ABSTRACT

WAGONER, VICTORIA. Computer Simulation Studies of Self-Assembly of Fibril-forming Peptides with an Intermediate Resolution Protein Model. (Under the direction of Carol K. Hall).

Assembly of normally soluble proteins into ordered aggregates, known as amyloid fibrils, is a cause or associated symptom of numerous human disorders, including Alzheimer's and the prion diseases. Recent experimental studies have offered tantalizing clues regarding the fibril structure, but our understanding of its assembly is still far from complete. The long term goal of our work is to determine the underlying physical forces responsible for the mis-folding and aggregation of proteins. Since the likely toxic species in the various amyloid diseases is believed to occur either on or off the fibrillization pathway, it is of interest to understand the connection between the protein sequence and the structure of the final product, the fibril. The focus here is on short fragments of amyloid proteins; these are believed to be the "Velcro" that holds the fibrillar structures formed by the parent protein together, and can in fact form fibrils themselves.

Accordingly, the aims of this work are to be able to predict and analyze: how variations in the sequence affect the likelihood that a given peptide will form a fibril, the structure of the fully formed fibril of a given sequence, what types of side chains disrupt assembly, and the kinetic events that occur along the fibrillization pathway. The method used is application of discontinuous molecular dynamics simulation to large systems of peptides that are modeled using PRIME20, a new intermediate–resolution protein force field developed in our group to describe the geometry and energetics for all twenty amino acids. Two different classes of peptides are studied: (1) palindromic aliphatic sequences

from the Syrian hamster and mouse prion proteins and variations thereof, and (2) short, truncated amyloid and amyloid-like peptides, some of whose fibril crystal structures have been measured.

We simulate the spontaneous assembly of several short prion and prion-like peptides starting from random initial configurations of random coils. We investigate fibril formation and structure of 48 peptides of palindromic prion sequences, AGAAAAGA (SHaPrP 113-120), VAGAAAAGAV (MoPr 111--120), and related variations, GAAAAAAG, $(AG)_4$, A8, GAAAGAAA, A10, V10, GAVAAAAVAG, and VAVAAAAVAV. We observe that as the chain length and the length of the stretch of hydrophobic residues increase, the ability to form fibrils increases. However as the hydrophobicity of the sequence increases, the ability to form well-ordered structures decreases. Thus, long hydrophobic sequences like VAVAAAAVAV and V10, form slightly disordered aggregates that are partially fibrillar and partially amorphous. Subtle changes in sequence result in slightly different fibril structures.

We study the spontaneous assembly of 48-peptide systems containing the short, amyloid peptide fragments and amyloid-like peptides: VEALYL, MVGGVV, SSTSAA, SNQNNF, GGVVIA, and the de novo designed peptides: STVIIE, STAIIE, STVIAE, STVIFE, STVIVE, STVIGE, and STVIEE. Fibril structure and formation kinetics are analyzed. The short amyloid peptide fragments SSTSAA and SNQNNF form fibrils at low temperatures; MVGGVV forms fibrils at intermediate temperatures, and GGVVIA forms β-sheets. For the *de novo* designed peptides, at high simulation temperatures, we observed fibrils for STVIIE, STVIFE and did not observe fibrils for STAIIE, STVIGE and STVIEE in agreement with experiments. The sequence STVIVE, formed fibrils in

our simulations but fibrils were not observed *in vitro* suggesting perhaps that PRIME20 is not robust enough to capture subtle difference in amino acid residues like isoleucine and valine.

Computer Simulation Studies of Self-Assembly of Fibril-Forming Peptides with an Intermediate Resolution Protein Model

by
Victoria Allen Wagoner

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Chemical Engineering

Raleigh, North Carolina

2010

APPROVED BY:

_____	_____
Carol K. Hall Committee Chair	Robert M. Kelly
_____	_____
Keith E. Gubbins	A. Clay Clark

DEDICATION

To my Granddaddy Tickle

BIOGRAPHY

Victoria Wagoner was born in Greensboro, NC on June 15, 1976. She is the granddaughter of Henry and Mozelle Tickle, Vernon Allen, and Dorothy and Jim Howell, and the daughter of Sonny and Eva Allen. She has one sister, Elizabeth Henley married to Jason Henley, who has three children: Kelly, Brittany and Noah. In December 1998, she received Bachelor's Degrees in Mathematics and Chemistry from the University of North Carolina at Chapel Hill. On April 24, 1999, she married Keith Wagoner. From 1998 to 2001 she worked as a research technician in both the UNC School of Medicine and the NCSU College of Veterinary Medicine where she participated in molecular biology and genetics research aimed at understanding the basis for diseases like breast cancer and Cystic Fibrosis. In addition she worked on understanding the immune response pathway in multiple sclerosis. In May 2001, she began work as an analytical chemist for ARCADIS, a contractor for the United States Environmental Protection Agency in RTP, NC. There she worked on a project to remove mercury from effluent coal-fired reactor stack gas and became interested in chemical engineering. In January 2003, she was admitted to North Carolina State University to pursue graduate studies in Chemical Engineering under the guidance of Carol K. Hall. She received her Master's of Science degree in Chemical Engineering from North Carolina State University in May 2005. In August 2010, she will begin her new career as a Consumer Safety Officer for the United States Food and Drug Administration.

ACKNOWLEDGEMENTS

It is my pleasure to acknowledge all the people who have contributed their wisdom, support and encouragement towards the completion of this thesis.

