with IDP-specific force fields is supported by recent high resolution NMR experiments on Aβ42 by Roche *et al.*[34], showing that Aβ42 resembles a random coil under native conditions with very little long-range order. In order to have an overall comparison metric for the different simulations, we calculated the χ^2 parameter (equation 3.1) from the simulated *J*-coupling constants (${}^3J_{H_N H_\alpha}$) plotted in Figure S3.3.

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(\langle J_i\rangle_{sim} - J_{i,expt}\right)^2}{\sigma^2} \tag{3.1}$$

J_i is the *J*-coupling constant for the i^{th} residue, N is the total number of experimental ${}^3J_{H_N H_\alpha}$ observables, σ^2 is the RMSD error associated with the empirical Vuister and Bax parameters[35] used in the calculations. The subscript *sim* and *expt* refers to the simulated and the experimental values respectively. Lower values of χ^2 metric indicate better agreement with experimental *J*-coupling constants.

The REMD simulations perform the worst for both the Aβ42 and Aβ43 ensembles, while the TCW simulations with Charmm36m perform the best. It is to be noted here that the *J*-coupling constants cannot fully discriminate between random and structured ensembles[5] because of the dynamic averaging during NMR experiments. In addition to *J*-coupling, chemical shifts are also an attractive metric to compare ensembles. Unfortunately, current state-of-the-art chemical shift calculators, developed specifically for protein applications have associated intrinsic errors, for example the RMS error for ${}^1H_N$ chemical shifts in the SHIFTX2[32] calculator is 0.17 ppm, which is much larger than the experimental difference (< 0.05 ppm) between the Aβ42 and Aβ43 ensembles. Thus, chemical shift is not an ideal metric to distinguish between simulated ensembles and as expected, none of the simulations is able to capture the subtle differences in the ${}^1H_N$ chemical shifts in the C-terminus of the two peptides[33] (Figure S3.4).

3.4 DISCUSSION AND CONCLUSIONS

There can be two types of error that can occur during computational studies of IDP structural ensembles, namely sampling error and systematic error. The sampling error occurs when the simulations have not been run sufficiently long to achieve convergence, while systematic error happens when the peptide or the water is not modeled with the correct physics, that is, by using

inaccurate force fields, which are biased towards folded protein environments. We show that the sampling error is significantly minimized when using the TCW technique, which converges much faster compared to REMD simulations, while the systematic error arising out of shortcomings in the protein and water model parameterizations is reduced after switching to the recently developed IDP-specific force fields.

Since REMD simulations allow exchanges with both higher and lower temperature replicas, this results to a phenomenon called diffusiveness[22] where replicas are stuck in a small temperature range or certain energy regions for a significant amount of time. The resulting simulations end up looking converged based on standard simulation metrics, when in reality it has sampled only limited regions of the energy landscape. This problem of diffusiveness, which we encountered in our REMD simulations, is significantly reduced in the TCW simulations, which uses only two temperature replicas to generate non-local trial moves.

On the other hand, IDP-specific force fields minimize the systematic error by taking into account the increased protein-water and backbone plasticity in the disordered peptides. In the TIP4P-D water model developed by the DE Shaw group (Piana *et al.*[14]), the London dispersion interactions were re-parametrized to be more favorable. The more recent Charmm36m[13] protein force field employs similar strategy by modifying the Lennard-Jones well depth. Additionally, it also includes refined backbone and salt-bridge interaction parameters. These modifications render the resulting ensemble to be more expanded with lower structural order compared to the standard protein force fields. However, these IDP-specific force fields come with their own limitations as they are applicable to disordered proteins only and cannot be used for simulating diverse protein environments like IDP interactions with folded proteins and order to disorder transitions, and there exists a lot of scope for further improvements in the fixed charge force fields used in peptide simulations as covered in Chapter 4.

3.5 REFERENCES

1. Uversky, V. N.; Dave, V.; Iakoucheva, L. M.; Malaney, P.; Metallo, S. J.; Pathak, R. R.; Joerger, A. C., Pathological Unfoldomics of Uncontrolled Chaos: Intrinsically Disordered Proteins and Human Diseases. *Chem. Rev.* **2014,** *114* (13), 6844.
2. Marsh, J. A.; Forman-Kay, J. D., Ensemble Modeling of Protein Disordered States: Experimental Restraint Contributions and Validation. *Proteins: Struct., Func., Bioinform.* **2012,** *80* (2), 556-572.
3. Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W.-Y.; Forman-Kay, J. D., Characterization of Disordered Proteins with Ensemble. *Bioinform.* **2013,** *29* (3), 398-399.
4. Bhowmick, A.; Brookes, D. H.; Yost, M. S. R.; Dyson, H. J.; Forman-Kay, J.; Gunter, D.; Head-Gordon, M.; Hura, G. L.; Pande, V. S.; Wemmer, D. E.; Wright, P. E.; Head-Gordon, T., Finding Our Way in the Dark Proteome. *J. Am. Chem. Soc. (Perspective)* **2016,** *138* (31), 9730-9742.
5. Ball, K. A.; Wemmer, D. E.; Head-Gordon, T., Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-β Peptides. *J. Phys. Chem. B* **2014,** *118*, 6405-6416.
6. Sugita, Y.; Okamoto, Y., Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999,** *314* (1), 141-151.
7. 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., Func., Bioinform.* **2006,** *65* (3), 712-25.

8. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R.; Tironi, I. G., Biomolecular Simulation: The Gromos96 Manual and User Guide. **1996**.
9. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and Reparametrization of the Opls-AA Force Field for Proteins Via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001,** *105* (28), 6474-6487.
10. Huang, J.; MacKerell, A. D., Charmm36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013,** *34* (25), 2135-2145.
11. MacKerell Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998,** *102* (18), 3586-3616.
12. Wang, W.; Ye, W.; Jiang, C.; Luo, R.; Chen, H. F., New Force Field on Modeling Intrinsically Disordered Proteins. *Chem. Bio. Drug Des.* **2014,** *84* (3), 253-269.
13. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D., Charmm36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2016**.
14. Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E., Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015,** *119* (16), 5113-5123.
15. Best, R. B.; Mittal, J., Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *J. Phys. Chem. B* **2010,** *114* (46), 14916-14923.
16. Nguyen, P. H.; Li, M. S.; Derreumaux, P., Effects of All-Atom Force Fields on Amyloid Oligomerization: Replica Exchange Molecular Dynamics Simulations of the Aβ16–22 Dimer and Trimer. *Phys. Chem. Chem. Phys.* **2011,** *13* (20), 9778-9788.
17. Siwy, C. M.; Lockhart, C.; Klimov, D. K., Is the Conformational Ensemble of Alzheimer's Aβ10-40 Peptide Force Field Dependent? *PLOS Comput. Bio.* **2017,** *13* (1), e1005314.
18. Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H., Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015,** *11* (11), 5513-5524.
19. Henriques, J.; Skepö, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: On the Accuracy of the Tip4P-D Water Model and the Representativeness of Protein Disorder Models. *J. Chem. Theory Comput.* **2016,** *12* (7), 3407-3415.
20. Henriques, J.; Cragnell, C.; Skepö, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J. Chem. Theory Comput.* **2015,** *11* (7), 3420-3431.
21. Brown, S.; Head‐Gordon, T., Cool Walking: A New Markov Chain Monte Carlo Sampling Method. *J. Comput. Chem.* **2003,** *24* (1), 68-76.
22. Lincoff, J.; Sasmal, S.; Head-Gordon, T., Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom. *J. Chem. Phys.* **2016,** *145* (17), 174107.
23. Case, D. A. B., V.;Berryman, J.T.; Betz, R.M.;Cai, Q.; Cerutti, D.S.; Cheatham, III, T.E.; Darden, T.A.; Duke, R.E.; Gohlke, H. et al. *Amber 14*, University of California, San Francisco, 2014.

24. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an Improved Four-Site Water Model for Biomolecular Simulations: Tip4p-Ew. *J. Chem. Phys.* **2004,** *120* (20), 9665-78.
25. Lindorff‐Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved Side‐Chain Torsion Potentials for the Amber Ff99sb Protein Force Field. *Proteins: Struct., Func., Bioinform.* **2010,** *78* (8), 1950-1958.
26. Andersen, H. C., Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980,** *72* (4), 2384-2393.
27. Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D., Openmm 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013,** *9* (1), 461.
28. Roe, D. R.; Cheatham III, T. E., Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9* (7), 3084-3095.
29. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen‐Bonded and Geometrical Features. *Biopolymers* **1983,** *22* (12), 2577-2637.
30. Ball, K. A.; Phillips, A. H.; Nerenberg, P. S.; Fawzi, N. L.; Wemmer, D. E.; Head-Gordon, T., Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-β Peptides. *Biochemistry* **2011,** *50* (35), 7612-28.
31. Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T., Differences in β-Strand Populations of Monomeric Aβ40 and Aβ42. *Biophys. J.* **2013,** *104* (12), 2714-2724.
32. Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S., SHIFTX2: Significantly Improved Protein Chemical Shift Prediction. *J. Biomol. NMR* **2011,** *50* (1), 43.
33. Conicella, A. E.; Fawzi, N. L., The C-Terminal Threonine of Aβ43 Nucleates Toxic Aggregation Via Structural and Dynamical Changes in Monomers and Protofibrils. *Biochemistry* **2014,** *53* (19), 3095-3105.
34. Roche, J.; Shen, Y.; Lee, J. H.; Ying, J.; Bax, A., Monomeric Aβ1–40 and Aβ1–42 Peptides in Solution Adopt Very Similar Ramachandran Map Distributions That Closely Resemble Random Coil. *Biochemistry* **2016,** *55* (5), 762.
35. Vuister, G. W.; Bax, A., Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J (H_NH_α) Coupling Constants in ^{15}N-Enriched Proteins. *J. Am. Chem. Soc.* **1993,** *115* (17), 7772-7777.

3.6 APPENDIX

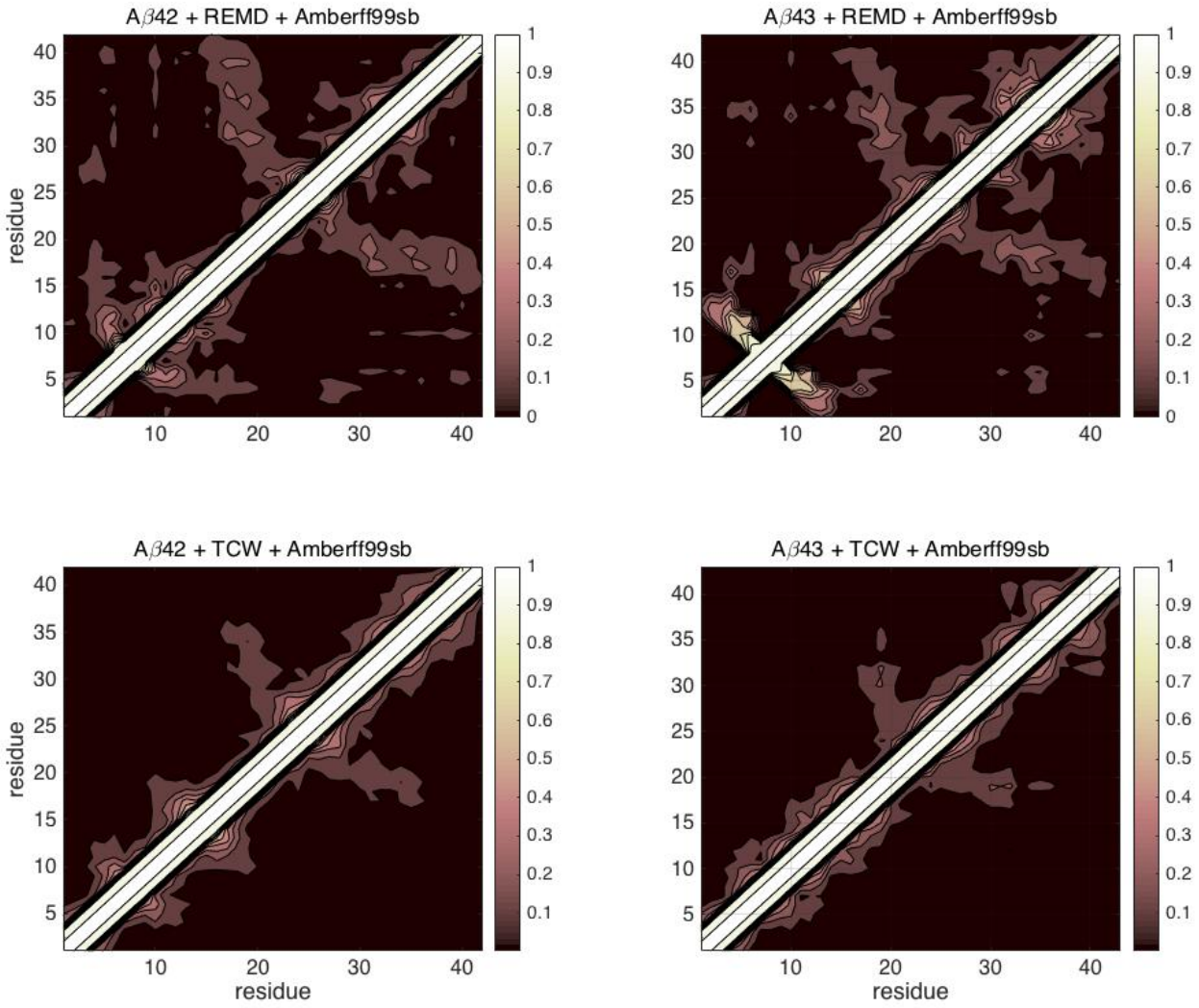


Figure S3.1: *Heavy-atom contact map of the simulated ensembles of Aβ42 and Aβ43* for REMD+Amberff99sb and TCW+Amberff99sb simulations. The contact maps show the probability of heavy atom contact between each pair of residues in peptide sequence. A contact between two residues is said to occur when any of their heavy atoms are less than 5 Å apart.

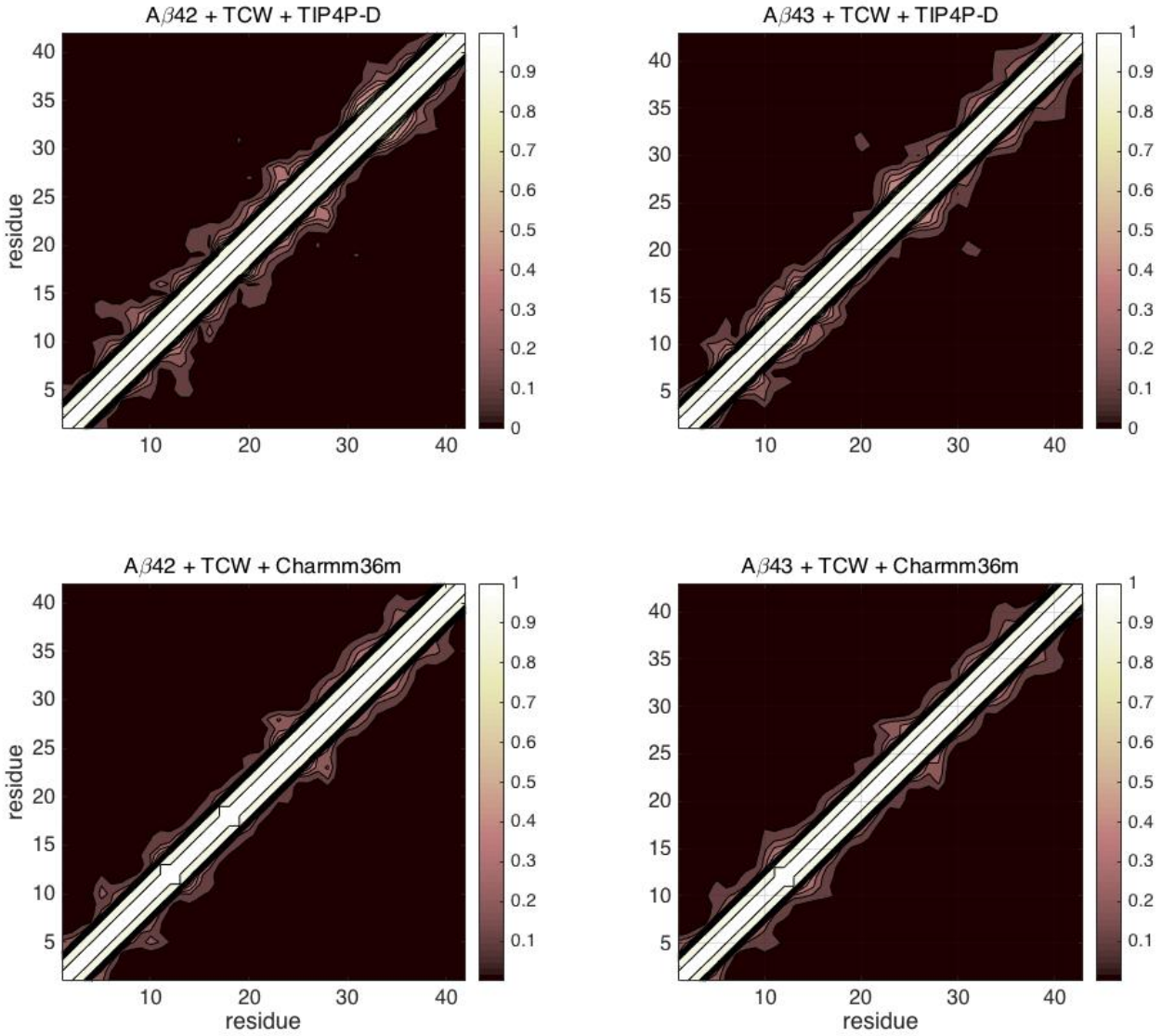


Figure S3.2: *Heavy-atom contact map of the simulated ensembles of Aβ42 and Aβ43* for TCW+TIP4P-D and TCW+Charmm36m simulations. The contact maps show the probability of heavy atom contact between each pair of residues in peptide sequence. A contact between two residues is said to occur when any of their heavy atoms are less than 5 Å apart.