I would like to thank my advisor and mentor Professor Carol K. Hall for her support, encouragement and wisdom, both professionally and personally. Professor Hall is truly devoted to her students and without her support I would not be defending my PhD today.

I would like to thank Professor Robert M. Kelly and the National Institutes of Health Biotechnology Traineeship for the education and financial support that made part of this research possible. I would also like to thank the U.S. Department of Education Graduate Assistance in Areas of National Need (GAANN) Computational Science Fellowship for providing financial support. Lastly, I would like to thank the National Institutes of Health Research Grant for funding the entire project.

I would like to thank the Hall research group members, past and present, for providing a community of support and collective wisdom that helped to propel this project. The study of protein aggregation in the Hall Group began many years ago and would not be where it is today without the combined efforts of many researchers. I would like to give special thanks to Dr. Hung Nguyen, Dr. Alex Marchut, and Ms. Erin Phelps, who have each made significant contributions to PRIME and provided me with many hours of stimulating discussion. I am especially indebted to Dr. Mookyung Cheon

who worked with me this last year to take PRIME to the next level and has provided me with gentle guidance and wisdom. I would like to thank Dr. Andrew Schultz and Dr. Steven Smith both of whom developed and improved the discontinuous molecular dynamics simulation method that now allows us to study large macromolecule systems in a reasonable time frame.

I would also like to thank all of my fellow computer system administrators Andrew Schultz, Aysa Akad, Arthi Jayaraman, Alex Marchut, Amit Goyal, Johnny Maury-Evertsz and Ravish Malik, for all of the hard work, time and energy that is required to maintain our computer system of over forty clusters and almost two hundred processors.

I would like to thank Professor Keith Gubbins, Professor Robert Kelly and Professor Clay Clark for serving on my committee and for their advice on my research. I would especially like to thank all of the faculty and staff of NCSU's Chemical and Biomolecular Engineering Department, and Todd Marcks and Dr. Shaffer of the North Carolina State Graduate School for everything they have done for me while I have been at NC State.

I am especially thankful to my husband, Keith, for the boundless love and support that he has given me since the first day I mentioned attending graduate school in Chemical Engineering. He is my best friend and biggest fan. His shoulders have allowed me to reach the sky. He continues to surprise me with the depths of his love and support. I thank my wonderful sister, Elizabeth, and my brother-in-law, Jason Henley, for all the

vacations and good times we have shared and for bringing three of the neatest people into this world: my nieces, Kelly and Brittany, and my nephew, Noah. I thank my parents, Sonny and Eva, for allowing me the freedom to reach my goals and for believing that I could do anything. I would also like to thank my in-laws Wayne and Daphna Wagoner who, aside from raising a wonderful son, have embraced me as a daughter and provided me with encouragement to follow my dreams.

Although they are no longer here to read and take pleasure in my achievements, I give thanks to Henry and Mozelle (Granddaddy and Grandma) for showing me that hard work really can make a difference for a country girl. My grandma only had an eighth grade education but she raised five children and eleven grandchildren to believe in their dreams. My granddaddy was lucky enough to attend one semester at State College before he had to return home to work on the farm. He made sure that I would never have to face the same choice. I know he would be so happy and proud to see me receive my doctorate from N. C. State. Thank you Granddaddy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Motivation

The hallmark of many neurodegenerative diseases, including Alzheimer's disease, is the accumulation and deposition of protein plaques in specific tissues within various organs in the body [1]. These plaques are composed of ordered protein aggregates known as amyloid fibrils which form when normally-soluble disease-specific proteins undergo a conformational change that leads to their aberrant assembly. Many of the twenty-four known so-called amyloid diseases are fatal.

In vitro experiments on the many different proteins associated with amyloid diseases, e.g. β-amyloid (Alzheimer's), prion protein (transmissible spongiform encephalopathies), and huntingtin (Huntington's), indicate that even though these proteins have no obvious sequence homology, the amyloid fibrils formed from these different proteins share similar morphological properties [2-3]. X-ray diffraction studies of amyloid fibrils show that they are straight, rigid structures of varying length that consist of two or more smaller fibrils, called protofilaments. The protofilament itself is a 'cross-β' structure with β strands perpendicular to the fibril axis connected by backbone

hydrogen bonds to form β-sheets parallel to the fibril axis. Interestingly, proteins other than those known to cause one of the amyloid diseases have been found to form fibrils when their protein's native state is disrupted, such as in the presence of denaturant or under concentrated conditions [4]. This suggests that it is the basic interactions experienced by all proteins e.g., hydrophobic interactions and hydrogen bonding, rather than the chemistry associated with specific sequences that drive amyloid formation [5-7]. This implies that fibril formation may be an intrinsic property of many different proteins not just those associated with the amyloid diseases. However, not all peptides are able to form amyloid fibrils and sometimes certain side chains can prevent amyloid formation [8] which suggests perhaps that side chain identity does in fact influence fibril formation [9]. Recent experimental advances in the areas of solid-state NMR, cryo-EM, and x-ray diffraction on microcrystals of amyloid-forming peptides have helped to elucidate the structure of certain fibrils: GNNQQNY [10], Aβ9-40 [11], and other amyloid fibrils [12-14]. We now know that at the core of the fibril, the cross-β-structure is uniquely defined by the protein sequence. In fact a complementary inter-digitation of side chains similar to a steric zipper has been proposed for many amyloid peptides [15]. However our understanding of these complex structures is far from complete and in many ways the latest experimental data has resulted in more questions than answers. Understanding of the molecular mechanisms that drive amyloid fibril formation is still inconclusive even with the detailed molecular structures of specific fibrils that are becoming more and more available. The role that peptide sequence and other conditions (temperature, pH, etc.)

ultimately play in the fibril formation pathway is just beginning to be understood. As a result, while experimental research into the formation of amyloid is vital, a number of investigators have turned to computer simulation to gain a better understanding of the molecular-level details associated with amyloid formation. For example, Ma and Nussinov used all-atom molecular dynamics simulation to study pre-formed model fibrils containing short alanine-rich fragments (5-40 residues) of proteins associated with disease. These model fibrils remained stable over the length of the simulation [16], indicating that their postulated fibril structure may indeed be that of the real fibril. Additionally, our group developed an intermediate-resolution protein model, called PRIME which, when combined with discontinuous molecular dynamics, allows the simulation of protofilament formation starting from a completely random initial configuration of 48 chains of Ac-$KA_{14}K$-NH_2 [17]. These studies indicate that we may be able to shed light on the complex problem of protein aggregation by studying simple sequences with computer simulations. Important lessons about the essential physical features necessary for the formation and structure of disease-related amyloid fibrils could be learned with the use of computer simulations, provided that the models employed are reasonable representations of the real molecule. More details on the application of computer simulations to the study of protein aggregation can be found in the following publications [18-19].

This thesis describes research aimed at using computer simulations to develop an understanding of protein aggregation into amyloid fibrils. We have developed a new implicit solvent force field, PRIME20, to describe the geometry and energetics of all twenty amino acids [20]. We use discontinuous molecular dynamics computer simulation to explore the self-assembly and structure of homogeneous systems of known amyloid fibril peptides in an attempt to elucidate the role of peptide sequence and backbone in fibril formation.

1.2 Overview

In this section, we summarize the remaining chapters of this thesis. All chapters contain their own literature review and bibliography.

Chapter 2 gives a detailed description of how computational approaches are being used to learn more about the formation of amyloid fibrils. We describe two computational approaches used to investigate fibril formation and structure: intermediate-resolution discontinuous molecular dynamics simulations and atomistic molecular dynamics simulations. Each method has its strengths and weaknesses, but taken together the two approaches provide a useful molecular-level picture of fibril structure and formation.

Chapter 3 describes our results for homogeneous systems of truncated aliphatic prion-related sequences. We selected ten different sequences that may or may not form amyloid fibrils to varying degrees based upon the Syrian Hamster prion peptide, SHaPrP 113-120 (AGAAAAGA) and the mouse prion peptide, MoPrP 111-120 (VAGAAAAGAV). The other eight sequences were: GA_6G (a longer uninterrupted alanine stretch flanked by glycine), $(AG)_4$ (a complete disruption of the hydrophobic stretch), GAAAGAAA (a mimic of Aβ 28-35), A8, VAVAAAAVAV (a MoPrP with reduced flexibility), GAVAAAAVAG (a highly hydrophobic sequence), A10 and V10. Simulations are performed on 48-peptide homogenous systems containing the listed sequences starting from random configurations of random coils at high temperatures. The systems are cooled slowly to the desired simulation temperature. We explore how changing sequence and temperature affects the pathway of aggregation. Each sequence has a different propensity to form fibrils. For those sequences that do form fibrils, fibrillization occurs at an optimum temperature above which amorphous aggregates or amorphous aggregates with some β-structure are observed and below which random coils are observed. Our findings somewhat support the importance of the role of the hydrophobic stretch in a peptide's ability to aggregate. Additionally, we observe the progress along the fibrillization pathway from free monomers to ordered aggregates (fibrils). At the beginning of the simulation, monomers associate together into β-sheets of various sizes (dimer, trimer, etc.). The β-sheets then self-assemble into fibrils made up of two, three and sometimes four β-sheets.

Chapter 4 describes our attempts to characterize the kinetic pathway of fibril formation by simulating the spontaneous assembly of several short, truncated amyloid and amyloid-like peptides. We selected five amyloid fibril sequences with a known crystal structure: VEALYL, MVGGVV, SSTSAA, SNQNNF and GGVVIA and seven amyloid-like sequences with varying degrees of ability to form fibrils: STVIIE, STAIIE, STVIAE, STVIFE, STVIVE, STVIGE, STVIEE. Our goals were to test the ability of PRIME20 to distinguish the role played by each of the twenty different amino acids in fibril formation, to validate PRIME20's ability to predict each sequence's propensity to form fibrils, and to contribute to the fundamental understanding of the fibril formation pathway. Simulations are performed on 48-peptide homogeneous systems containing each of the peptides mentioned above starting from random configurations of random coils at high temperatures. The systems are cooled slowly to the desired simulation temperatures. For some sequences, we are able to observe β-sheet and fibril formation for a given temperature. For other sequences, we are unable to capture the fibril assembly process and further investigation is needed to determine if this is a limitation of our model or simply the inability of certain sequences to form fibrils. Each sequence follows a unique aggregation towards the final structure which indicates that PRIME20 is discerning enough to capture the differences in side chain identities.

Chapters 2, 3 and 4 are adapted from the following publications:

Chapter 2: V.A. Wagoner and C.K. Hall, "Computational approaches to fibril formation and structure", Methods in Enzymology, 2006. 412: p. 338-365.