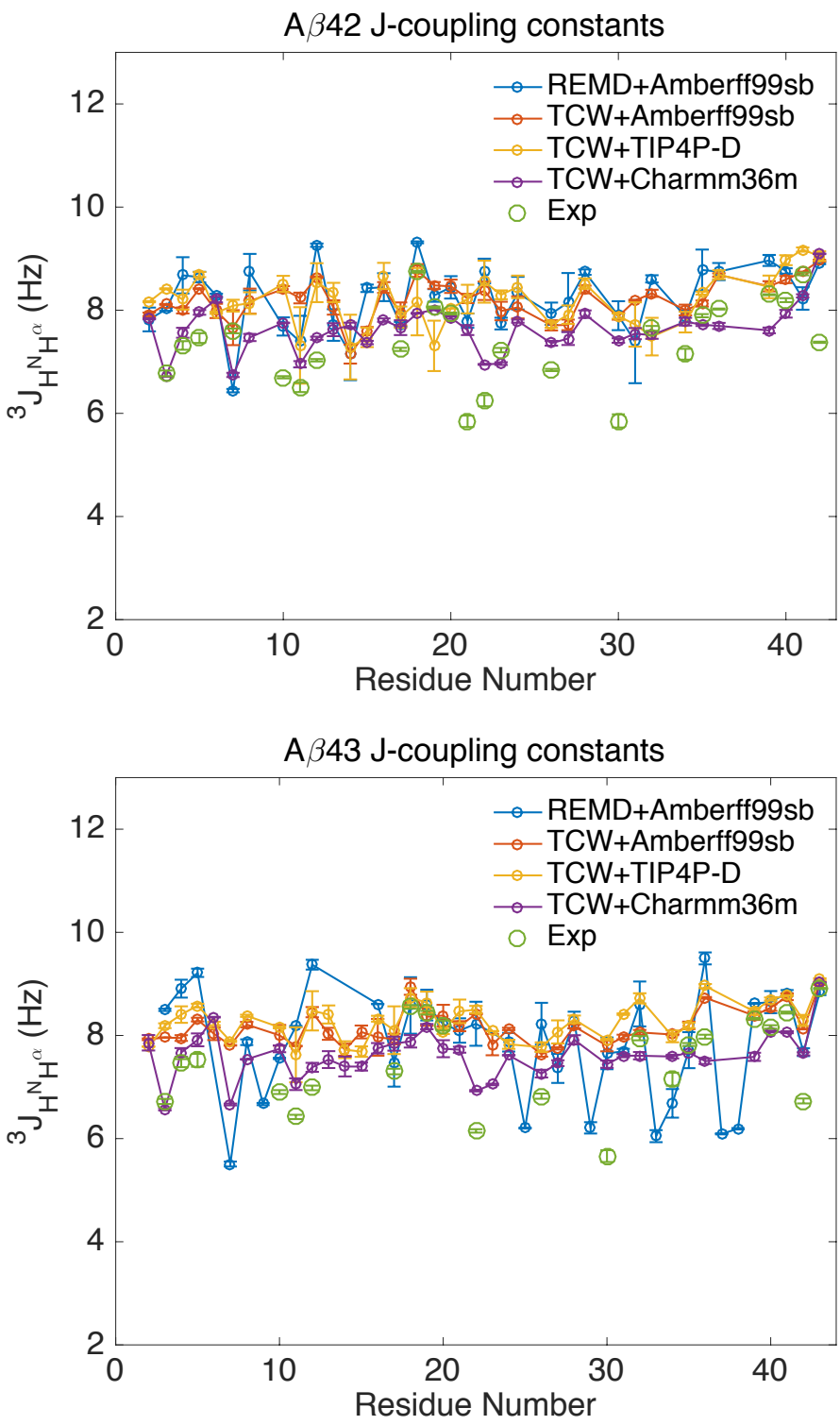


Figure S3.3: *Comparison of simulated J-coupling constants for backbone amides for the Aβ42 and Aβ43 monomeric structural ensembles with experimental data from Conicella et.al.* [33]. We have corrected the simulated values (obtained using the Vuister and Bax parameters[35]) by 10% to make the simulated data consistent with the H_NH_α 3D experiments.

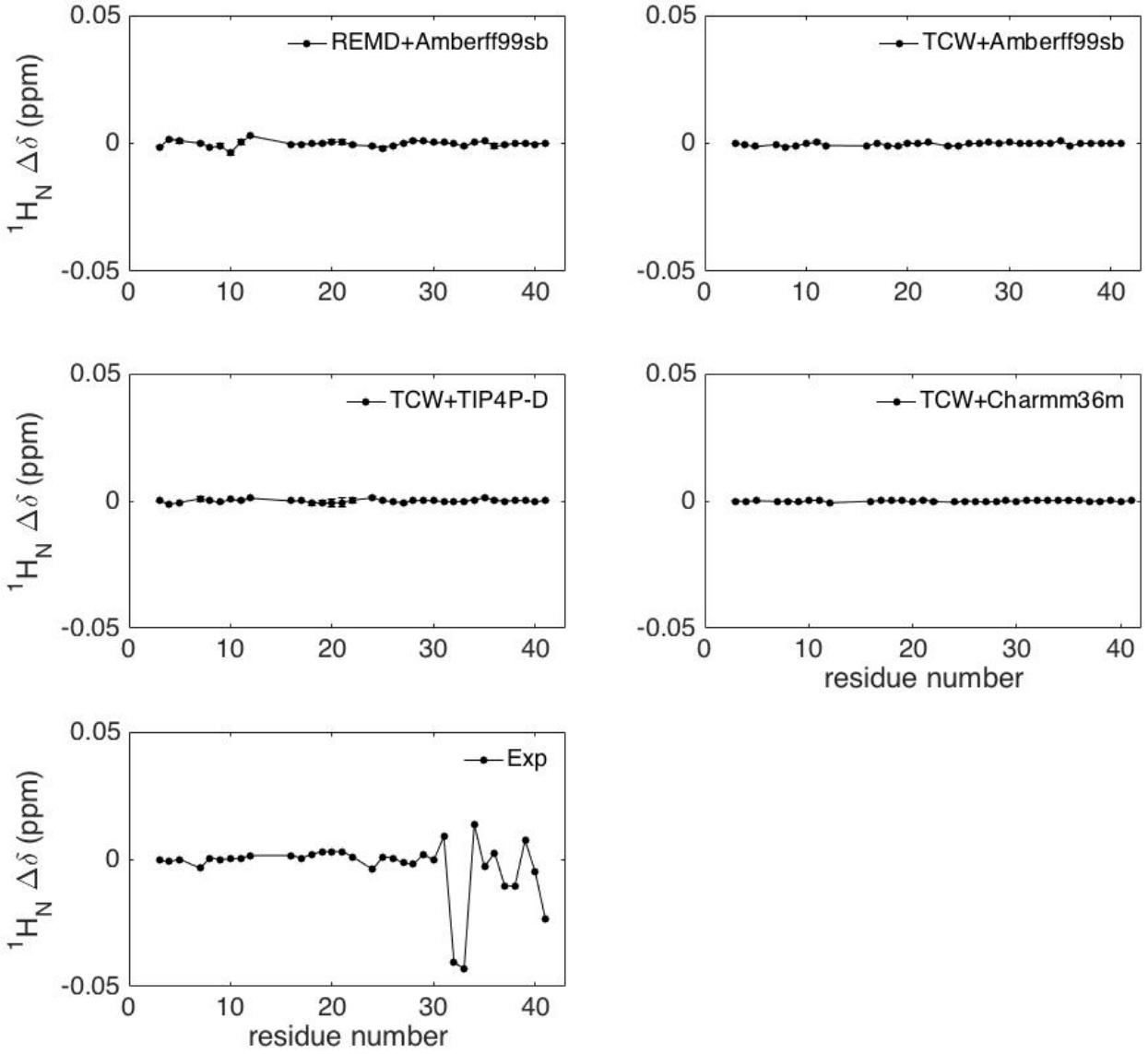


Figure S3.4: *Comparison of simulated chemical shifts difference between the Aβ43 and Aβ42 structural ensembles with experiments*[33]. The simulated values were calculated using the SHIFTX2 calculator[32].

Chapter 4

Improved Sampling of Disordered Peptide Structural Ensembles Using the Polarizable AMOEBA Force Field

Current pairwise additive force fields are found to predict more collapsed and ordered structures for intrinsically disordered peptides (IDPs), than found by experiment, due to the parameterization of a limited functional form largely designed to reproduce folded protein structures. Recent modifications of peptide-water and water-water interactions in these standard force fields have sought to increase the structural plasticity consistent with IDP behavior, but have the limitation that they no longer reproduce folded structures. Many-body potentials that include polarization should provide a better physical model for capturing both folded proteins and proteins with intrinsic disorder. The polarizable AMOEBA force field is shown here to best reflect the structural ensemble of the disordered Histatin 5 peptide, assessed by comparison with well-defined small angle X-ray scattering measurements and secondary structure propensities evaluated from circular dichroism and NMR, when compared to different fixed-charge force fields containing IDP-specific parameter modifications.

4.1 INTRODUCTION

Intrinsically disordered peptides (IDP) are a class of peptides lacking a well-defined equilibrium structure under native conditions[1], instead rapidly interconverting between different conformations in solution. While experimental techniques like nuclear magnetic resonance (NMR) spectroscopy can provide restraints on the ensemble-averaged structural properties for IDPs, a complete characterization of the structural substates of the IDPs requires further elaboration with computational models[1-4], which can be verified with or refined against experimental observables[5-7]. At present all IDP ensembles have been generated with pairwise additive force fields whose limited functional forms have been parameterized to reproduce well-folded protein data[8-10]. The structural, dynamic and solvation properties of IDPs are quite different from folded proteins, and the pairwise additive models have been found to generate ensembles for IDPs that are more structured and compact than would be consistent with experimental observations[11-14].

To make better predictions, a few research groups have proposed parameter changes to standard force fields for IDP applications. In the TIP4P-D water model, Piana *et al.* have increased the dispersion coefficient C_6 by 50% to make London dispersion interactions more favorable[15], and when combined with Amberff99sb-ildn[9] (A99SBildn) for the protein, resulted in more expanded peptides with improved agreement with experimental NMR and small angle X-ray scattering (SAXS) data from small globular proteins. Best and Mittal[14] have shown that the combination of the TIP4P/2005 water model and backbone parameter modifications of the Amberff03ws (A03WS) protein model produces better temperature dependence of chemical shifts and better reproduction of helix-coil transition thermodynamics for the 15-residue peptide Ac-$(AAQAA)_3$-NH_2 peptide. Independently, Henriques *et al.*[13, 17] have shown that both A99SBildn/TIP4P-D and A03WS/TIP4P/2005 perform better than the unmodified pairwise additive force fields for the disordered Histatin 5 peptide, although in both cases they exhibit too high a secondary structure content such that the radius of gyration (Rg) is smaller

than that found from small angle X-ray scattering. More recently, a modification to the standard CHARMM force field was introduced (CHARMM36m[16]) with refined peptide backbone parameters and salt-bridge interactions, as well as an increased Lennard-Jones (LJ) well depth, again which better reproduce unstructured ensembles of arginine–serine peptide, the FG-nucleoporin peptide, N-terminal fragment of white lysozyme and the N-terminal domain of HIV-1 integrase. Even so, Huang *et al.* found that while developing the CHARMM36m force field[16] that they could not determine a universal LJ well depth parameter (ε_H) that was applicable to all IDP systems, thus highlighting the need of a more general force field.

The primary problem with "IDP-tailored" force fields is that they are not transferable to different IDPs and any protein that has structural order – i.e. they can't be applied when simulating interactions of IDPs with folded proteins[18], disorder-to-order transition in peptides[19] and folded proteins with only intrinsically disordered regions[20]. Because IDPs can adopt a range of structures from highly solvent-exposed environments to ordered states when folding upon binding to other proteins, they require a force field that is many-body in character. Polarization the leading order many-body interaction that can tune its electrostatic response depending on evolving environments, and thus are expected to be more accurate in simulating both folded proteins as well as IDP systems. In this work, we have investigated the polarizable AMOEBA force field[23, 27, 28], with no parameter modifications, to simulate the short 24-residue disordered Histatin 5 peptide. The resulting Histatin 5 structural ensemble shows better agreement with experimental radius of gyration[29] and secondary structure data[30-32] as compared to different fixed-charge force fields that have been modified for IDP simulations[14, 15].

4.2 METHODS

The initial extended structure of Histatin-5 (Asp_1-Ser-His-Ala-Lys_5-Arg-His-His-Gly-Tyr_{10}-Lys-Arg-Lys-Phe-His_{15}-Glu-Lys-His-His-Ser_{20}-His-Arg-Gly-Tyr_{24}) initial was created using the *Leap*[33] module of the Amber14 simulation package[3]. Five Cl^- atoms were added to maintain charge neutrality. The AMOEBA 2003 water model[23] and AMOEBA 2013[28] protein parameters were used to model the system. All equilibration steps were performed using standard molecular dynamics (MD) simulation with a 1 femtosecond (fs) time-step in the OpenMM package[34]. The peptide was initially solvated using 1 nanometer (nm) water padding on each side, and simulated at 500 K for 1 nanosecond (ns) in the NVT ensemble to collapse the peptide, after which the peptide was re-solvated using a smaller cubic box with side lengths of 5.91 nm; the simulation box contained 6166 water molecules. The re-solvated peptide was equilibrated in the NVT ensemble first at 500K for 1 ns, then at 300K for 1 ns. In the final equilibration step, the peptide was run in the NPT ensemble at 300K to achieve the correct density, and structures at the maximum probable density were chosen for subsequent production simulations.

Polarizable force fields such as AMOEBA are computationally slower because of the additional expense of solving the polarizable point dipoles to self-consistency[23], thereby exacerbating the already high computational cost of sampling using the replica exchange molecular dynamics[24] (REMD) method across ~50 temperature replicas. We have recently advanced a new enhanced sampling procedure, Temperature Cool Walking[25] (TCW), which is a non-equilibrium alternative to REMD simulations. Since TCW has been shown to exhibit faster convergence for ensemble-averaged properties compared to REMD for different peptide

systems[26], and is computationally cheaper because it utilizes only two temperature replicas, its use for IDP simulations makes the study of polarizable force fields feasible.

We have performed three independent TCW simulations, using a low and high temperature replica at 300K and 456.21K, respectively, with a cooling (heating) schedule that samples 50 intermediate temperatures. The acceptance ratio between the replicas was maintained close to ~68%. The peptide was simulated for 50 fs of cooling at each intermediate temperature during the annealing protocol. We use a RESPA[35] integrator with 2 fs and 1 fs time-steps for the non-bonded and bonded interactions respectively. Exchange between replicas was attempted every 5 ps. Three independent production runs were performed with different starting structures and velocities. The resulting Histatin-5 structural ensemble was equivalent to a total 100 ns of simulation time, with the first 10 ns being discarded as equilibration for the TCW simulations. More details about the TCW method can be found in a previous publication[26]. TCW is implemented in the OpenMM package[34]. The cpptraj[37] module of Amber14 was used for analyzing the radius of gyration of the peptide. DSSP criterion[38] was used to analyze secondary structures for the peptide.

4.3 RESULTS

In what follows we compare our AMOEBA simulation results with the already published results generated using a standard fixed-charge force field and force fields that have been modified for IDPs for Histatin-5 structural ensembles[17]. Figure 4.1 shows the R_g probability distribution for Histatin-5 for these different force fields, while Table 4.1 provides the mean and standard deviation for R_g.

Table 4.1: Mean and standard deviation for radius of gyration for Histatin-5.

Method	$\langle R_g \rangle$ (nm)
Experimental[29]	1.38
MD + A99SBildn +TIP3P[17]	1.00 ± 0.15
MD + A03WS + TIP4P/2005[17]	1.30 ± 0.20
MD + A99SBildn + TIP4P-D[17]	1.32 ± 0.20
TCW + AmoebaPRO13	1.40 ± 0.22

We see that the standard pairwise additive force field A99SBildn/TIP3P performs the worst since it doesn't have any parameter modifications, and predicts more compact globular structures than all of the other force fields. For the other three force fields, A03WS/TIP4P/2005, A99SBildn/TIP4P-D and AMOEBA, the shape and the spread of the R_g distribution is similar, however the AMOEBA $\langle R_g \rangle$ prediction (1.40 ± 0.22 nm) is closest to the experimental $\langle R_g \rangle$ (1.38 nm) obtained from SAXS[29].

Experimental CD[31, 32] and NMR[30, 32] studies have found that in aqueous solutions Histatin-5 doesn't display any helical conformations, although Histatin-5 prefers α-helical conformations in non-aqueous solvents. From Figure 4.2, we can see that the resulting structural ensemble from all the IDP-modified fixed-charge force fields have too much helical content, while the structural ensemble generated using AMOEBA has close to zero percent

helical content. Thus, the AMOEBA force field does remarkably well in secondary structure signatures that are in better agreement with the NMR and CD data on Histatin-5.

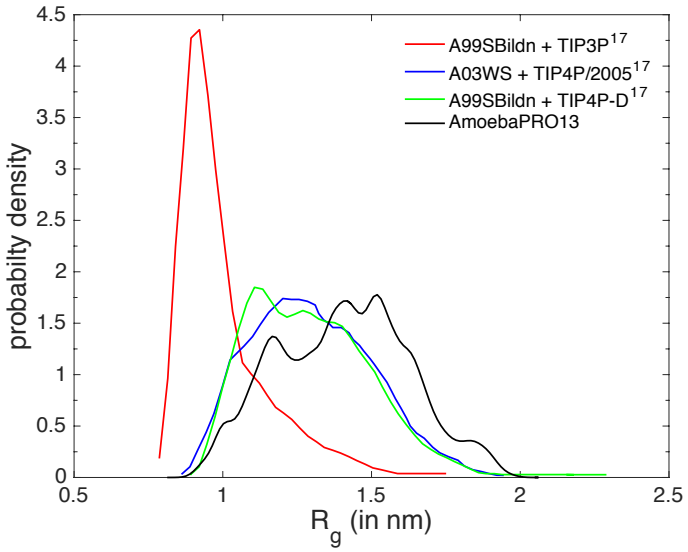


Figure 4.1: Probability density estimates for Histatin-5 radius of gyration. The data for the fixed-charge force fields was obtained from Henriques *et al.*[13, 17].