Chapter 3: V.A. Wagoner and C.K. Hall, "Computer Simulation of Amyloid Fibril Formation by Palindromic Sequences in Prion Peptides", in preparation.

Chapter 4: V.A. Wagoner and C.K. Hall "Understanding the Molecular Assembly of Small Fibril-forming Peptides: A Computer Simulation Study", in preparation.

1.3 References

1. Koo, E.H., P.T. Lansbury, and J.W. Kelly, *Amyloid diseases: Abnormal protein aggregation in neurodegeneration.* Proc Natl Acad Sci USA, 1999. **96**: p. 9989-9990.

2. Blake, C. and L.C. Serpell, *Synchotron x-ray studies suggest that the core of transthyretin amyloid fibril is a continuous β-sheet helix.* Structure, 1996. **4**: p. 989.

3. Sunde, M., et al., *Common core structure of amyloid fibrils by synchotron x-ray diffraction.* J. Mol. Biol., 1997. **273**: p. 729.

4. Chiti, F., et al., *Designing conditions for in vitro formation of amyloid protofilaments and fibrils.* Proc. Natl. Acad. Sci. USA, 1999. **96**: p. 3590.

5. Kelly, J.W., et al., *Transthyretin quaternary and tertiary structural changes facilitate misassembly into amyloid.* Adv. Protein Chem., 1997. **50**: p. 161.

6. Kelly, J.W., *Towards an understanding of amyloidogeneisis.* Nat. Struct. Biol., 2002. **9**: p. 323.

7. Rochet, J.C. and P.T. Lansbury, *Amyloid fibrilogenesis: Themes and variations.* Curr. Opin. Struct. Biol., 2000. **10**: p. 60.

8. Monsellier, E. and F. Chiti, *Prevention of amyloid-like aggregation as a driving force of protein evolution.* European Molecular Biology Organization reports 2007. **8**(8): p. 737-742.

9. Fandrich, M. and C.M. Dobson, *The behavior of polyamino acids reveals an inverse side chain effect in amyloid structure formation.* European Molecular Biology Organizatin, 2002. **21**: p. 5682-5690.

10. Nelson, R., et al., *Structure of the cross-beta spine of amyloid-like fibrils.* Nature, 2005. **435**(7043): p. 773-8.

11. Petkova, A.T., W.-M. Yau, and R. Tycko, *Experimental constraints on quaternary structure in alzheimer's beta-amyloid fibrils.* Biochem, 2006. **45**: p. 498-512.

12. Makin, O.S., et al., *Molecular basis of amyloid fibril formation and stability.* Proc Natl Acad Sci U S A, 2005. **102**(2): p. 315-320.

13. Rambaran, R.N. and L.C. Serpell, *Amyloid fibrils.* Prion, 2008. **2**(3): p. 112-117.

14. Tycko, R., *Molecular structure of amyloid fibrils: Insights from solid-state nmr.* Quarterly Reviews in Biophysics, 2006: p. 1-55.

15. Sawaya, M.R., et al., *Atomics structures of amyloid cross-beta spines reveal varied steric zippers.* Nature, 2007. **447**: p. 453-457.

16. Ma, B. and R. Nussinov, *Molecular dynamics simulations of alanine rich β-sheet oligomers: Insight into amyloid formation.* Prot. Sci., 2002. **11**: p. 2335-2350.

17. Nguyen, H.D. and C.K. Hall, *Molecular dynamics simulations of spontaneous fibril formation by rando-soil peptides.* Proc Natl Acad Sci USA, 2004. **101**(46): p. 16180-16185.

18. Hall, C.K., et al., *Simulations of protein aggregation: A review*, in *Misbehaving proteins*, R.F. Murphy and A.M. Tsai, Editors. 2006, Springer New York.

19. Hall, C.K. and V.A. Wagoner, *Computational approaches to fibril formation and structure.* Methods in Enzymology, 2006. **412**: p. 338-365.

20. Cheon, M., I. Chang, and C.K. Hall, *Extending the prime model for protein aggregation of all twenty amino acids.* PRoteins, 2010. **in print**.

CHAPTER 2

Computational Approaches to Fibril Structure and Formation

2.1 Introduction

The hallmark of many neurodegenerative diseases, including Alzheimer's disease, is the accumulation and deposition of protein plaques in specific tissues within various organs in the body [1]. These plaques are composed of ordered protein aggregates known as amyloid fibrils which form when normally-soluble disease-specific proteins undergo a conformational change that leads to their aberrant assembly. Many of the twenty-four known so-called amyloid diseases are fatal. [2-5].

In vitro experiments on the many different proteins associated with amyloid diseases, e.g. β-amyloid (Alzheimer's), prion protein (transmissible spongiform encephalopathies), and huntingtin (Huntington's), indicate that even though these proteins have no obvious sequence homology, the amyloid fibrils formed from these different proteins share similar morphological properties [6-10]. X-ray diffraction studies of amyloid fibrils show that they are straight, rigid structures of varying length that consist of two or more smaller fibrils, called protofilaments. The protofilament

itself is a 'cross-β' structure with β strands perpendicular to the fibril axis connected by backbone hydrogen bonds to form β-sheets parallel to the fibril axis. Interestingly, proteins other than those known to cause one of the amyloid diseases have been found to form fibrils when their protein's native state is disrupted, such as in the presence of denaturant or under concentrated conditions [11-12]. As a consequence, a number of researchers have begun to think that it is the basic interactions experienced by all proteins e.g., the hydrophobic interactions and hydrogen bonding, rather than the chemistry associated with specific sequences, that drive amyloid formation [3, 11]. This implies that fibril or protofibril [13] formation may be an intrinsic property of many different proteins [4, 14-16] not just those associated with the amyloid diseases.