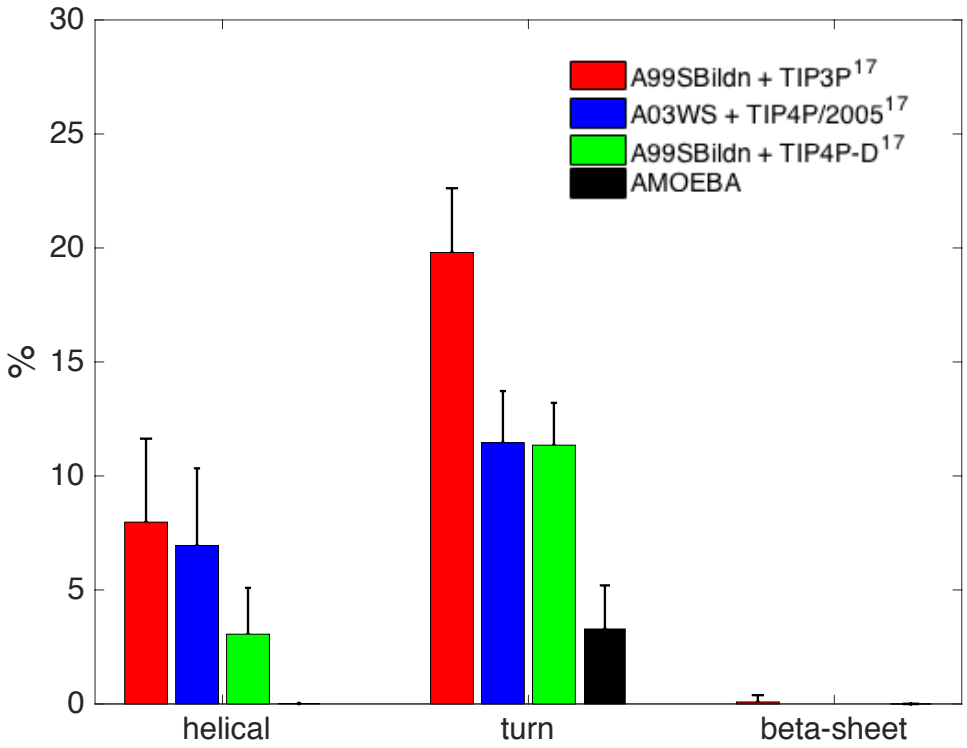


Figure 4.2: Average percentages of different structural features for the disordered Histatin-5 peptide. The data for the fixed-charge force fields was obtained from Henriques *et al.*[13, 17].

4.4 DISCUSSION AND CONCLUSIONS

Because of their structural plasticity, the free IDP state has more solvent-accessible surface-area than folded proteins, but have the additional complexity of being able to fold upon binding or form complexes with other disordered partners. Given this range of protein environments, it is apparent that pairwise additive potentials and their IDP-specific modifications have inadequate physics to capture this diversity in IDP behavior. Our analysis shows that simulations using the general and transferable AMOEBA polarizable model generates more accurate structural ensembles for the Histatin-5 test system than do any fixed-charge force fields, including those that have been specifically modified to better reproduce IDP behavior. We are currently further verifying the accuracy of the AMOEBA force field using the 15 residue peptide Ac-$(AAQAA)_3$-NH_2[14] as a test system.

In summary, the polarizable AMOEBA force field presents a significant advance over fixed-charge force field for IDP simulations, as it does not require any problem-specific parameterization for IDPs, and can be used as a general force field for different types of IDPs and their complexes. Our ultimate goal is to be able to generate accurate structural ensembles for physiologically relevant disordered peptides like the Alzheimer's Amyloid-β peptide[5] using the combination of TCW and the AMOEBA force field.

4.5 REFERENCES

1. Tompa, P., Intrinsically Disordered Proteins: A 10-Year Recap. *Trends Biochem. Sci.* **2012,** *37* (12), 509-516.
2. Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M., Mapping Long-Range Interactions in α-Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005,** *127* (2), 476-477.
3. Nerenberg, P. S.; Yedvabny, E.; Ball, K. A.; So, C.; Head-Gordon, T., Computational Characterization of the Disordered Ensembles of Vasopressin and Oxytocin. *Biophys. J.* **2014,** *106* (2), 481a-482a.
4. Ball, K. A.; Phillips, A. H.; Nerenberg, P. S.; Fawzi, N. L.; Wemmer, D. E.; Head-Gordon, T., Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-β Peptides. *Biochemistry* **2011,** *50* (35), 7612-28.
5. Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T., Differences in β-Strand Populations of Monomeric Aβ40 and Aβ42. *Biophys. J.* **2013,** *104* (12), 2714-2724.
6. Ball, K. A.; Wemmer, D. E.; Head-Gordon, T., Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-β Peptides. *J. Phys. Chem. B* **2014,** *118*, 6405-6416.
7. Marsh, J. A.; Forman-Kay, J. D., Ensemble Modeling of Protein Disordered States: Experimental Restraint Contributions and Validation. *Proteins: Struct., Func., Bioinform.* **2012,** *80* (2), 556-572.
8. Huang, J.; MacKerell, A. D., Charmm36 All‐Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013,** *34* (25), 2135-2145.
9. Lindorff‐Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved Side‐Chain Torsion Potentials for the Amber ff99sb Protein Force Field. *Proteins: Struct., Func., Bioinform.* **2010,** *78* (8), 1950-1958.

10. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F., Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *European Biophys. J.* **2011,** *40* (7), 843.
11. Siwy, C. M.; Lockhart, C.; Klimov, D. K., Is the Conformational Ensemble of Alzheimer's Aβ10-40 Peptide Force Field Dependent? *PLOS Comput. Bio.* **2017,** *13* (1), e1005314.
12. Nerenberg, P. S.; Head-Gordon, T., Optimizing Protein− Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J. Chem. Theory Comput.* **2011,** *7* (4), 1220-1230.
13. Henriques, J.; Cragnell, C.; Skepö, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J. Chem. Theory Comput.* **2015,** *11* (7), 3420-3431.
14. Best, R. B.; Mittal, J., Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *J. Phys. Chem. B* **2010,** *114* (46), 14916-14923.
15. Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E., Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015,** *119* (16), 5113-5123.
16. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D., Charmm36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2016**.
17. Henriques, J.; Skepö, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: On the Accuracy of the Tip4p-D Water Model and the Representativeness of Protein Disorder Models. *J. Chem. Theory Comput.* **2016,** *12* (7), 3407-3415.
18. Wang, Y.; Chu, X.; Longhi, S.; Roche, P.; Han, W.; Wang, E.; Wang, J., Multiscaled Exploration of Coupled Folding and Binding of an Intrinsically Disordered Molecular Recognition Element in Measles Virus Nucleoprotein. *Proc. Natl. Acad. Sci.* **2013,** *110* (40), 3743-3752.
19. Moritsugu, K.; Terada, T.; Kidera, A., Disorder-to-Order Transition of an Intrinsically Disordered Region of Sortase Revealed by Multiscale Enhanced Sampling. *J. Am. Chem. Soc.* **2012,** *134* (16), 7094-7101.
20. Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R., Structure of Tumor Suppressor P53 and Its Intrinsically Disordered N-Terminal Transactivation Domain. *Proc. Natl. Acad. Sci.* **2008,** *105* (15), 5762-5767.
21. Johnson, M. E.; Malardier-Jugroot, C.; Head-Gordon, T., Effects of Co-Solvents on Peptide Hydration Water Structure and Dynamics. *Phys. Chem. Chem. Phys.* **2010,** *12* (2), 393-405.
22. Johnson, M. E.; Malardier-Jugroot, C.; Murarka, R. K.; Head-Gordon, T., Hydration Water Dynamics near Biological Interfaces. *J. Phys. Chem. B* **2008,** *113* (13), 4082-4092.
23. Ren, P.; Ponder, J. W., Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003,** *107* (24), 5933-5947.
24. Sugita, Y.; Okamoto, Y., Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999,** *314* (1), 141-151.
25. Brown, S.; Head‐Gordon, T., Cool Walking: A New Markov Chain Monte Carlo Sampling Method. *J. Comput. Chem.* **2003,** *24* (1), 68-76.

26. Lincoff, J.; Sasmal, S.; Head-Gordon, T., Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom. *J. Chem. Phys.* **2016,** *145* (17), 174107.
27. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A., Current Status of the Amoeba Polarizable Force Field. *J. Phys. Chem. B* **2010,** *114* (8), 2549.
28. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., The Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013,** *9* (9), 4046.
29. Cragnell, C.; Durand, D.; Cabane, B.; Skepö, M., Coarse‐Grained Modelling of the Intrinsically Disordered Protein Histatin 5 in Solution. Monte Carlo Simulations in Combination with Saxs. *Proteins: Struct., Func., Bioinform.* **2016**.
30. Raj, P. A.; Marcus, E.; Sukumaran, D. K., Structure of Human Salivary Histatin 5 in Aqueous and Nonaqueous Solutions. *Biopolymers* **1998,** *45* (1), 51-67.
31. Raj, P. A.; Edgerton, M.; Levine, M., Salivary Histatin 5: Dependence of Sequence, Chain Length, and Helical Conformation for Candidacidal Activity. *J. Biol. Chem.* **1990,** *265* (7), 3898-3905.
32. Brewer, D.; Hunter, H.; Lajoie, G., NMR Studies of the Antimicrobial Salivary Peptides Histatin 3 and Histatin 5 in Aqueous and Nonaqueous Solutions. *Biochem. Cell Biol.* **1998,** *76* (2-3), 247-256.
33. 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., Func., Bioinform.* **2006,** *65* (3), 712-25.
34. Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D., Openmm 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013,** *9* (1), 461.
35. Tuckerman, M.; Berne, B. J.; Martyna, G. J., Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992,** *97* (3), 1990-2001.
36. Andersen, H. C., Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980,** *72* (4), 2384-2393.
37. Roe, D. R.; Cheatham III, T. E., Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9* (7), 3084-3095.
38. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen‐Bonded and Geometrical Features. *Biopolymers* **1983,** *22* (12), 2577-2637.

Chapter 5

Mechanism of Nucleation and Growth of Aβ40-Fibrils from All-Atom and Coarse-Grained Simulations

In this work, we characterize the nucleation and elongation mechanisms of the "disease" polymorph of the amyloid-β 40 (Aβ40) fibril by using an off-lattice coarse-grained (CG) protein model. After determining the nucleation size and subsequent stable protofibrillar structure from the CG model, validated with all-atom simulations, we consider the "lock and dock" and "activated monomer" fibril elongation mechanisms for the protofibril by statistical additions of a monomer drawn from four different ensembles of the free Aβ40 peptide to grow the fibril. Our CG model shows that the dominant mechanism for fibril elongation is the "lock and dock" mechanism across all monomer ensembles, even when the monomer is in the activated form. Although our CG model finds no thermodynamic difference between the two fibril elongation mechanisms, the "activated" monomer is found to be kinetically faster by a factor of two for the "locking" step compared with any other structured or unstructured monomer ensemble.

Sasmal, S.; Schwierz, N.; Head-Gordon, T., Mechanism of Nucleation and Growth of Aβ40 Fibrils from All-Atom and Coarse-Grained Simulations. *J. Phys. Chem. B* **2016**, 120, 12088-12097.

5.1 INTRODUCTION

The ability of the short 39-42 residue intrinsically disordered peptide amyloid-β (Aβ) to self-assemble into larger aggregates has long been ascertained to be a hallmark of Alzheimer's disease [1]. The peptide in its monomeric form is non-toxic, but it is prone to form both small soluble (oligomers) and large insoluble aggregates (fibrils) that get deposited in the brain of Alzheimer's patients as extracellular plaques [2]. Although there is mostly consensus that the lower molecular weight oligomers are the neurotoxic species [3, 4], other work has argued that cytotoxicity of the insoluble fibrils could arise depending on the type of polymorph formed [5]. Furthermore, recent work has indicated that the distinction between the two disease hypotheses may not be absolute, since oligomerization and early fibril formation pathways may be coupled. Chimon et al. [6] and Ahmed and co-workers [7] have demonstrated that the early fibrillization pathway steps are connected through oligomerization intermediates, and fibrils have also been shown to catalyze toxic oligomer formation via secondary nucleation mechanisms [8]. Thus, it remains important to understand the early stages of the Aβ fibril nucleation and elongation pathway, given its potential connection to the disease state. . For our work, protofibril refers to fibrils, much smaller than micrometer in length and composed of three protofilaments of parallel intermolecular β sheets

Amyloid fibril elongation is a complex process that can occur either by addition of structured/unstructured monomers at the fibril tips[9] or by association of two smaller protofibrils to form a larger one [10]. Recent experiments on the amyloidogenic α-synuclein fibril [9] using single-molecule localization microscopy have shown that monomer addition is the primary mechanism of fibril elongation, and amyloid fibrils have been experimentally shown to exhibit

both unidirectional [11] and bidirectional growth [12, 13]. In addition, fibrils do not have a uniform growth rate [14], which can be explained by the "stop-and-go" model [15]. According to this model, the fibrils interchange between the active "go" and inactive "stop" model, with only the active form taking part in amyloid elongation; any incorrect monomer addition makes the fibril template temporarily inactive, and monomer addition can resume only after the incorrectly added monomer has dissociated or attached itself in the correct alignment. The different spatial arrangements of the protofilaments that give rise to variations in the fibril structures have also been attributed to non-uniform fibril growth rates, highlighting the importance of understanding fibril polymorphism and the structural fluctuations that might contribute to different growth rates and mechanisms [9].

For a fibril in the "go" or active state, there are two proposed mechanisms for fibril elongation by monomer addition: the two-step "dock-lock" mechanism[10, 16-18] and the one-step "activated" monomer mechanism [17, 18]. The mechanistic model of the "dock-lock" mechanism is

$$A\beta_{monomer} \rightleftharpoons A\beta_{docked} \rightarrow A\beta_{fibril} \quad (1)$$

In the first step of the "dock-lock" mechanism, a structured/unstructured monomer binds reversibly, or "docks" to the fibril surface, and then, in the "locking step", the monomer undergoes structural rearrangements to form the native contacts present in a mature fibril. In previous studies, the "locking" step is shown to be at least 2 orders of magnitude slower than the "docking" step [19].

Unlike the "dock-lock" mechanism where the monomer goes through structural rearrangements after binding, in the "activated" monomer mechanism the free "monomer" in the solution is assumed to be present in the strand-loop-strand "U" shaped configuration adopted by peptide chains in mature fibrils. Schematically this process can be shown as:

$$A\beta^{*}_{monomer} \rightarrow A\beta_{fibril} \quad (2)$$

where * represents the "activated" form. This "activated" monomer then irreversibly binds with the fibril surface and forms all the necessary native contacts in a one-step mechanism with a much faster time-scale.

There have been many computational studies looking at different aspects of Aβ fibril stability and growth processes using both atomistic and coarse-grained models to provide more molecular details that are outside of experimental purview. While all-atom simulations are able to look into specific residue and water interactions that may contribute to fibril stability and fibril growth mechanisms[10, 16, 20-23], the timescale of fibril growth (based on addition of a single monomer) is on the order of milliseconds to several minutes [24, 25] which makes atomistic studies largely intractable for kinetic questions. Coarse-grained simulations [26-29] and hybrid approaches [30] have been shown to be promising in fibrillization studies, by providing for more statistically relevant data, after appropriate validation.

In this work, we have employed a coarse-grained (CG) protein model previously developed in our lab[31-34] in order to model the AD protofibril interacting with the Aβ monomer sampled from four different monomer ensembles: 1) monomers equilibrated under fibril potential, 2) a random coil monomer ensemble, 3) monomers obtained using a combination of all-atom simulations and experimental NMR data and 4) a so called "activated" monomer ensemble, which are described in more detail below. The CG model is able to capture the most important aspects of the complex physico-chemical properties of the amino acid beads and their sequence specific interactions, as well as the secondary structure propensities and tertiary organization stabilized by van der Waals forces and short-ranged, anisotropic interactions that describe backbone hydrogen bonding. Early validation studies of our CG model demonstrated its

ability to fold globular proteins with an RMSD of ~3 Å with respect to the native experimental structure[31-33, 35-39], which is the limit possible under the coarse-graining procedure.

In the current study, we have extended the coarse-grained protein model to study the disease polymorph of amyloid fibril that has been recently solved by solid-state NMR[40]. Unlike the structural details of the oligomeric species which are mostly unknown, most amyloid fibrils are structurally defined by stacked in-register intermolecular β sheet filaments arranged with two-fold or three-fold symmetry about the fibril axis to create the insoluble fibril [41]. *In vitro*, amyloid fibrils have been shown to be polymorphic, i.e. giving rise to structures that vary with different growth conditions [42-44]. In fact we used the CG model to analyze the stability and growth mechanisms of the "agitated" and "quiescent" polymorphs of 2-stack Aβ40 fibrils [42], and familial Alzheimer's disease (FAD) mutants[26, 27, 45] in previous work. However recently Lu et al. [40] showed that fibrils grown from fibril seeds taken from the brain tissue of Alzheimer's patients are not polymorphic, i.e. when exposed to different growth conditions, there is a single thermodynamically stable structural form, which we label the AD form. The solved AD fibril structure based on solid-state NMR consists of a 3-fold symmetric arrangement of filaments, of which a few cross-sections involving only the backbone are shown in Figure 5.1.