While experimental research into the formation of amyloid is vital, a number of investigators have turned to computer simulation to gain a better understanding of the molecular-level details associated with amyloid formation. For example, Ma and Nussinov recently used all-atom molecular dynamics simulation to study pre-formed model fibrils of short alanine-rich fragments (5-40 residues) of proteins associated with disease. These model fibrils remained stable over the length of the simulation [17], indicating that their postulated fibril structure may indeed be that of the real fibril. Additionally, our group has developed an intermediate-resolution protein model which, when combined with discontinuous molecular dynamics, allows the simulation of protofilament formation starting from a completely random initial configuration of 48 chains of Ac-$KA_{14}K$-NH_2 [18]. These studies indicate that we may be able to shed light

on the complex problem of protein aggregation by studying simple sequences with computer simulations. Important lessons about the essential physical features necessary for the formation and structure of disease-related amyloid fibrils could be learned with the use of computer simulations, provided that the models employed are reasonable representations of the real molecule.

In this paper we focus on the two methods mentioned in the previous paragraph, molecular dynamics simulations based on all-atom protein models and discontinuous molecular dynamics simulation based on intermediate resolution protein models, and their application to protein aggregation. Both methods are rooted in molecular dynamics simulation. In molecular dynamics, the trajectories of studied atoms are computed by solving Newton's equation of motion at regularly spaced time intervals, called time steps. At the beginning of each time step the net force acting on each molecule due to all the other molecules is calculated. Knowledge of this force allows the determination of each molecule's acceleration (F=ma), which in turn allows the prediction of the position and velocity of each molecule at the beginning of the next time step. At each time step, the instantaneous values of thermodynamic properties are computed and recorded. Thermodynamic properties are obtained by averaging the instantaneous properties over time. [19-21]

Discontinuous molecular dynamics (DMD) is a variant on standard molecular dynamics that is applicable to systems of molecules interacting via discontinuous

potentials (e.g., hard sphere and square-well potentials). Unlike soft potentials such as the Lennard-Jones potential, discontinuous potentials exert forces only when particles collide, enabling the exact (as opposed to numerical) solution of the collision dynamics. This imparts great speed to the algorithm, allowing sampling of longer time scales and larger systems than traditional molecular dynamics. The particle trajectories are followed by analytically integrating Newton's equations of motion, locating the time between collisions and then advancing the simulation to the next collision (event) [22-23]. DMD on chain-like molecules is generally implemented using the "bead string" algorithm introduced by Rapaport [24-25] and later modified by Bellemans et al. [26]. Chains of square-well spheres can be accommodated in this algorithm by introducing well-capture, well-bounce, and well-dissociation "collisions" when a sphere enters, attempts to leave, or leaves the square well of another sphere. In canonical ensemble DMD simulations (constant system size, volume and temperature), the temperature is maintained constant by implementing the Andersen thermostat method [27]; all spheres are subjected to random, infrequent collisions with ghost particles whose velocities are chosen randomly from a Maxwell Boltzmann distribution centered at the system temperature.

Perhaps the most important aspect of molecular simulations, whether we are speaking of continuous or discontinuous molecular dynamics, is what the simulation can teach us about the proteins we are studying. Simulations provide molecular snapshots of a protein that currently cannot be obtained from experiments. The snapshots tell us how proteins (or a system of proteins) arrange themselves, which residues are interacting via a

hydrogen bond, the distance between two key amino acids, etc. Structural properties can also be determined such as the radius of gyration which is a measure of the distance of particle i from the center of mass of the entire molecule

$$R_g^* = \frac{<1/N\sum_{i=1}^{N}[(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2]>}{\sigma^2 N} \quad (1)$$

In addition, knowing the internal energy (E) of the system allows us to calculate many other properties such as the melting temperature, free energy, and the heat capacity

$$C_v^* = \frac{C_v}{k_B} = \frac{[<E^2> - (<E>)^2]}{(\varepsilon T^*)^2} \quad (2)$$

by employing statistical mechanical relationships. These properties can be compared with experimentally-determined properties.

2.2 Protein Models

Protein folding models range from very simple lattice models or single-sphere-per-amino-acid continuum models to very complex all-atom continuum models in explicit solvent. All-atom simulation packages such as CHARMM[28], AMBER[29], DISCOVER[30], ENCAD[31], and ECEPP[32] provide explicit detail on the protein's geometry and its interactions with itself and solvent molecules, thereby accounting for all of the interactions involved in protein folding. The atomistic detail provided by these software packages comes at a cost in terms of the time scale that can be examined during

a simulation. The longest all-atom simulation to-date was a 1 μs simulation of the 36 residue villin headpiece starting from an initial configuration with some native state characteristics performed by Duan and Kollman [33]. (In comparison the time scale required for a protein to fold from a random coil is 10 – 100 μs [21].) This simulation took two full months on 256 dedicated parallel supercomputers yet they only achieved partial folding of the protein. Therefore, to simulate the complete folding process from a random coil configuration to the protein's native (folded) state, simplified representations (models) of proteins must be used. The big question is: How much detail is enough?

2.2.1a All-atom Molecular Dynamics

All-atom molecular dynamics simulations represent the protein at atomic resolution, generally accounting for every atom on the protein and on the solvent molecules. These simulations can provide crucial information regarding the stability of a final fibril structure. They can also provide a detailed description of the events occurring during the initial stages of aggregation, most commonly the formation of β-sheets which is believed to be a key step in amyloid fibril formation [34]. As mentioned earlier, in traditional molecular dynamics simulations Newton's equations of motion are numerically integrated to obtain the trajectories of each participating molecule. A necessary input into these calculations is a description of the molecular geometry and of the forces acting upon each molecule. The latter is expressed in terms of an energy function which is usually empirically based as opposed to being derived directly from

quantum calculations. The energy is then the summation of the intra and intermolecular interactions between all of the molecules. There are many all-atom molecular dynamics computer programs (CHARMM, AMBER, and ECEPP to name a few). The programs differ primarily in terms of the description of the forces acting on each atom (the forcefield). We will focus on the CHARMM energy function for illustrative purposes.

2.2.1b CHARMM

One of the most popular computer programs used to perform all-atom molecular dynamics simulations is CHARMM. In any of the all-atom MD packages, a molecular model (physical description of the protein) is needed before dynamic calculations can be made. In CHARMM, the molecular model includes information such as the protein sequence and the residue topology which are stored in the protein structure file (PSF). By residue topology, we mean a file containing data on the twenty amino acid residues including the identities of the constituent atoms, the values of the bond and torsional angles, the charge on each residue, and the identities of the atoms that are capable of hydrogen bonding. The purpose of the topology file is to enable the computer to generate a representation of the protein in question from its sequence alone [35]. CHARMM also incorporates another kind of file called a patch, which contains for example information about which residues can form disulfide bridges. In addition to the protein structure file, CHARMM requires a coordinate file, which contains the relative positions of all atoms (protein and solvent) in the input configuration, and a parameter file, which contains

information necessary for calculating the energy of the molecule such as the force constants, the equilibrium bond distance, bond angle, and dihedral angles, and the van der Waals radii. The cartesian coordinates of a particular protein can also be generated from x-ray crystallography data or NMR data provided by the protein databank. CHARMM uses periodic boundary conditions in order to limit the number of particles necessary to model the protein in solution. A summary of the files and information required in CHARMM to calculate the energy of a molecule in a particular conformation is shown in Figure 2.1 [28, 35]. These features are common to all CHARMM codes regardless of the type of simulation being performed.

The next component of CHARMM is the empirical energy function. There are many different forms for the empirical energy function but all can be expressed as a sum of the following types of terms

$$E = E_b + E_\theta + E_\varphi + E_\omega + E_{vdW} + E_{el} + E_{hb} + E_{cr} + E_{c\varphi} . \qquad (3)$$

The terms in the energy function include potential energies for the bond length, E_b, bond angle, E_θ, dihedral angle, E_φ, improper torsions, E_ω, electrostatic interactions, E_{el}, hydrogen bonding interactions, E_{hb}, and structural constraints, E_{cr} and $E_{c\varphi}$. Each term in Eq. (3) is a continuous potential that represents either a bond length, bond angle, torsional angle or nonbonded pair-wise interaction term. One example of a nonbonded interaction is the van der Waals energy, E_{vdW}, which is often represented with a Lennard-Jones potential. Another nonbonded interaction is the electrostatic potential, E_{el}, which can be represented in a number of different ways including having a constant dielectric

(Coulomb's Law), a distance-dependent dielectric, a shifted dielectric, a dielectric that depends on the distance between the center of geometry of the two groups containing electrostatic atoms i and j, or an extended dielectric that is the sum of a constant dielectric term and a second term that accounts for the approximate potential and field at atom i due to all other atoms not connected to atom i by angles and bonds outside a predetermined cutoff radius. [28]

An important consideration in computer simulations of protein folding and aggregation is the treatment of the solvent. There are many different options for representing solvent and/or its impact (e.g., the hydrophobic effect) in CHARMM. One of the most common potentials used in CHARMM to describe water-water interaction is the ST2 potential [28] but other models have been used successfully. The detailed description of water-water interactions that characterizes explicit-solvent all-atom molecular dynamics simulations severely limits the time scale that can be accessed. For example, in simulating a single protein molecule solvated with 3000 water molecules most of the computational time is dedicated to keeping track of the water-water interactions. The point is that all-atom explicit-solvent calculations are very expensive computationally and significantly reduce the accessible time scales even for a single protein.

In an effort to access longer time scales and to study more complex systems, several energy functions have been developed that model the solvent implicitly as a potential of mean force. A potential of mean force accounts for both the hydrophobic

effect and electrostatic screening effects. One popular implicit-solvent model is the CHARMM PARAM19 model [36-37]. In the CHARMM PARAM19 model the potential of mean force is described as the sum of both the intra-solute interactions and a mean solvation term. The solvation term is the sum over all the atoms of the product of an atomic solvation parameter and the solvent-accessible surface area [37]. This model assumes that the major contributor to the solvation energy is the interaction between the protein and its first shell of solvent molecules [38]. In an effort to increase computational speed, Ferrara et al. used an approximate method developed by Hasel and coworkers [39] to obtain an analytical expression for the solvent accessible surface area. Additionally, this model enforces a free energy penalty when a charged residue is buried in the protein interior and accounts for electrostatic screening effects with a distance-dependent dielectric. The CHARMM PARAM19 implicit-solvent model is much faster computationally than the explicit-solvent models and requires a mere 50% increase in computational time as compared to *in vacuo* simulations [37]. However, even with the use of an implicit-solvent model, the timescales accessible in all-atom protein simulations are limited because of the molecular detail of the model.