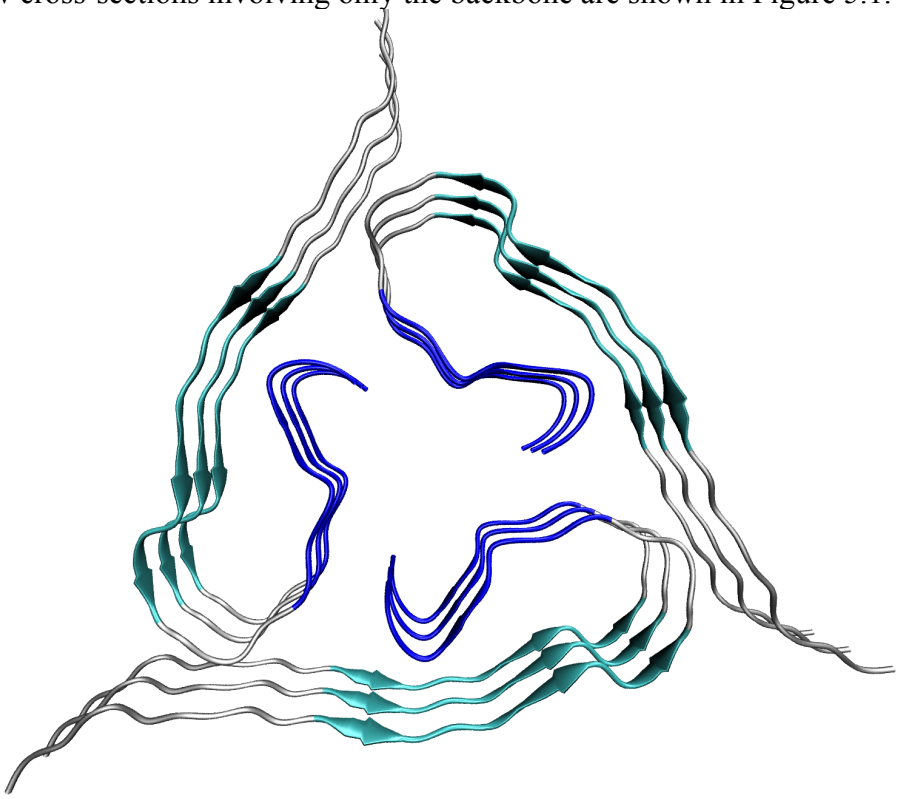


Figure 5.1: *Structure of the Aβ40 polymorph isolated from the brain tissues of an Alzheimer's patient, with 3-fold symmetry based on solid state NMR studies by Lu et al. [40].* The aqua region is the central hydrophobic core consisting of residues 12-24 and the blue region highlights the C-terminal hydrophobic residues 30-40.

In this work, we use the CG model to evaluate the nucleation and elongation free energy curves to determine the minimum number of monomers needed to stabilize the 3-stack AD Aβ40 protofibril. Atomistic simulations have been used to validate the critical nucleus size obtained from the coarse-grained simulations, and are found to be in near quantitative agreement. Based on the minimal stable size of the AD protofibril, containing 15 peptide chains or 5 cross-sections,

we have evaluated the fibril addition mechanism using four different monomer ensembles with the CG model. Our free energy of monomer addition is found to be in excellent agreement with experiment, and the 2-step "dock and lock" is the main mechanism for fibril elongation via monomer addition, irrespective of the monomer configuration ensemble. Kinetically, we find that the docking step is anywhere between 4X and 28X faster than the locking step depending on monomer ensemble, with the activated monomer exhibiting much faster growth on the C-terminus of the fibril compared to all other monomer conformational ensembles.

5.2 METHODS

Details of the Coarse-Grained Model. Each amino acid of the CG protein model is represented by a single bead corresponding to C_α carbon atom position. The C_α - C_α distances are scaled by the characteristic length scale σ_o, which is set to a fixed value of 3.808 Å, corresponding to the average distance between two adjacent C_α atoms when the intervening peptide bond is in the trans state. The energy function used in the CG model is

$$E_{total} = \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} \begin{bmatrix} A[1 + cos(\phi + \phi_0)] + B[1 - cos(\phi + \phi_0)] \\ +C[1 + cos\ 3(\phi + \phi_0)] + D\left[1 + cos\left(\phi + \phi_0 + \frac{\pi}{4}\right)\right] \end{bmatrix} + \sum_{\substack{i,j \geq 3\ for\ intrachain \\ i,j\ for\ interchain}} 4\varepsilon_H S_1 \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - S_2 \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] + \sum_{Hbonds} U_{HB} \qquad (3)$$

where the first term represents the angle-bending potential, modeled as a stiff harmonic potential with a spring constant $k_\theta = 20\ \varepsilon_H$ /rad^2 and θ is the bond angle defined by three consecutive C_α beads. The optimal bond angle θ_o is set to 95° if bead i is involved in helical secondary structure, 120° if in beta sheets and 105° otherwise. The second term is the dihedral angle potential, where A, B, C, D and ϕ_0 are constants whose values are adjusted to get the desired minima and energy barrier for different dihedral types (Table S5.1 and Figure S5.1) formed by beads i-1, i, i+1 and i+2. There are six different types of dihedrals – helical (H), extended (E), floppy turn (T), +/- 90° turns (P/Q) and 0° dihedral turns (U).

The third term describes the potential involving non-local interactions (Figure S5.2), where r_{ij} is the distance between beads i and j, and σ is set to 1.16 σ_o to account for the excluded volume of the side chains. There are four different bead flavors – strong attraction (B), weak attraction (V), weak repulsion (N) and strong repulsion (L), and the values of S correspond to different bead–bead interactions. $S_2 = -1$ for the attractive interaction B-B, B-V and V-V, while $S_1 = 1.4$, 0.7 and 0.35 respectively. $S_1 = 1/3$ and $S_2 = -1$ for repulsive interactions L-L, L-V and L-B; and $S_1 = 0.35$ and $S_2 = 0$ for all repulsive N-X interactions.

The last term U_{HB} is a distance and orientation dependent short-ranged function for hydrogen bonding interactions. The energy U_{HB} between beads i and j is given by:

$$U_{HB} = -\varepsilon_{HB}\, F(r_{ij} - r_{HB})\, G(|t_{HB,i}.\hat{r}_{ij}| - 1)\, G(|t_{HB,j}.\hat{r}_{ij}| - 1) \qquad (4a)$$

$$F(r_{ij} - r_{HB}) = exp\left[-\frac{(r_{ij} - r_{HB})^2}{\sigma^2_{HBdist}}\right] \qquad (4b)$$

$$G\left(\left|t_{HB,i}.\hat{r}_{ij}\right|-1\right)$$

where F is a Gaussian distance dependent term centered at the ideal hydrogen bond distance r_{HB}, and $r_{HB} = 1.35\ \sigma_o$ for helices and $1.25\ \sigma_o$ for all other secondary structure types. The directional dependent terms, G, have an exponential form, whose arguments $t_{HB,i}$ and $t_{HB,j}$ are unit normal vectors to the planes described by beads i-1, i, i+1 and beads j-1,j,j+1 respectively. The width of the functions F and G are both set to $\sigma_{HBdist} = \sigma_{HB} = 0.5$. The energy scale is set by the hydrophobic strength ε_H.

The hydrogen bonding capabilities of a given amino acid are classified as follows: helical (A), extended β-strand that can form both intramolecular and intermolecular hydrogen-bonds (B), both helical/extended (X), extended β-strand that can form only intermolecular hydrogen-bonds (Y), and none (C). In this work we have extended the intramolecular hydrogen-bond interactions to now describe intermolecular hydrogen-bonding through the new classification Y. For beads assigned as helix, the U_{HB} term is evaluated between i and i+3 bead pairs, and for beads assigned as beta sheets (B/Y) or both (X), the term is evaluated for every bead pair within a cutoff distance of $3.0\ \sigma_o$.

AD Fibril Model and the different Aβ40 Monomer Ensembles. The starting structure of the CG Aβ40 AD fibril was modeled on the 3-stack fibril structure determined by Lu et al. (PDB ID 2M4J [40]). Each amino acid of Aβ40 is represented by a single bead corresponding to it's C_α carbon atom positions, and Table 5.1 provides the sequence mapping for the Aβ40 peptide, and additional information about the secondary structure propensities and hydrogen bonding capabilities of each residue. Details about mapping the amino acids to CG beads are provided in the SI.

In order to study fibril elongation by monomer addition at the fibril ends, we have adopted 4 different coarse-grained monomer ensembles. Details about the force field used for the monomer ensemble can be found in Table 5.2, but they are summarized here.

(1) For the first monomer ensemble, referred to as the *Fibril* ensemble, we have used individual monomers derived from the middle cross-sections of protofibrils equilibrated with the coarse-grain potentials. The monomer potential is then exactly the same as the fibril potential (Table 5.1). However, in the absence of other "U-shaped" adjacent monomers with which to form intermolecular interactions, the monomers can instead form intra-molecular β-strand when they are equilibrated alone in the box.

(2) The second ensemble, the *Random* ensemble, has monomers whose secondary structure for all the beads is designated as a "floppy turn T", but with the ability to form intramolecular or intermolecular hydrogen bonds with any other bead.

(3) The third ensemble, the *IDP* ensemble, is a CG ensemble based on the all-atom simulations by Ball and co-workers, which was shown to form a diverse set of partial secondary structures [46]. The all-atom *IDP* ensemble was generated from a Replica Exchange molecular dynamics simulations [46], and then using the ENSEMBLE conformer selection program[47] to select the best set of structures that conform to experimental NMR data, including chemical shifts and J-couplings [48]. While the bead flavor for each residue is (by definition) the same as the fibril monomers, another new advance of the CG model methodology is that each individual monomer now has distinct tertiary conformations, and thus different bead secondary structure propensities and hydrogen bonding abilities. In this study the individual monomer potentials are determined using the DSSP criterion [49].

(4) Finally, the *Activated* monomer ensemble was restrained to conform to the same shape found in mature AD fibrils, such that they are always in the "activated" form. The barriers

for the intramolecular β-strand potential have been increased (Y dihedral), so that the monomers have higher probability of being in the strand-loop-strand fibrillar form, and they are less likely to form intra-molecular hydrogen bonds.

Table 5.1: Sequence mapping from all-atom AD fibril structure to CG fibril model.

Amino acid sequence	DAEFRHDSGY EVHHQKLVFF AEDVGSNKGA IIGLMVGGVV
CG bead sequence	LVLBLNBBBB LVNNLNBVBB VLLVBBBNNB BBNBBVNNBB
Secondary Structure	TTTTTTTTE EEEETEEEEEE EETTTTTEE EEEEEEEE
Hydrogen bonding	CCCCCCCBB BBBBBBBBBB BBBCCCCCBB BBBBBBBBB

Table 5.2: Monomer potentials for different monomer ensembles used in fibril addition studies

Fibril ensemble	Bead flavor	same as protofibril model
	Secondary Structure	same as protofibril model
	Hydrogen bonding	same as protofibril model
Random ensemble	Bead flavor	same as protofibril model
	Secondary Structure	TTTTTTTTTT TTTTTTTTTT TTTTTTTTT TTTTTTTT
	Hydrogen bonding	XXXXXXXXXX XXXXXXXXXX XXXXXXXXXX XXXXXXXX
IDP ensemble	Bead flavor	same as protofibril model
	Secondary Structure	based on the secondary structure propensity of each structure
	Hydrogen bonding	based on the secondary structure propensity of each structure
Activated ensemble	Bead flavor	LVLBLNBBBB LVNNLNBVBB VLLVBBBNNB BBNBBVNNBB
	Secondary Structure	TTTTTTTYYY YYYYYYYYYY YPTYYTQYYY YYYYYYY
	Hydrogen bonding	CCCCCCCXXX XXXXXXXXXX XCCCCCCXXX XXXXXXXX

Simulation protocol for the CG model. Constant temperature Langevin dynamics was used to propagate the simulation. The simulations are done in reduced units with mass m, energy ε_{HB}, length σ_o and k_B being set to unity. The bond lengths between adjacent beads are held constant using RATTLE algorithm [50]. Reduced time and temperature are given by $\tau = (m\sigma_o/\varepsilon_H)^{1/2}$ and $T^* = \varepsilon_H / k_B$. We simulated the systems at reduced temperature T^*=0.45 (~ 337 K) with a timestep of 0.005τ and friction parameter ζ=0.05. The simulation box size is 150 reduced units in all dimensions.

Protofibril Stability Metric. For the fibril stability studies, a long AD amyloid fibril consisting of 24 cross-sections (72 monomer chains total) was first equilibrated for 1.3M time steps in the *NVT* ensemble at the reduced temperature $T^* = 0.45$ ($T \approx 337$K). Models for different fibril sizes from 6 monomer chains (2 cross-sections) to 30 monomer chains (10 cross-sections) were made from the inner-most center of the equilibrated long AD fibril. We performed 24 independent equilibration runs for the longer fibrils, followed by 3 simulations for each of the smaller protofibrils obtained from the larger fibrils in the *NVT* ensemble at the same T^*. We thus collected production statistics of 120 independent runs for each fibril size, with the simulation time for each run being 65000τ or 13M time steps, corresponding to roughly 13 ms for each simulation.

Protofibril structural stability is measured from the order parameter, χ.

$$\chi = \frac{1}{M}\sum_{\alpha=1}^{N_C}\sum_{\beta>\alpha}^{N_C}\sum_{i}^{N}\sum_{j}^{N} h\left(\varepsilon - \left|r_{\alpha,i;\beta,j} - r^{0}_{\alpha,i;\beta,j}\right|\right) \quad (5)$$

where the outer two sums are over all chains α and β of the outermost two cross-sections of the two ends of the protofibrils, and the inner sums are over bead i of chain α and bead j of chain β, and M is the total number of pair distances. The scalar $r_{\alpha,i;\,\beta,j}$ corresponds to the distance between bead i on chain α and bead j on chain β of a given protofibrils configuration, while $r^{0}_{\alpha,i;\,\beta,j}$ defines the same distance but for the reference AD protofibril. The χ value gives an estimate of the structural deviations from the ideal AD protofibril structure, and since it is based on the Heaviside function, *h*, it ranges from 0 to 1, with a value of 1 indicating perfect structural agreement with the ideal AD reference within allowed thermal fluctuations, *ε* (set to 0.6 distance units or 2.3 Å). Since the order parameter only takes into account the outermost two cross-sections, it can be used for direct comparison between AD protofibrils containing any larger number of cross-sections.

We use two forms of the χ parameter defined in Eq. (5), by restricted sums over beads and chains in Eq. (5). The first metric P_f measures the local structural similarity of the individual protofilaments to the AD reference protofilament, by looking at chain pairs α and β belonging to only one protofilament. It thus sums over 3 independent chain pairs at each end, and only considers beads i and j for residues 17-35 residues which includes the two β strands and the connecting turn region. The global χ_f metric measures how the 3 protofilaments are oriented with respect to one another and with the fibril axis. In this case the chains α and β correspond to the six monomer chains on each end, and the sum over beads i and j are restricted to the β strand regions 17-21 and 31-35. In both cases, statistics were collected about after every 50 τ (10,000 steps).

Free Energy Protocol for the CG model. Based on the fibril ensembles generated for the AD protofibrils of different numbers of cross-sections, we can calculate equilibrium populations of structurally stable and unstable protofibrils based on population differences measured by P_f and/or χ_f. We have determined that there are two well-defined and reasonably well-separated populations, one of which is defined by P_f>0.7 and χ_f>0.7 which measures a population, C_n, of n-ordered monomers in a protofibril with intact end monomers and a well-defined fibril axis. This population is in equilibrium with the remaining fraction of trajectories corresponding to a protofibril with loss of structural order corresponding to P_f<0.7 or χ_f<0.7, and thus measures the population C_{n-1}.

Based on thermodynamic arguments advanced by Ferrone[51] for nucleation-polymerization reactions, we can estimate the change in free energy, ΔG, per unit cross-section (consisting of 3 peptides), *n*, as

$$\frac{d\Delta G}{dn} = k_B T\, ln\left(\frac{[C_{n-1}]}{[C_n]}\right) \quad (6)$$

where k_B is the Boltzmann constant and T is the temperature. When Eq. (6) is integrated over all oligomer sizes, we can generate a free energy curve based on C_n and C_{n-1} populations measured in our model, to determine the critical nucleus size and size regime in which the fibril is thermodynamically stable.

All-atom simulation protocol. The fibrils for the structural stability analysis in the all-atom MD simulations consisted of 3-15 Aβ40 peptide chains, based on the 2M4J AD amyloid

fibril [40]. All amino acid termini are capped using an uncharged amino group at the N-terminus and an uncharged carboxyl group at the C-terminus. In our model, residues K16 and K28 are positively charged and residues E11, E22 and D23 are negatively charged based on neutral pH, physiological salt concentration and the pKa values of the amino acids. This results in one excess negative charge for each peptide, and the system is neutralized with Na+, leading to a system size of up to 224,000 atoms. The force field parameters for the peptides are taken from Amber99sb-star-ildn [52], TIP4P-Ew is used for the water molecules [53]. The simulation box has the size of 120 Å in all directions.

The molecular dynamics simulations at fixed particle number *N,* pressure *P* and temperature *T* are performed using the Gromacs simulation package, version 4.6.1[54]. Periodic boundary conditions are applied and the particle-mesh Ewald method is used for the periodic treatment of Coulomb interactions. The bonds to hydrogen atoms are constrained by the LINCS algorithm [55] and a 2 fs time step is used. To equilibrate the system, we first perform an energy minimization with the steepest descent algorithm. We employ 200 ps first in the NVT and then in the NPT ensembles as equilibration using the Berendsen scheme. For the production run, we perform 100 ns simulations employing Nosé-Hoover temperature coupling with a time constant of $\tau_T = 0.5 s^{-1}$ and isotropic Parrinello-Rahman pressure coupling with a time constant of $\tau_p = 5\ s^{-1}$. Conformational transitions are quantified using the same two order parameters used for the CG model.

Molecular Graphics: The molecular graphics were created using the Visual Molecular Dynamics (VMD) software package [56] *http://www.ks.uiuc.edu/Research/vmd/*

5.3 RESULTS

Protofibril stability and critical nucleus of the AD prtotofibril. We examined AD protofibrils of different numbers of cross-sections ranging from 2 to 10 (i.e. 6-30 monomer chains) to determine the critical nucleus and the size regime where the AD fibril becomes stable. In each case the fibril order was tracked as a function of time using the fibril structural similarity parameters χ_f and P_f as described in the Methods section. Figure 5.2 shows the average of the two metrics, $<\chi_f>$ and $<P_f>$ taken over the last 5000τ of simulation for the two fibril ends. Both $<\chi_f>$ and $<P_f>$ show an increasing trend of stability with increasing number of cross-sections in the protofibrils, which flattens out beyond 5 cross-sections. As expected, $<P_f>$ has a higher value than the $<\chi_f>$ metric, indicating that there is less disorder at the individual protofilament level as compared to the quaternary arrangement of the protofilaments in the protofibril.

A characteristic of the amyloid fibril is interdigitation [42] of the N-terminal and C-terminal β-strands of two distinct chains to form side-chain hydrophobic contacts, known as the "stagger" in amyloid fibril terminology. For the agitated polymorph of the 2-stack fibril, the side-chains of the N-termini of chain *i* interacts with the C termini of monomer *i*+2 to yield a stagger value of +2. We have shown in previous work that the fibril stagger results in asymmetry in the fibril ends[26], such that one end has the C-termini exposed while and the other end exposes the N-terminus. For the AD fibrils, the side-chains of the N-termini of chain *i* interacts with the C termini of monomer *i*+1, i.e. a stagger value of +1.

Our analysis shows that the smaller protofibrils have similar fibril similarity parameter values at each of the ends, until 4 cross-sections are formed, after which the $<\chi_f>$ and $<P_f>$ values start showing differences between the two protofibril ends when the fibril stagger becomes more prominent. Furthermore, the most common signature of instability at the fibril ends in our simulations is when the end monomers form intra-molecular β-hairpins instead of inter-molecular β-strands (Figure 5.3). Both explicit [10] and implicit water simulations [20] have

confirmed that one of the possible intermediates during fibril elongation is the formation of such β-hairpins. Our CG simulations thus suggest similar intermediates for fibril dissociation at the tips.

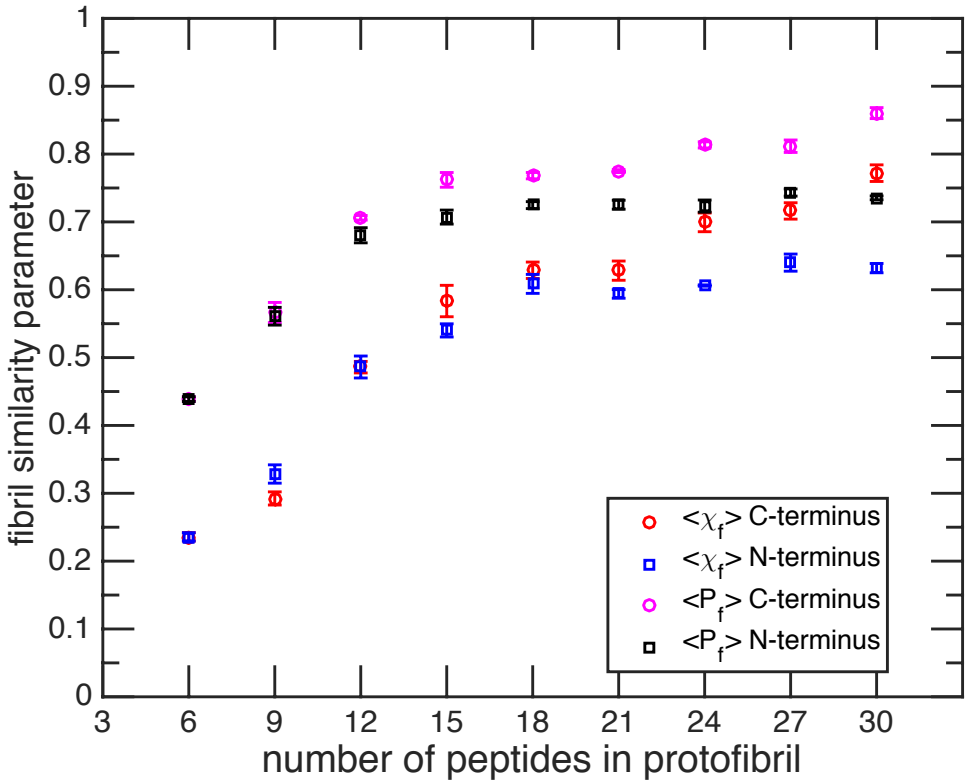


Figure 5.2: *Fibril stability parameters $<\chi_f>$ and $<P_f>$ for equilibrated protofibrils consisting of 6-30 peptide chains.* The parameters give a measure of how much the protofibril structure has deviated from the ideal fibril, with a value of 1 representing perfect fibril structure. Statistics are based on 3 different sets of simulations for each initial protofibril structure.

Figure 5.4 plots the free energy as a function of the number of cross-sections in the protofibrils based on integration of Eq. (6). The nature of the free energy profile corresponds to what one would expect for nucleated growth polymerization, with the critical nucleus being the most unstable species. In our CG model the critical nucleus size is ~12 peptide chains or 4 cross-sections, in which the protofibril has well-formed intermolecular β sheets, but either exhibits point chain defects (like the β-hairpin) or lacks the specific quaternary interactions present in the mature fibrils such as a well-defined stagger. Below the critical nucleus size, the equilibrium shifts towards free monomers while beyond the critical nucleus size the protofibril is dominated by a population that is able to maintain the AD fibril structure, and subsequent addition of cross-sections increases the thermodynamics stability of the fibril species. It should be noted that we are not on an absolute scale for free energy, and therefore impose the definition of zero at the minimum protofibril size.

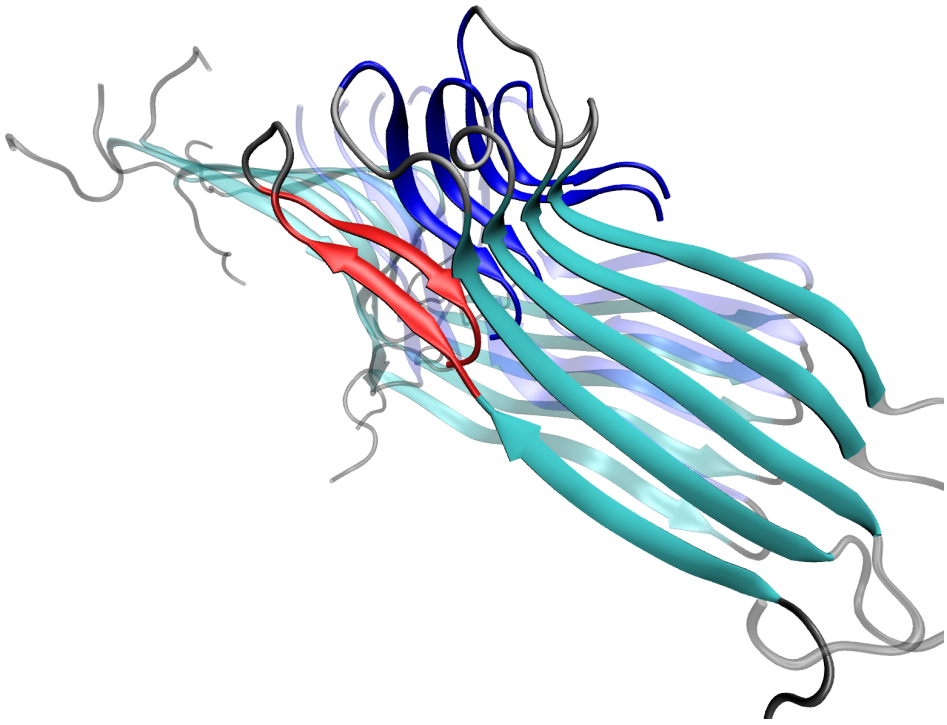


Figure 5.3: *Snapshot of an unstable protofibril containing 15 peptide chains*, where one of the monomers in the outermost cross-section has formed an intra-molecular β -strand (red strand) with the C-terminal residues (blue region), by breaking a part of the N-terminal inter-molecular β -sheet formed by residues 10-22.

In order to validate the free energy results obtained from our coarse-grain model, we performed 100 ns simulations of 2 to 5 cross-sections (6-15 peptide chains) to see whether we obtain similar structural stability trends using all-atom simulations with explicit water. We used the same global χ_f and local P_f metrics (Eq. (5)) calculated for the N-terminus and C-terminus, with ε set to 2.3 Å, which is equivalent to the CG value after adjustments from reduced units. Figure 5.5 shows $<\chi_f>$ and $<P_f>$ metrics averaged over the last 30 ns, and Figures S5.3 and S5.4 show the full 100ns time course using these same metrics. Like the CG model, individual protofilaments are better ordered with respect to the overall protofibril metric, and in general the C-terminus has higher values than the N-terminus. The global χ metric shows an increasing trend to stability, and reaches an abrupt convergence at 15 peptide chains. The large variation in the order parameter at 12 chains is consistent with the CG model, the size at which the protofibril corresponds to the most unfavorable free energy in the thermodynamics of fibril stability. Thus the all-atom model strongly supports the size at which we can investigate the thermodynamics and kinetic mechanisms of fibril addition in what follows.

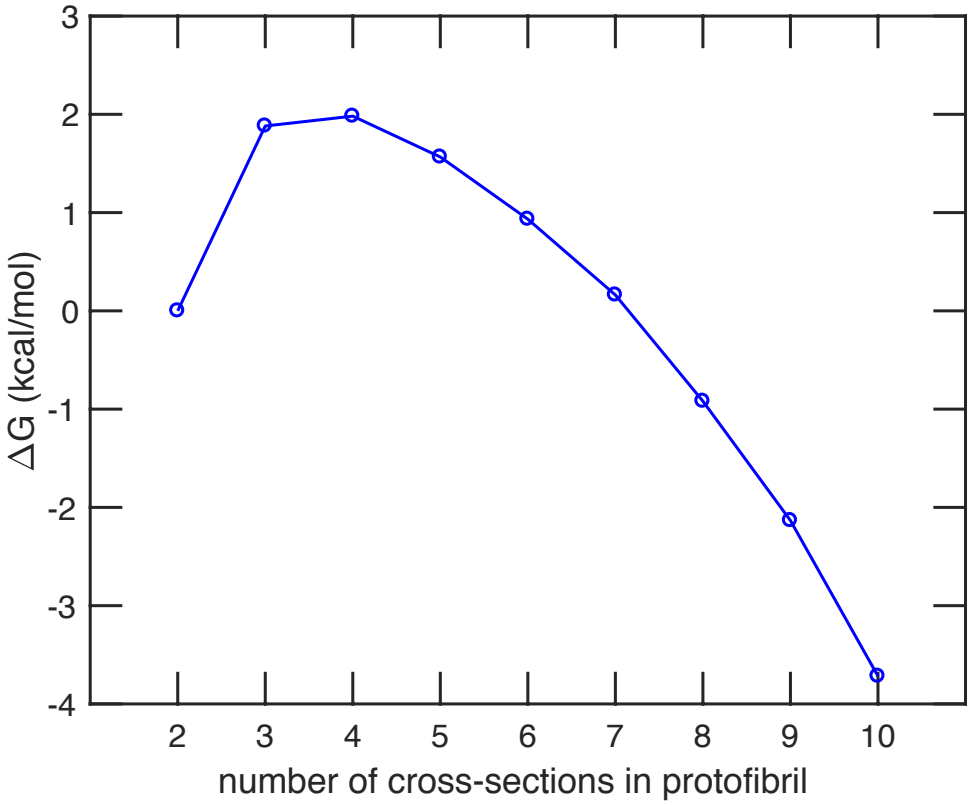


Figure 5.4: *Free energy profile for the protofibril and free monomer equilibrium.* The critical nucleus size is about 12 peptide chains with the constant negative slope beyond the critical nucleus size indicating favorable region of fibril growth via monomer addition.

Thermodynamics of protofibril Elongation. In order to evaluate the thermodynamics of fibril elongation for the CG model, and its dependence on the monomer ensemble, we picked the stable form of the model protofibril containing 5 cross-sections and added it to the simulation box populated randomly with 10 monomers drawn from each of the four monomer ensembles. The peptides were allowed to diffuse for 65M time steps (estimated to be 65 milleseconds) at T^*=0.45 during which we collected statistics on the probability function $p(r_C)$, the probability of finding the center of mass of a "free" monomer at a distance r_C from the mid-point of the outermost fibril cross-section. The mid-point of fibril cross-section is defined as the centroid of the triangle formed by bead 33 of the three peptides chains forming the cross-section. The probability $p(r_C)$ was normalized by the probability of finding a monomer in the bulk. We considered only the monomers that add on to the fibril tips, and ignore monomers that take part in lateral addition to the fibril surface, by constructing a half-spherical shell with the mid-point of the outermost fibril cross-section as the origin in $p(r_C)$ calculation. This definition is consistent with the experiment in reference 57, which monitored the increase in Thioflavin T fluorescence resulting from binding of the fluorescent dye to β-sheets present in amyloid structure; thus, the experimental technique reports only on monomer binding at the fibril tips, and not on the lateral edges. The other half-sphere, which we have ignored in our calculations, contains the middle cross-sections of the fibril where monomers are more likely to take part in lateral addition; our calculations showed that only 1-2% of the monomers ended up in the ignored half-spherical region, and thus the numbers we have presented in this paper for binding affinities won't be significantly influenced by the choice of our shell.

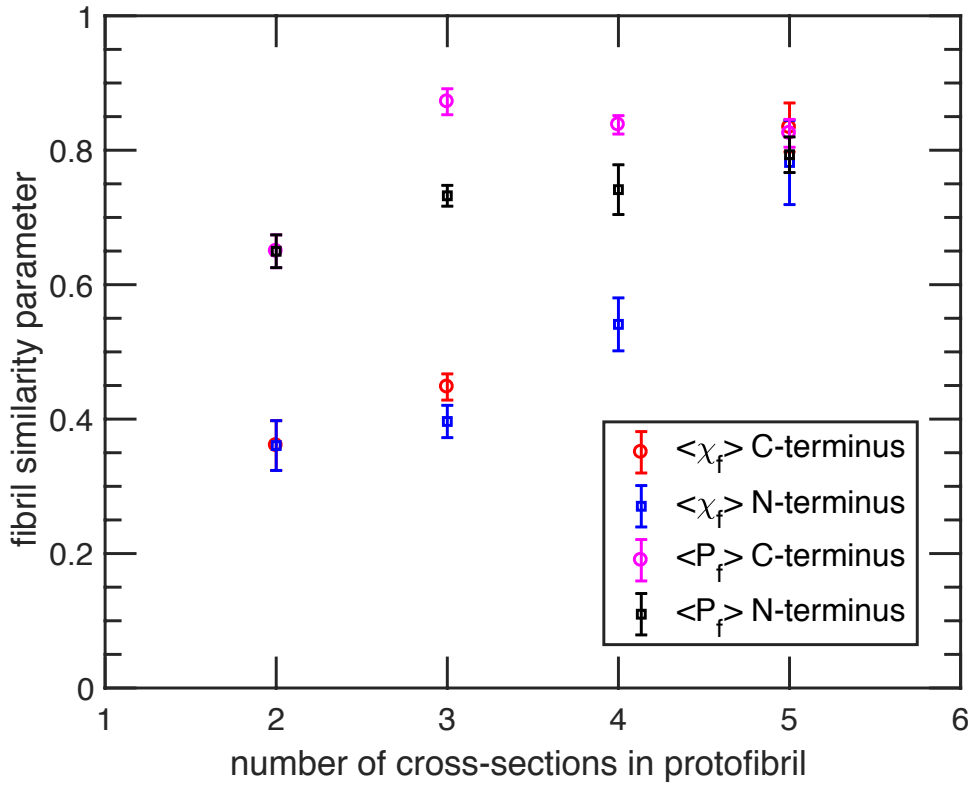


Figure 5.5: *Fibril stability parameters <χ_f> and <P_f> for equilibrated protofibrils using an all-atom simulation.* Statistics are based on the last 30 ns of 100 ns simulations for 2-5 cross-sections.

One dimensional free energy profile (so called potential of mean force, PMF) sampled along the radial distance r from the fibril tip, is then calculated as

$$F(r_C) = -k_B T \, ln[p(r_C)] \quad (7)$$

Standard deviations are based on five different sets of runs, with each set of run consisting of 300 separate simulations. Figure 5.6 shows the reversible work needed for a monomer to move from the fibril surface to a separation distance r_C between the fibril end and the monomers sampled from the *Fibril* ensemble (Fig. 6a) and from the *Activated* ensemble (Fig. 6b); we include the same PMF plots for the *Random* and *IDP* monomer ensembles in the SI material. In all four cases the PMF plot has a single minimum and shows no difference at the two fibril ends. The one-dimensional PMF, based solely on the distance of separation between the monomer and the fibril surface, largely captures the free energy of the "docking" phase.