Once the energy function is fully defined and the protein structure file is generated, the starting initial configuration is subjected to an energy minimization step to find the nearest local free energy minimum configuration. The initial structure configuration is cooled to a temperature at or close to 0K (absolute zero) in which case the entropy is zero and so the internal energy is equivalent to the free energy. CHARMM

has five different energy minimization techniques: steepest descent (SD), conjugated gradient (CONJ), Powell (POWE), Newton-Raphson (NRAP), and truncated-Newton minimization package (TNPACK). These methods have the common goal of finding a set of coordinates which corresponds to a molecular conformation whose potential energy is at a minimum. [35] Energy minimization can also be used to eliminate structural defects caused by overlapping atoms and distorted bond and torsional angles. Energy minimization is performed not only in the case of a single molecule, but also in the case where the initial configuration is a pre-formed fibril or aggregate structure.

The next step after energy minimization is the calculation of the dynamics during which Newton's second law is numerically integrated, most often with the Verlet algorithm [19-20]. There are six basic steps in a typical dynamics run: the preliminary step (discussed above) which involves generating the protein structure file (PSF), the energy minimization step (also discussed above), the heating step, the equilibration step, the production step, and the quenching step. In the heating step, the velocities, which are randomly assigned to the atoms in the system from a Gaussian distribution, are increased at preset time intervals during dynamics calculations until the temperature of interest is reached. The heating step is followed by a period of equilibration until a stable temperature and structure are reached. The production step starts with the equilibrated structure; the particle trajectories (the dynamics) are followed by numerically integrating Newton's laws of motion with a time step on the order of 1-2 fs and for timescales of 10ns. [35] While the overall procedure seems quite straightforward, it can take months to

reach some biologically relevant time scales, like β-sheet formation, on a system of parallel supercomputers [33].

2.2.1c Simulation Results on Fibril Structure Using All-atom MD

Nussinov and coworkers have used the CHARMM polar hydrogen force field package in the presence of solvent to monitor the motions of 24 chains of Aβ(16-22), KLVFFAE, initially arranged in a double-layered antiparallel β-sheet configuration[40]. At the end of a 2.5 ns simulation these short highly hydrophobic sequences formed a double-layered β-sheet conformation. The strands within each β-sheet were antiparallel to each other with a 15° twist as shown in Figure 2.2. A more recent study by Haspel and coworkers on systems containing six or nine copies of a short sequence of the amyloidogenic human calcitonin hormone peptide (15-19) [41], DFNKF, led to the observation of a stable parallel β-sheet conformation in a single sheet layer.

The atomic-resolution simulations approach discussed thus far provides valuable information on the orientation of the peptides within a fibril but no information on the misfolding events that precede the formation of the fibril. In other words, the model fibril structure is postulated and is not based upon known or experimental information. In contrast the initial configuration in a molecular dynamics simulation of $A\beta_{1-40}$ by Petkova et al.[42] was taken from their group's solid-state NMR data on phi-psi angles. When experimental data was not available, the phi-psi angles were set to -140˚ and 140˚, respectively, the dihedral angles ω were set to 180˚, and the signs for the dihedral angles

were chosen to allow for intermolecular backbone hydrogen bonding. The CHARMm forcefield was used to model 5 copies of $A\beta_{1-40}$ starting from the in-register parallel (within each β-sheet) cross-β arrangement obtained from the NMR data as described above and shown in Figure 2.3A. After energy minimization, they observed a fibril structure with the following characteristics: a salt bridge between residues K28 and D23, an offset horseshoe structure for each strand which was held together by intra-chain association of hydrophobic side chains, and inter-chain backbone hydrogen bonding along the fibril axis as shown in Figure 2.3B.

Information regarding fibril structure can also be gleaned from atomic resolution simulations of the initial stages of aggregation. Gsponer et al.[36] explored the initial steps of yeast prion aggregation using CHARMM PARAM19, an implicit solvent model. They performed 20 MD simulations on a system containing three copies of the heptapeptide sequence, GNNQQNY, residues 7-13 of the yeast protein Sup35, starting from random conformations at 330K for a total of 20 μs. They found that the three peptides arranged themselves in a parallel orientation with predominantly β-structure. Santini et al.[43] studied a system of three $A\beta_{16-22}$ peptides arranged in six different starting configurations, including random unfolded chains, and an antiparallel β-sheet, to explore the early stages of aggregation. They used the ART-OPEP model in conjunction with the GROMACS simulation package with implicit solvent for a total simulation time of 20 ns. They found that the system preferred to be in an ordered β-sheet configuration. The system size was too small to make any conclusive remarks about parallel versus

antiparallel arrangements.

Atomic resolution simulations have also been used to study destabilizing events that are believed to happen early on during the fibrillization process, such as the α-helix to β-strand conversion or the partial unfolding of the native state into a conformation likely to aggregate into ordered structures. Recently, Tarus, Straub and Thirumalai[44] explored the initial stage of aggregation of $A\beta_{10-35}$, which is thought to be a dimerization based on experimental observations. They used the GRAMM energy minimization program to form two types of dimers, the φ-dimer which is stabilized by hydrophobic contacts, and the ε-dimer which is stabilized by electrostatic interactions. Each of the two dimers was then simulated for 10 ns using CHARMM PARAM22 with explicit solvent to determine its structural stability. The φ-dimer did not dissociate while the ε-dimer did, indicating that the expulsion of water molecules due to the interaction between hydrophobic residues on the φ-dimer may play a significant role in its structural stability.