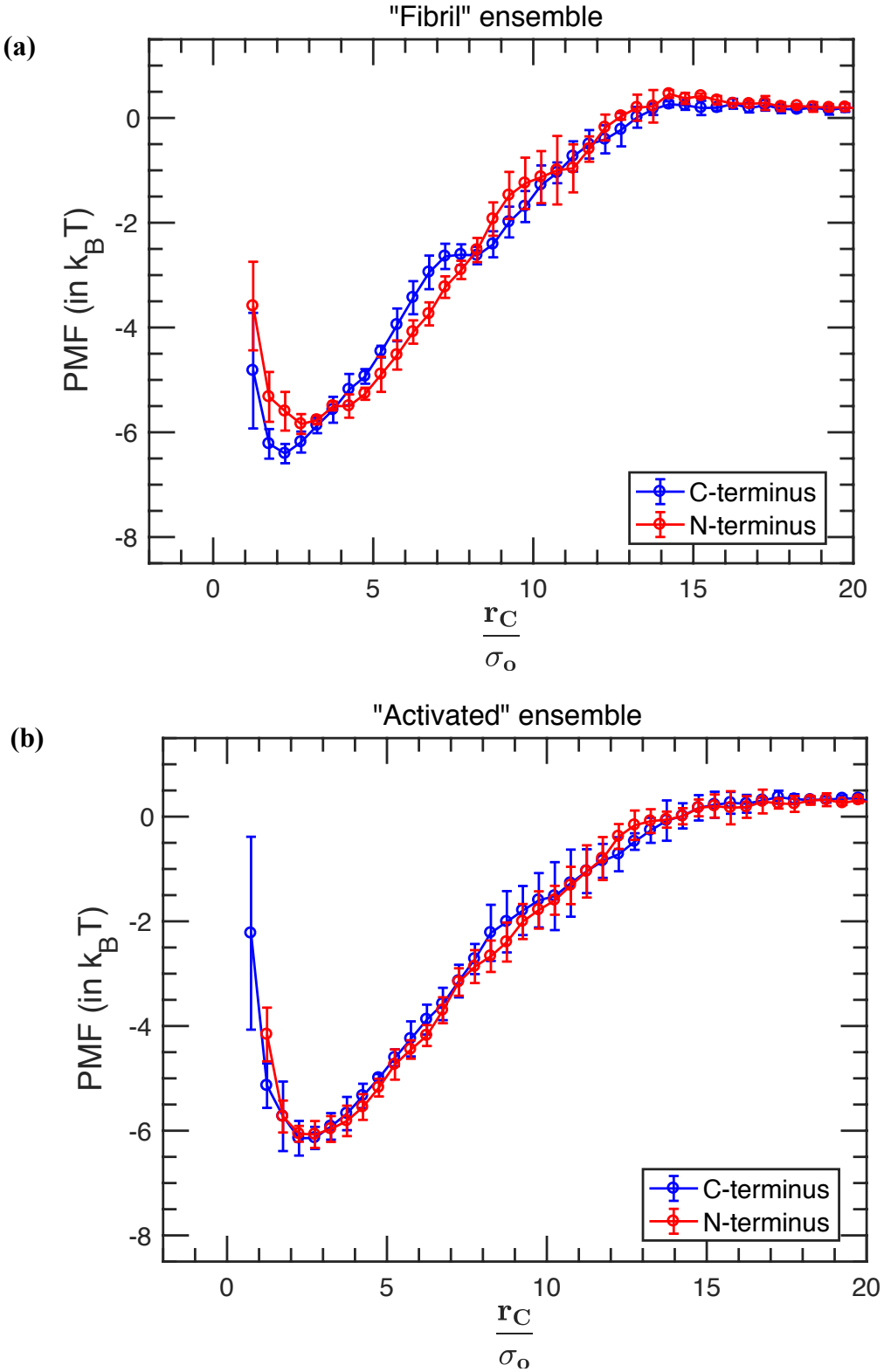


Figure 5.6: *The potential of mean force profile for monomer addition at the two protofibril tips as a function of separation distance r_C (a) from monomer configurations drawn from the Fibril ensemble, and (b) Activated ensemble.* The free energy profile is very similar for the all four monomer ensembles, with a single minimum.

We next proceed to estimating the free energy of binding (ΔG_{Bind}) derived by Schwierz et al. [10] using Eq. (8) .

$$k_{EQ} = L_x L_y \int_{r_B}^{r_S} dr_C \frac{P(r_C)}{P(r_S)} \tag{8a}$$

$$\Delta G_{Bind} = k_B T \, ln\left[c^0 k_{EQ}\right] \tag{8b}$$

where k_{EQ} is the equilibrium constant for peptide binding, L_x and L_y are the dimensions of the simulation box, r_S is the smallest separation distance where $F(r_C) \approx 0$, r_B is the separation distance corresponding to minimum value of $F(r_C)$, $P(r_C)$ is the simulated probability density distribution, and c^0 is standard state concentration (1 mol/L). Table 5.3 provides the calculated free energy of binding for the different monomer ensembles using the CG model, which varies from 13.3-14.3 k_BT. These values are in very good agreement with experimental binding affinities of ~15.2 k_BT [57], and suggests that the lack of explicit interactions due to water and side chain degrees of freedom are not strongly effecting the results, in which all atom simulations determined binding free energies of ~8-10 k_BT [10].

Since all four ensembles have similar PMF values and free energy of binding within thermal fluctuations, the "docking" phase is found to not be sensitive to the monomer conformational ensemble. The optimal r_C for the C-terminus is slightly lower than the N-terminus by ~ 2.3 Å for all the ensembles except the *Activated* ensemble. Since the C-terminal tip has more hydrophobic residues, which manifests as a higher density of attraction in our model, the approaching monomer has more non-specific surface area to explore when "docking" and is more collapsed when it docks. On the other fibril end, the N-terminus has more sequence specificity and patterning of both attractive and repulsive beads, hence the monomer requirements for finding optimal interactions is more entropically unfavorable. Because the monomers of the *Activated* ensemble are very rigid with a fixed shape, these monomers don't collapse upon docking on the C-terminus, and thus have similar separation distance r_C at the two fibril ends.

Table 5.3: Free energy of binding for fibril elongation at the two fibril tips by monomer addition.

	Monomer Ensemble (in units of k_BT)			
	Fibril	*Random*	*IDP*	*Activated*
C-terminus	13.7 ± 0.74	14.3 ± 0.33	14.2 ± 0.18	13.6 ± 0.22
N-terminus	13.3 ± 0.30	14.2 ± 0.31	13.6 ± 0.23	13.5± 0.13

Kinetics of Fibril Elongation. We next performed simulations to evaluate the mean first passage time (MFPT) for fibril elongation by monomer addition. The simulation conditions comprised an equilibrated stable protofibril with 15 peptide chains as before, but now with two monomers placed randomly on either side of the fibril such that the separation between bead 20 of the monomer and the mid-point of the outermost fibril cross-section was 5 reduced distance units (~19 Å). This initial condition minimized any lag time resulting from the time to encounter the fibril tip by diffusion, and the large simulation box size ensured that the two monomers at the two opposite ends of the protofibril did not interact with each other. The upper bound for the

MFPT was 65M time steps (~65 ms) at T^*=0.45, and again the 1500 independent trajectories were divided into five sets to calculate standard deviations.

It is difficult to observe a complete "dock-lock" or "activated" monomer addition mechanism using even CG computer simulations owing to the slow time scale in the range of minutes[19]. To overcome this problem and at the same time get some meaningful information about the mechanisms, we looked at the MFPT required to form the first 5 non-native contacts as an estimate of the docking time, and 5 native contacts as a lower bound to the locking time scale. Since the "docking" step is a reversible process, with monomers forming and breaking contacts, we found that the probability of monomer dissociation from the fibril tip was more than 50% when fewer than 5 contacts were formed between the free monomer and fibril tip, hence we selected a lower bound of 5 non-native contacts. We choose the same number for the native contacts in the "locking" step to be consistent with our definition of "docking" step, where in this case the first 5 native contacts are restricted to residues 8-40, which take part in inter-molecular β-sheets and also in hydrophobic contacts between different protofilament layers.

Figure 5.7a shows the MFPT needed to form the first five non-native in the docking step and the first five native contacts in the locking step between the free monomer and the two fibril ends. The time scales for the formation of non-native contacts at the two fibril tips are very similar for all monomer ensembles except the *Activated* monomer ensemble, which shows addition to the C-terminus that is 3-4 times faster than addition to the N-terminus. Figure 5.7b shows the MFPT needed to form the first 10 non-native in the docking step and the first 10 native contacts in the locking step between the free monomer and the two fibril ends, confirming that the results are not changed by the choice of the number of native/non-native contacts. Figure S5.6 and S5.7 shows representative snapshots of fibril structures during elongation process, with monomers forming five native and non-native contacts.

Furthermore there are distinctly different time scales associated with the "docking" phase, which is anywhere between 4 and 30 times faster than the start of the "locking" phase, depending on monomer ensemble, with the smallest difference observed for *Activated* monomer addition to the C-terminus of the fibril (Table 5.4). These time scales are distinct enough to state that even the *Activated* monomer first "docks" onto the fibril surface at nonspecific locations and has to search on the fibril surface to form the "correct" specific contacts. For the "activated monomer" mechanism, one would expect more commensurate time scales for the formation of the first five non-native and native contacts, something that we do not strictly observe in our simulations. Even so, the start of the locking step is 2 times faster for the *Activated* monomer ensemble compared to that for other monomer ensembles, as it does not require any internal reconfiguration to properly "lock", especially on the C-terminus end of the fibril.

Table 5.4: Ratio of locking to docking time scales for monomer addition at the two fibril ends.

	Monomer Ensemble τ_{lock}/τ_{dock}			
	Fibril	*Random*	*IDP*	*Activated*
C-terminus	11.8 ± 5.0	9.4 ± 4.0	6.5 ± 2.4	4.1± 1.0
N-terminus	29.3 ± 28.2	8.9 ± 3.2	13.7± 7.7	15.2 ± 15.1

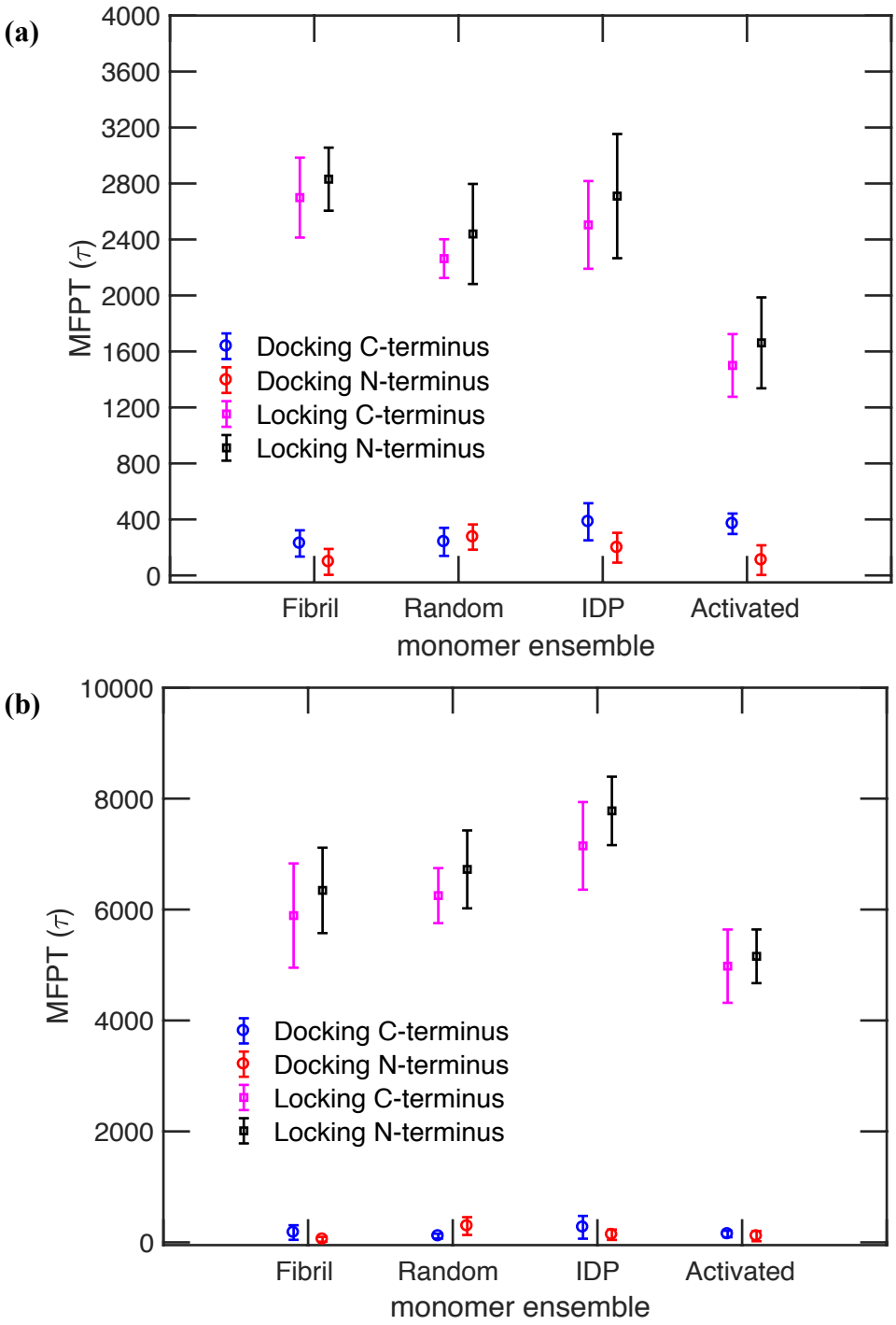


Figure 5.7: *MFPT required to form the first non-native (docking) and the first native contacts (locking) for different monomer ensembles*. All the four monomer ensembles have two different time-scales for forming the first non-native and native contacts, indicating that the "lock-dock" mechanism is the more common mechanism for fibril elongation, irrespective of the monomer structure. Nonetheless the activated form has the smallest separation of dock and lock timescales. (a) formation of 5 non-native and 5 native contacts. (b) formation of 10 non-native and 10 native contacts, which shows that our conclusions are not dependent on number of contacts.

5.4 DISCUSSION AND CONCLUSIONS

A detailed understanding of the mechanism underlying nucleation and growth of amyloid fibrils is fundamental for determining the possible disease origins of neurodegenerative conditions, such as Alzheimer's disease. However, because of the slow millisecond to second experimental time scale involved in amyloid fibril formation, all-atom computer simulations that can provide molecular interpretations are challenged by the restrictions of the computational time scale of microseconds at most. In this case, CG simulations have proven very useful, by integrating out uninteresting fast motions to reach effective time scales of tens of milliseconds in obtaining information on the nucleation and first events of fibril elongation of the disease polymorph of amyloid fibrils composed of the Aβ40 peptide.

Both the CG and all-atom simulations corroborate that the critical nucleus has ~12 Aβ peptide chains, and that the minimal size at which AD protofibrils become stable correspond to 15 Aβ peptide chains arranged in 5 cross-sections. For the fibril elongation process, we tested the two-step dock-lock mechanism, in which the docked peptide undergoes significant conformational rearrangements to fit precisely on the fibril template, and the activated monomer mechanism, in which the peptide is already in the U-shaped conformation observed in a mature fibril and is hypothesized to add onto the fibril in a single step. Our results show that there is no significant thermodynamic difference between the two scenarios and that the kinetics in both cases is always a two-step process involving a docking timescale that is 4-30 times faster than the locking step, depending on which end of the fibril addition takes place, and the type of monomer ensemble.

Formation of at least five native contacts between the fibril and monomer in the locking step indicates that conformational rearrangement of a nonactivated peptide is ~2 times slower than rotational rearrangement of the activated but ridged U-shaped peptide. Although the activated monomer conformation is rare, constituting <1% of the conformational ensembles derived from all-atom simulations that best conform to the NMR data for Aβ monomers in solution[48], the CG model suggest that the kinetics of fibrillization would be faster with an activated form of the monomer.

5.5 REFERENCES

1. Karran, E.; Mercken, M.; De Strooper, B., The Amyloid Cascade Hypothesis for Alzheimer's Disease: An Appraisal for the Development of Therapeutics. *Nature Rev. Drug Discovery* **2011,** *10* (9), 698-712.

2. Glabe, C. G., Structural Classification of Toxic Amyloid Oligomers. *J. Bio. Chem.* **2008,** *283* (44), 29639-29643.

3. Walsh, D. M.; Klyubin, I.; Fadeeva, J. V.; Cullen, W. K.; Anwyl, R.; Wolfe, M. S.; Rowan, M. J.; Selkoe, D. J., Naturally Secreted Oligomers of Amyloid β Protein Potently Inhibit Hippocampal Long-Term Potentiation in Vivo. *Nature* **2002,** *416* (6880), 535-539.

4. Haass, C.; Selkoe, D. J., Soluble Protein Oligomers in Neurodegeneration: Lessons from the Alzheimer's Amyloid β-Peptide. *Nature Rev. Mol. Cell Bio.* **2007,** *8* (2), 101-112.

5. Petkova, a. T.; Leapman, R. D.; Guo, Z. H.; Yau, W. M.; Mattson, M. P.; Tycko, R., Self-Propagating, Molecular-Level Polymorphism in Alzheimer's β-Amyloid Fibrils. *Science* **2005,** *307* (5707), 262-265.

6. Chimon, S.; Shaibat, M. A.; Jones, C. R.; Calero, D. C.; Aizezi, B.; Ishii, Y., Evidence of Fibril-Like β-Sheet Structures in a Neurotoxic Amyloid Intermediate of Alzheimer's β-Amyloid. *Nature Struct. Mol. Bio.* **2007,** *14* (12), 1157-1164.

7. Ahmed, M.; Davis, J.; Aucoin, D.; Sato, T.; Ahuja, S.; Aimoto, S.; Elliott, J. I.; Van Nostrand, W. E.; Smith, S. O., Structural Conversion of Neurotoxic Amyloid-β (1-42) Oligomers to Fibrils. *Nature Struct. Mol. Bio.* **2010,** *17* (5), 561-567.

8. Cohen, S. I.; Linse, S.; Luheshi, L. M.; Hellstrand, E.; White, D. A.; Rajah, L.; Otzen, D. E.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P., Proliferation of Amyloid-β42 Aggregates Occurs through a Secondary Nucleation Mechanism. *Proc. Natl. Acad. Sci.* **2013,** *110* (24), 9758-9763.

9. Pinotsi, D.; Buell, A. K.; Galvagnion, C.; Dobson, C. M.; Kaminski Schierle, G. S.; Kaminski, C. F., Direct Observation of Heterogeneous Amyloid Fibril Growth Kinetics Via Two-Color Super-Resolution Microscopy. *Nano Lett.* **2013,** *14* (1), 339-345.

10. Schwierz, N.; Frost, C. V.; Geissler, P. L.; Zacharias, M., Dynamics of Seeded $A\beta_{40}$-Fibril Growth from Atomistic Molecular Dynamics Simulations: Kinetic Trapping and Reduced Water Mobility in the Locking Step. *J. Am. Chem. Soc.* **2016,** *138* (2), 527-539.

11. Ban, T.; Hoshino, M.; Takahashi, S.; Hamada, D.; Hasegawa, K.; Naiki, H.; Goto, Y., Direct Observation of Aβ Amyloid Fibril Growth and Inhibition. *J. Mol. Biol.* **2004,** *344* (3), 757-767.

12. Pinotsi, D.; Buell, A. K.; Galvagnion, C.; Dobson, C. M.; Kaminski Schierle, G. S.; Kaminski, C. F., Direct Observation of Heterogeneous Amyloid Fibril Growth Kinetics Via Two-Color Super-Resolution Microscopy. *Nanoletters* **2013,** *14* (1), 339-345.

13. Blackley, H.; Sanders, G.; Davies, M.; Roberts, C.; Tendler, S.; Wilkinson, M., In-Situ Atomic Force Microscopy Study of β-Amyloid Fibrillization. *J. Mol. Biol.* **2000,** *298* (5), 833-840.

14. Kellermayer, M. S.; Karsai, Á.; Benke, M.; Soós, K.; Penke, B., Stepwise Dynamics of Epitaxially Growing Single Amyloid Fibrils. *Proc. Natl. Acad. Sci.* **2008,** *105* (1), 141-144.

15. Ferkinghoff-Borg, J.; Fonslet, J.; Andersen, C. B.; Krishna, S.; Pigolotti, S.; Yagi, H.; Goto, Y.; Otzen, D.; Jensen, M. H., Stop-and-Go Kinetics in Amyloid Fibrillation. *Phys. Rev. E* **2010,** *82* (1), 010901.

16. Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D., Monomer Adds to Preformed Structured Oligomers of Aβ-Peptides by a Two-Stage Dock–Lock Mechanism. *Proc. Natl. Acad. Sci.* **2007,** *104* (1), 111-116.

17. Straub, J. E.; Thirumalai, D., Toward a Molecular Theory of Early and Late Events in Monomer to Amyloid Fibril Formation. *Ann. Rev. Phys. Chem.* **2011,** *62*, 437-463.

18. Massi, F.; Straub, J. E., Energy Landscape Theory for Alzheimer's Amyloid β-Peptide Fibril Elongation. *Proteins: Struct., Func., Bioinform.* **2001,** *42* (2), 217-229.

19. Esler, W. P.; Stimson, E. R.; Jennings, J. M.; Vinters, H. V.; Ghilardi, J. R.; Lee, J. P.; Mantyh, P. W.; Maggio, J. E., Alzheimer's Disease Amyloid Propagation by a Template-Dependent Dock-Lock Mechanism. *Biochem.* **2000,** *39* (21), 6288-6295.

20. Gurry, T.; Stultz, C. M., Mechanism of Amyloid-β Fibril Elongation. *Biochem.* **2014,** *53* (44), 6981-6991.

21. Buchete, N-V.; Tycko, R.; Hummer, G., Molecular Dynamics Simulations of Alzheimer's β-Amyloid Protofilaments. *J. Mol. Biol.* **2005,** *353* (4), 804-821.

22. Buchete, N-V.; Hummer, G., Structure and Dynamics of Parallel β-Sheets, Hydrophobic Core, and Loops in Alzheimer's Aβ Fibrils. *Biophys. J.* **2007,** *92* (9), 3032-3039.

23. Horn, A. H.; Sticht, H., Amyloid-β 42 Oligomer Structures from Fibrils: A Systematic Molecular Dynamics Study. *J. Phys. Chem. B* **2010,** *114* (6), 2219-2226.

24. Cannon, M. J.; Williams, A. D.; Wetzel, R.; Myszka, D. G., Kinetic Analysis of β-Amyloid Fibril Elongation. *Anal. Biochem.* **2004,** *328* (1), 67-75.

25. Qiang, W.; Kelley, K.; Tycko, R., Polymorph-Specific Kinetics and Thermodynamics of β-Amyloid Fibril Growth. *J. Am. Chem. Soc.* **2013,** *135* (18), 6860-6871.

26. Fawzi, N. L.; Okabe, Y.; Yap, E.-H.; Head-Gordon, T., Determining the Critical Nucleus and Mechanism of Fibril Elongation of the Alzheimer's $A\beta_{1-40}$ Peptide. *J. Mol. Biol.* **2007,** *365* (2), 535-550.

27. Fawzi, N. L.; Kohlstedt, K. L.; Okabe, Y.; Head-Gordon, T., Protofibril Assemblies of the Arctic, Dutch, and Flemish Mutants of the Alzheimer's $A\beta_{1-40}$ Peptide. *Biophys. J.* **2008,** *94* (6), 2007-2016.

28. Pellarin, R.; Caflisch, A., Interpreting the Aggregation Kinetics of Amyloid Peptides. *J. Mol. Biol.* **2006,** *360* (4), 882-892.

29. Abeln, S.; Vendruscolo, M.; Dobson, C. M.; Frenkel, D., A Simple Lattice Model That Captures Protein Folding, Aggregation and Amyloid Formation. *PloS One* **2014,** *9* (1), e85185.

30. Han, W.; Schulten, K., Fibril Elongation by $A\beta_{17-42}$: Kinetic Network Analysis of Hybrid-Resolution Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2014,** *136* (35), 12450-12460.

31. Sorensen, J. M.; Head-Gordon, T., Toward Minimalist Models of Larger Proteins: A Ubiquitin-Like Protein. *Proteins: Struct., Func., Bioinform.* **2002,** *46* (4), 368-379.

32. Sorenson, J.; Head-Gordon, T., Protein Engineering Study of Protein L by Simulation. *J. Comp. Bio.* **2002,** *9* (1), 35-54.

33. Brown, S.; Fawzi, N. J.; Head-Gordon, T., Coarse-Grained Sequences for Protein Folding and Design. *Proc. Natl. Acad. Sci.* **2003,** *100* (19), 10712-10717.

34. Yap, E. H.; Fawzi, N. L.; Head-Gordon, T., A Coarse-Grained α-Carbon Protein Model with Anisotropic Hydrogen-Bonding. *Proteins: Struct., Func., Bioinform.* **2008,** *70* (3), 626-638.

35. Brown, S.; Head-Gordon, T., Intermediates and the Folding of Proteins L and G. *Prot. Sci.* **2004,** *13* (4), 958-970.

36. Fawzi, N. L.; Chubukov, V.; Clark, L. A.; Brown, S.; Head-Gordon, T., Influence of Denatured and Intermediate States of Folding on Protein Aggregation. *Prot. Sci.* **2005,** *14* (5), 1380-1380.

37. Marianayagam, N.; Fawzi, N.; Head-Gordon, T., Protein Folding by Distributed Computing and the Denatured State Ensemble. *Proc. Natl. Acad. Sci.* **2005,** *102*, 16684-16689.

38. Sorenson, J.; Head-Gordon, T., Matching Simulation and Experiment: A New Simplified Model for Protein Folding. *J. Comp. Bio.* **2000,** *7* (3/4), 469-481.

39. Sorenson, J. M.; Head-Gordon, T., Redesigning the Hydrophobic Core of a Model β-Sheet Protein: Destabilizing Traps through a Threading Approach. *Proteins: Struct., Func., Bioinform.* **1999,** *37* (4), 582-591.

40. Lu, J.-X.; Qiang, W.; Yau, W.-M.; Schwieters, C. D.; Meredith, S. C.; Tycko, R., Molecular Structure of β-Amyloid Fibrils in Alzheimer's Disease Brain Tissue. *Cell* **2013,** *154* (6), 1257-1268.

41. Tycko, R., Amyloid Polymorphism: Structural Basis and Neurobiological Relevance. *Neuron* **2015,** *86* (3), 632-645.

42. Petkova, A. T.; Yau, W.-M.; Tycko, R., Experimental Constraints on Quaternary Structure in Alzheimer's β-Amyloid Fibrils. *Biochemistry* **2006,** *45* (2), 498-512.

43. Paravastu, A. K.; Leapman, R. D.; Yau, W.-M.; Tycko, R., Molecular Structural Basis for Polymorphism in Alzheimer's β-Amyloid Fibrils. *Proc. Natl. Acad. Sci.* **2008,** *105* (47), 18349-18354.

44. Xiao, Y.; Ma, B.; McElheny, D.; Parthasarathy, S.; Long, F.; Hoshi, M.; Nussinov, R.; Ishii, Y., Aβ(1-42) Fibril Structure Illuminates Self-Recognition and Replication of Amyloid in Alzheimer's Disease. *Nature Struct. Mol. Bio.* **2015,** *22* (6), 499-505.

45. Fawzi, N. L.; Yap, E. H.; Okabe, Y.; Kohlstedt, K. L.; Brown, S. P.; Head-Gordon, T., Contrasting Disease and Nondisease Protein Aggregation by Molecular Simulation. *Acc. Chem. Res.* **2008,** *41* (8), 1037-47.

46. Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T., Differences in β-Strand Populations of Monomeric Aβ40 and Aβ42. *Biophys. J.* **2013,** *104* (12), 2714-2724.

47. Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W.-Y.; Forman-Kay, J. D., Characterization of Disordered Proteins with Ensemble. *Bioinformatics* **2013,** *29* (3), 398-399.

48. Ball, K. A.; Wemmer, D. E.; Head-Gordon, T., Comparison of Structure Determination Methods for Intrinsically Disordered Amyloid-β Peptides. *J. Phys. Chem. B* **2014,** *118*, 6405-6416.

49. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983,** *22* (12), 2577-2637.

50. Andersen, H. C., Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *J. Comp. Phys.* **1983,** *52* (1), 24-34.

51. Ferrone, F., Analysis of Protein Aggregation Kinetics. *Meth. Enzym.* **1999,** *309*, 256-274.

52. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved Side-Chain Torsion Potentials for the Amber Ff99sb Protein Force Field. *Proteins: Struct., Func., Bioinform.* **2010,** *78* (8), 1950-1958.

53. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an Improved Four-Site Water Model for Biomolecular Simulations: Tip4p-Ew. *J. Chem. Phys.* **2004,** *120* (20), 9665-9678.

54. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., Gromacs: Fast, Flexible, and Free. *J. Comp. Chem.* **2005,** *26* (16), 1701-1718.

55. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald: An N·Log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.

56. Humphrey, W.; Dalke, A.; Schulten, K., Vmd: Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33-38.

57. O'Nuallain, B.; Shivaprasad, S.; Kheterpal, I.; Wetzel, R., Thermodynamics of Aβ(1-40) Amyloid Fibril Elongation. *Biochemistry* **2005,** *44* (38), 12709-12718.

5.6 APPENDIX

METHODS

Mapping of amino-acid sequence to CG beads

In the first step, the amino acids were mapped to the 4-letter CG beads using a preliminary mapping rule given in Table S5.1. The mapping rule presented in Table S5.1 is based on extensive parameterization of the coarse-grain model based on thermodynamic studies of proteins L and G [1-5]. The residues are assigned strongly attractive to repulsive beads based on their hydrophobicity. However, one also needs to take care of the amino acid environment when designing the CG bead-sequence. In the next step, we scanned through the protein structure and assigned strongly attractive bead flavor B to residues, which were responsible for hydrophobic contacts between the protofilament layers (residues 7-10, 25-27, 30-32, 39-40). This step was necessary to keep the different layers of protofilament together.

REFERENCES:

1. Sorensen, J. M.; Head-Gordon, T., Toward minimalist models of larger proteins: A ubiquitin-like protein. *Proteins: Struct., Func., Bioinform.* **2002,** *V46* (N4), 368-379.
2. Sorenson, J.; Head-Gordon, T., Protein engineering study of protein L by simulation. *J. Comp. Bio.* **2002,** *9* (1), 35-54.
3. Brown, S.; Fawzi, N. J.; Head-Gordon, T., Coarse-grained sequences for protein folding and design. *Proc. Natl. Acad. Sci.* **2003,** *100* (19), 10712-10717.
4. Yap, E. H.; Fawzi, N. L.; Head‐Gordon, T., A coarse‐grained α‐carbon protein model with anisotropic hydrogen‐bonding. *Proteins: Struct. Funct. and Bioinf.* **2008,** *70* (3), 626-638.
5. Brown, S.; Head-Gordon, T., Intermediates and the folding of proteins L and G. *Prot. Sci.* **2004,** *13.*

Table S5.1: Mapping of the amino acids (AA) to coarse-grained beads of four different flavors (B/V/N/L)

AA	CG Bead	AA	CG bead	AA	CG Bead	AA	CG Bead
Trp	B	Met	B	Gly	N	Glu	L
Cys	B	Tyr	B	Ser	N	Asp	L
Leu	B	Val	V	Thr	N	Gln	L
Ile	B	Ala	V	Lys	N	Asn	L
Phe	B	Pro	N	His	N	Arg	L

Table S5.2: Parameters for different dihedral angle propensities

Dihedral Type	A(ε_H)	B(ε_H)	C(ε_H)	D(ε_H)	σ_o (rad)	Local minima (global minima in **bold**)
H (helical)	0	1.2	1.2	1.2	0.17	-65°, **50°**, 165°
E (Extended)	0.9	0	1.2	0	-0.35	**-160°**, -45°, 85°
T (turn)	0.2	0.2	0.2	0	0	**-60°**, **60°**, **180°**
P (+90°)	0.36	0	0.48	0	+1.57	-155°, -25°, **90°**
Q (-90°)	0.36	0	0.48	0	-1.57	**-90°**, 25°, +155°
U (0°)	0.36	0	0.48	0	+3.14	-115°, **0°**, 115°
Y (Extended)	1.8	0	2.4	0	-0.35	**-160°**, -45°, 85°

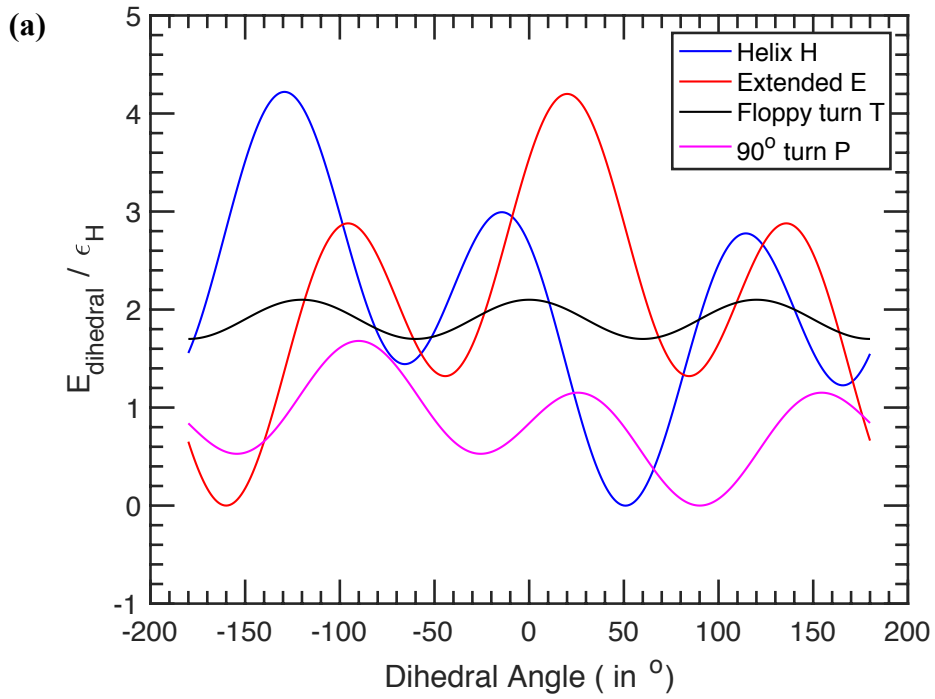


Figure S5.1: Dihedral angle potential as a function of dihedral angle for the dihedral types helix H, extended E, turn T and +90° turn P.

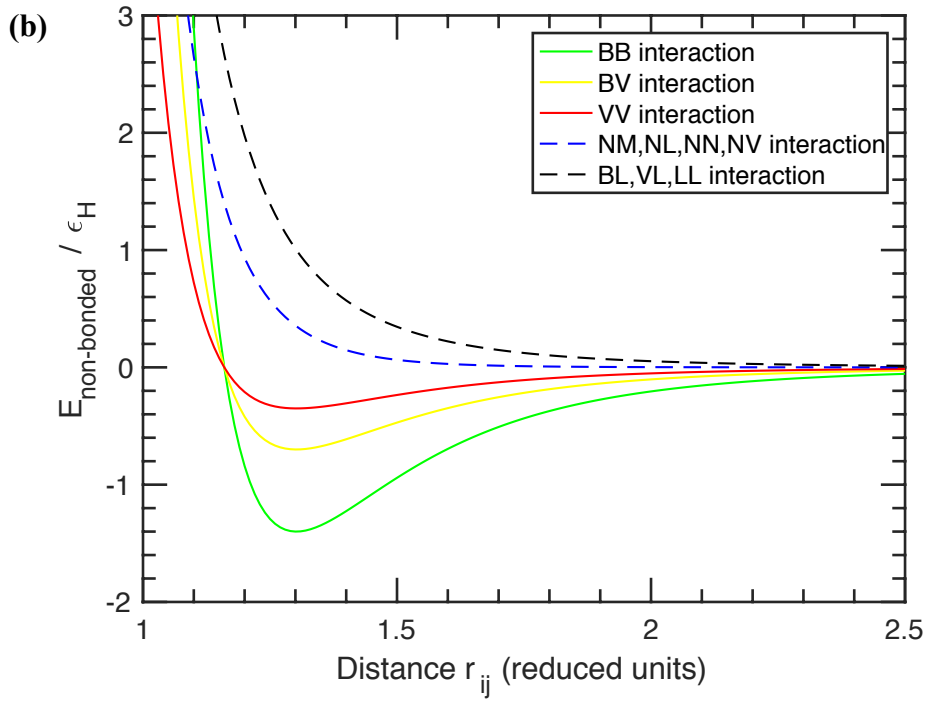


Figure S5.2: Non-bonded interaction potential as a function of reduced distance r_{ij} between bead i and j. Interactions BB, BV, and VV are attractive in nature with minima at r_{ij} = 1.3; while interactions NX and BL/VL/LL are purely repulsive.

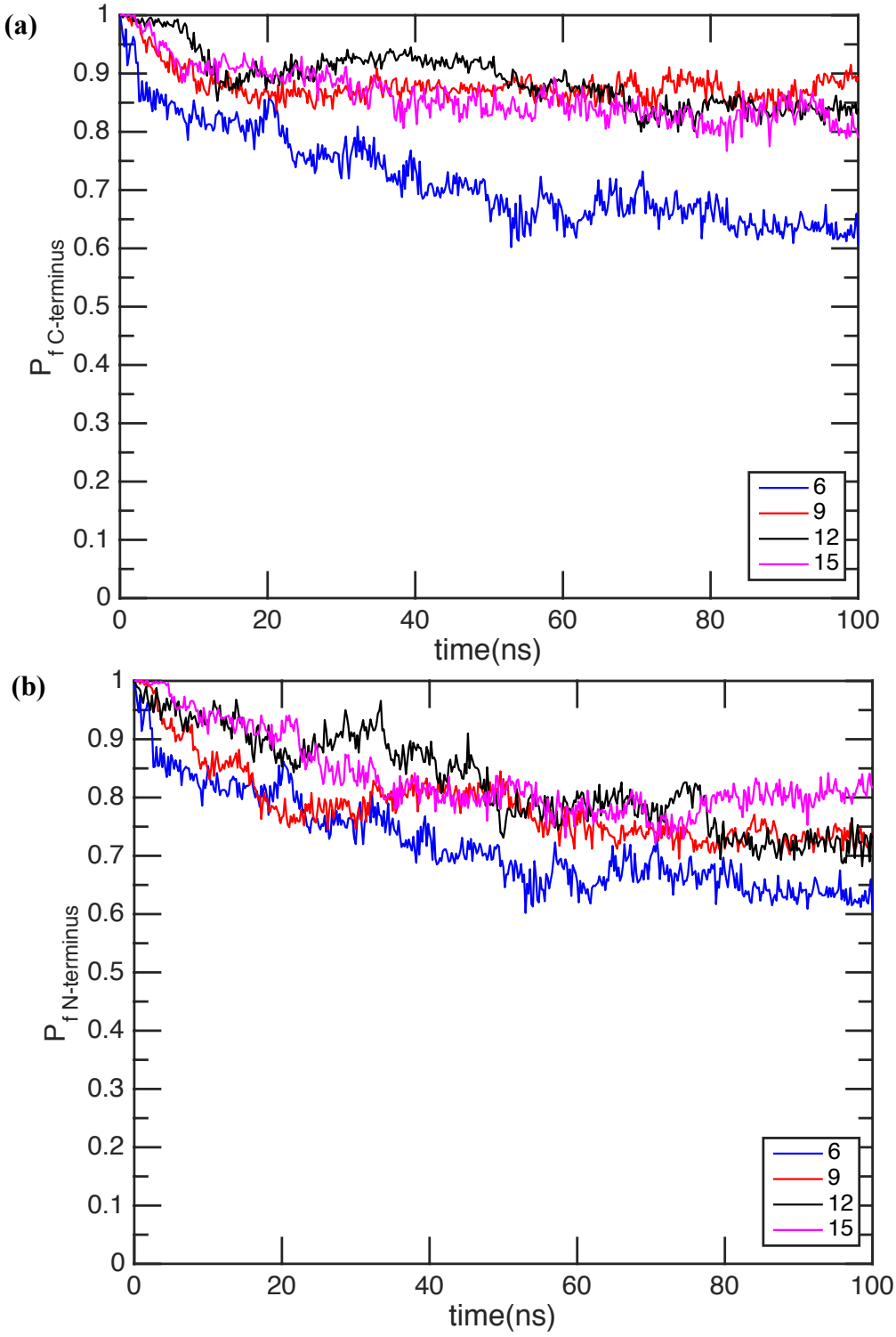


Figure S5.3: Time course of the fibril similarity parameter P_f based on all-atom simulations for different protofibril sizes for the two fibril tips - C-terminus (a) and N-terminus (b).

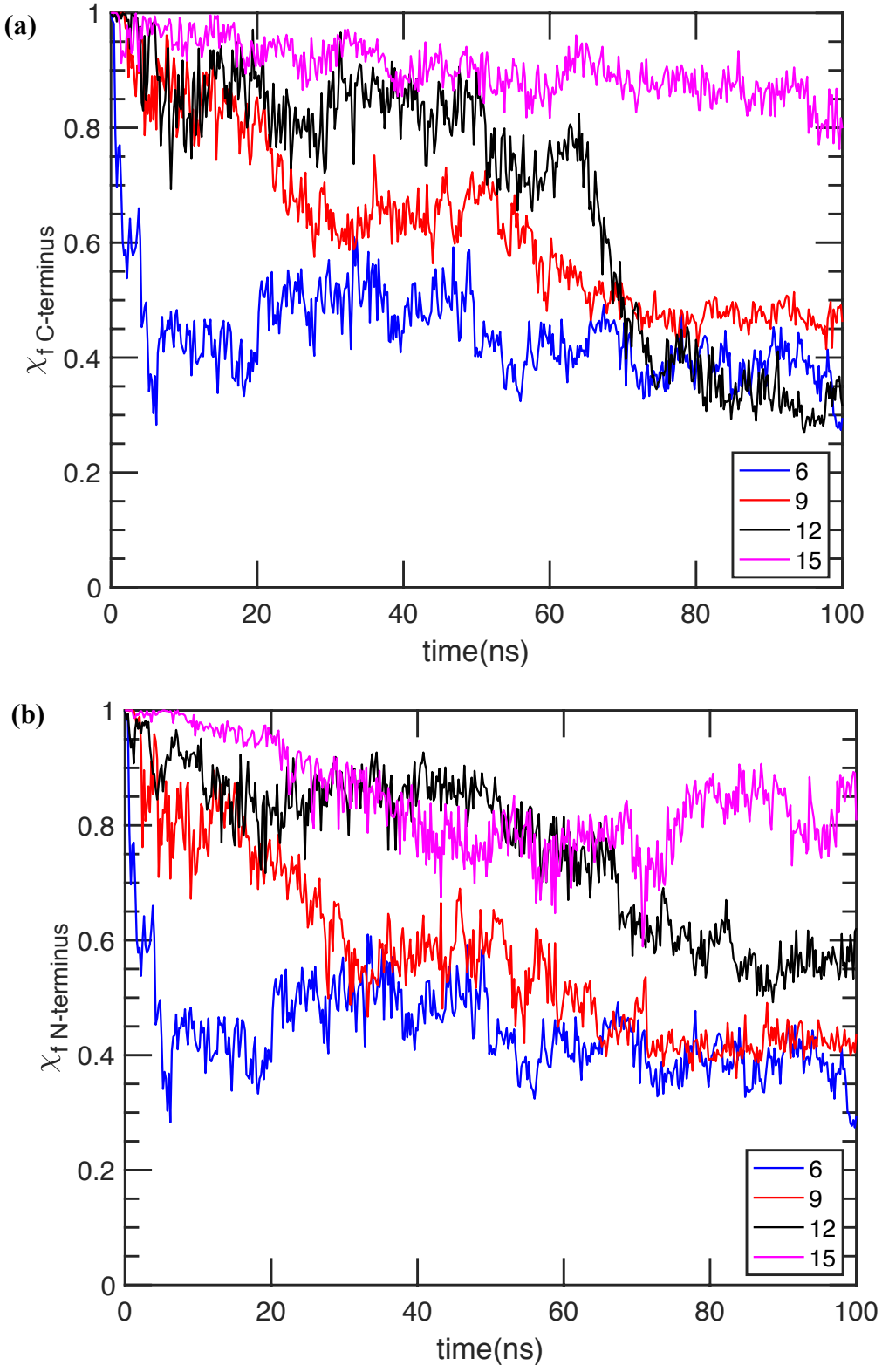


Figure S5.4: Time course of the fibril similarity parameter χ_f based on all-atom simulations for different protofibril sizes for the two fibril tips - C-terminus (a) and N-terminus (b).

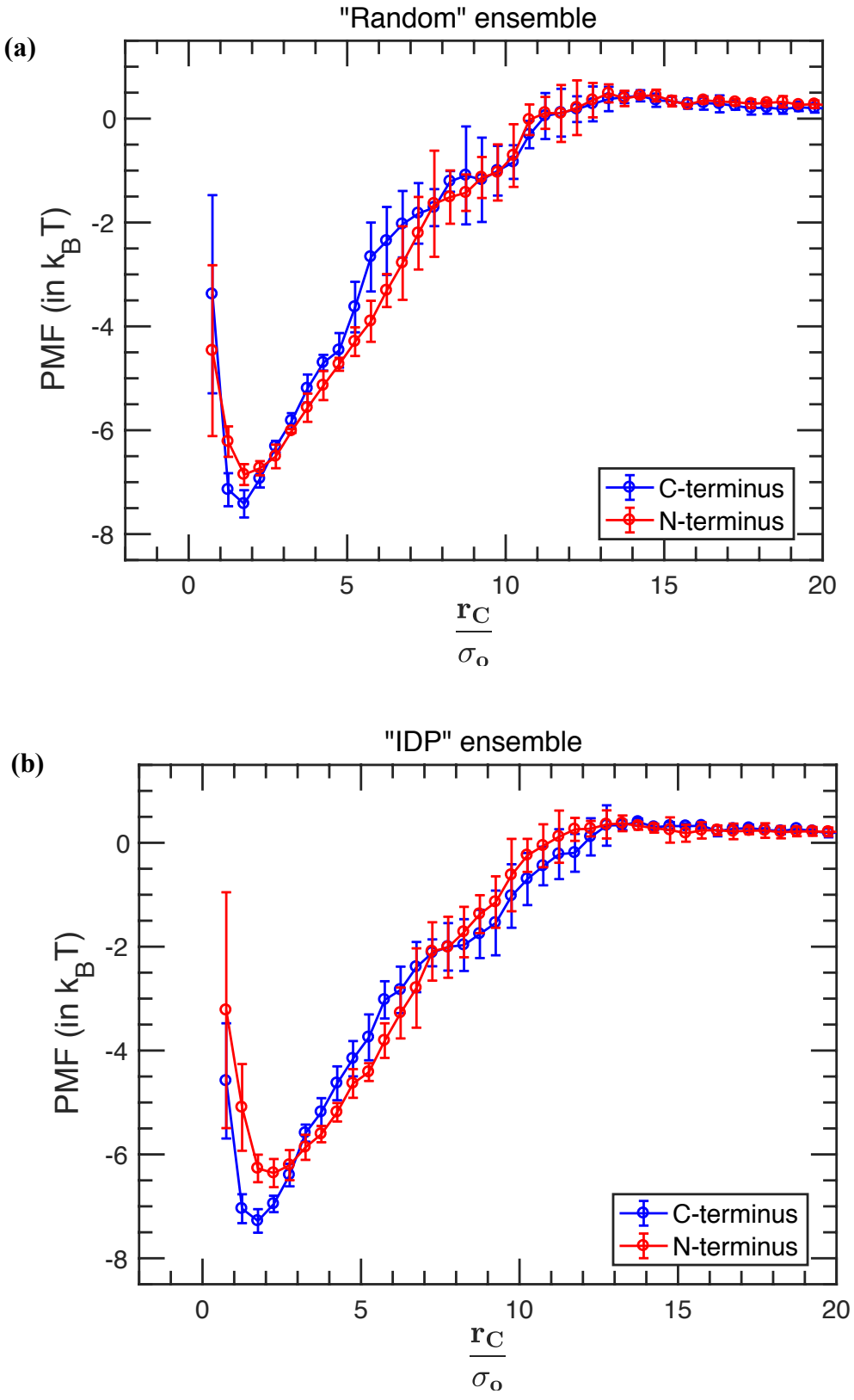


Figure S5.5: The potential of mean force ($F(\delta_C)$) profile for monomer addition at the two fibril tips as a function of separation distance δ_C for *Random* ensemble (a) and *IDP* ensemble (b).

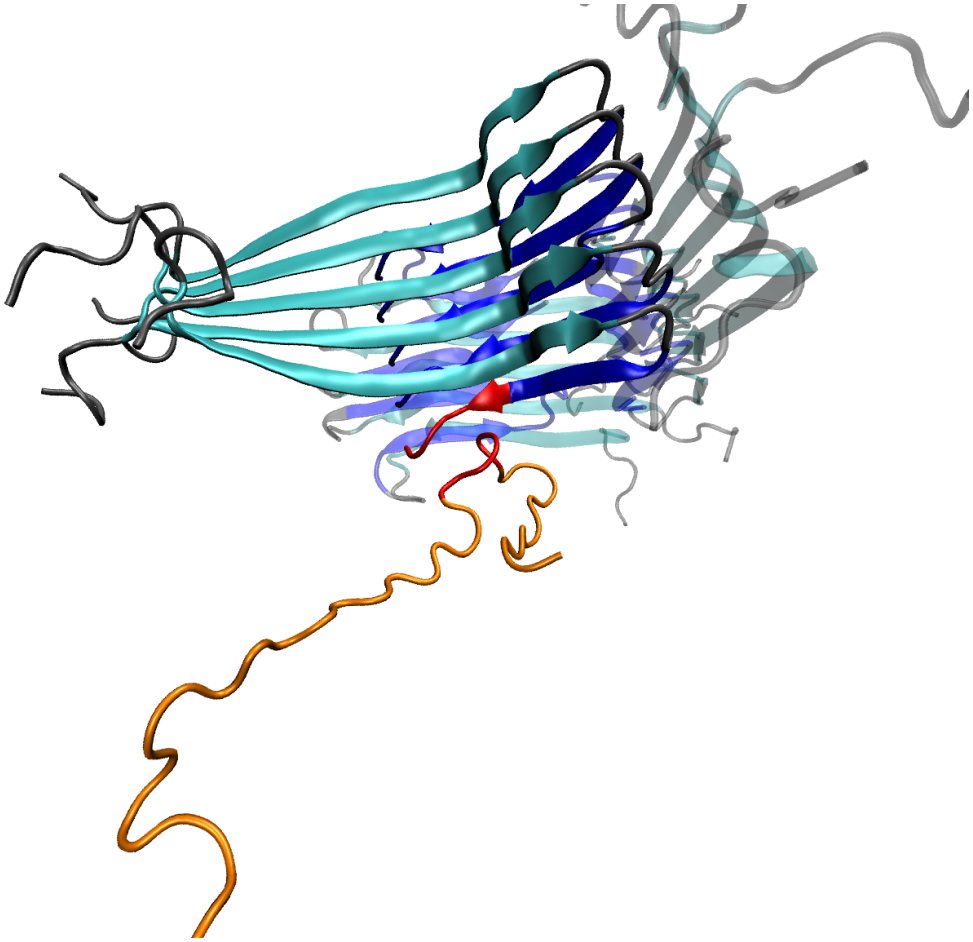


Figure S5.6: *Snapshot of a monomer forming 5 non-native contacts with a fibril tip during the "docking" step*. The non-native contacts are formed by residues 22-26 between the monomer (orange) and the fibril tip and are shown in red.

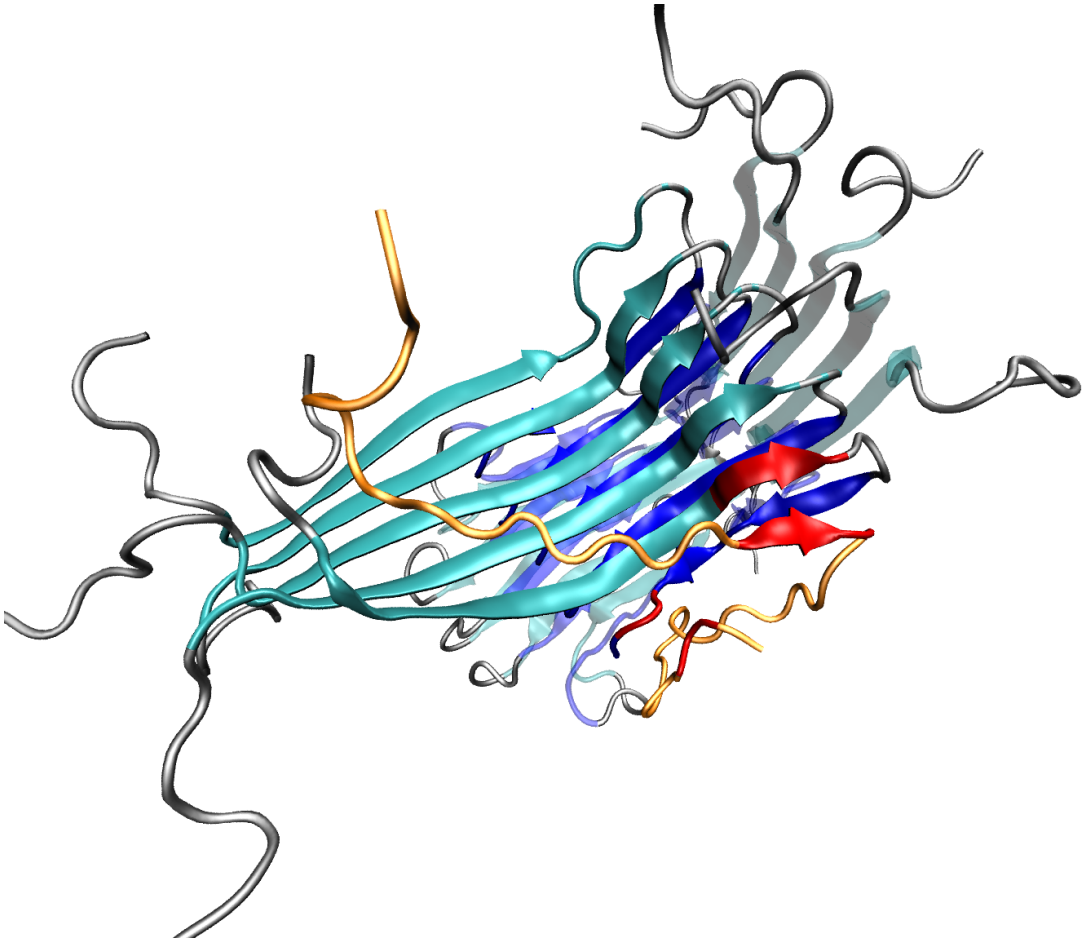


Figure S5.7: *Snapshot of the same monomer in figure S5.6 forming 5 native contacts with a fibril tip during the "locking" step*. The native contacts are formed by residues 20-22 and 36-37 between the monomer (orange) and the fibril tip and are shown in red.