2.2.2a Intermediate-resolution protein model with discontinuous molecular dynamics

Given the computational difficulties and limitations associated with simulating the folding of just a single protein, one would imagine that these problems would be even greater when the number of chains in the system is sizeable as in aggregation simulations. There is also the added difficulty of the long time scales involved. For example, it can take hours for proteins to aggregate *in vitro* [45-46]. Certain aspects of protein

aggregation (for example the cross-β structure in protofilaments) seem to be less dependent on the protein sequence than protein folding [11-12]. In fact, Wetzel and coworkers have shown experimentally that mutations in the hydrophobic core of Aβ1-40, residues 17-20, destabilizes the fibril structure but this effect is often counterbalanced by the formation of additional hydrogen bonds [47-48]. These observations suggest that protein aggregation may be less sensitive to the details of the intra-and inter-molecular interactions than protein folding. Therefore, simplified approaches have gained in popularity for studies of aggregation in multi-peptide systems.

There are two avenues by which all-atom molecular dynamics can be simplified. One approach is to change the manner in which the dynamics is calculated, and the secondapproach is to reduce the detail involved in the description of the molecule. These two simplifications can either be used together or separately. In our group's work we do both. We simplify the dynamics calculation by using discontinuous molecular dynamics and we reduce the detail in the protein's representation by using an intermediate resolution protein model.

2.2.2b Discontinuous Molecular Dynamics

DMD simulations begin by placing the model molecules into a virtual simulation box that is replicated in all dimensions to eliminate wall effects; this is known as periodic boundary conditions. The volume fraction of molecules in the simulation box for the case in which the molecules are spherical is given by $\varphi = \pi N\sigma^3 / 6V$, where N is the

number of molecules in the box, σ is the particle diameter, and V is the volume of the simulation box. In protein simulations one more commonly refers to the concentration of the system which is given by $c = N/L^3$ where L is the simulation box length. The temperature is set according to the equipartition theorem which relates the system kinetic energy to the temperature

$$\frac{1}{2}(m_i v_i)^2 = \frac{3}{2} N k_B T \qquad (4)$$

where m is the particle mass, v is the particle velocity, and k_B is the Boltzmann constant. As mentioned previously, the velocities are assigned randomly from a Maxwell-Boltzmann distribution about the system temperature. Protein aggregation simulations are often performed in the canonical ensemble, which means that the number of particles, volume, and temperature are held constant. Constant temperature is maintained by implementing the Anderson thermostat method [27].

The key steps in a DMD simulation are the calculation of collision times, t_{ij}, between particles i and j, and the calculation of the post-collision velocities for the colliding pair. DMD proceeds by calculating the collision times for all possible pairs of particles, determining which collision pair, i and j, has the smallest collision time, advancing the system to that event, and computing the system dynamics. A simple procedure for DMD is shown in Figure 2.4. For two particles, i and j, with diameter σ undergoing a hard sphere collision, the collision will occur at time t_{ij} that satisfies the condition

$$| \mathbf{r}_{ij}(t+t_{ij}) | = | \mathbf{r}_{ij} + \mathbf{v}_{ij}t_{ij} | = \sigma \qquad (5)$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$. The collision time t_{ij} is obtained by squaring equation (5) and solving the resulting quadratic equation

$$t_{ij} = \frac{-b_{ij} \pm \sqrt{b_{ij}^{\,2} - v_{ij}^{\,2}(r_{ij}^{\,2} - \sigma^2)}}{v_{ij}^{\,2}} \qquad (6)$$

where b_{ij} is defined as $b_{ij} = \mathbf{r}_{ij} \cdot \mathbf{v}_{ij}$. This equation must meet certain criteria for the calculated t_{ij} to be correct. If b_{ij} is greater than 0 then the spheres are moving away from each other and no collision will occur. If b_{ij} is less than 0 and the square-root discriminant is positive (to give a real solution) the spheres will collide [22-23, 49]. Equation 6 can be used to generate a list of all collision times for all possible colliding pairs, i and j. The next collision time t_c is the minimum collision time in this list. The system is advanced by t_c to the new position $\mathbf{r}_i(t+t_c)$, at which a single pair of particles is in contact preparing to undergo a collision

$$\mathbf{r}_i(t + t_c) = \mathbf{r}_i(t) + \mathbf{v}_i t_c \quad . \qquad (7)$$

After the collision occurs, the post-collision velocities are calculated. The new velocities for interacting particles i and j are determined by imposing conservation of kinetic energy and conservation of linear momentum. The new velocities are given by

$$v_i(after) = v_i(before) + \Delta\mathbf{v}_i \qquad (8)$$

$$v_j(after) = v_j(before) - \Delta\mathbf{v}_i \qquad (9)$$

where v_i represents each component of the velocity vector and the velocity change, $\Delta\mathbf{v}_i$, for a hard core collision between collision partners i and j separated by a distance σ is

$$\Delta \mathbf{v}_i = -\Delta \mathbf{v}_j = -\left(\frac{b_{ij}}{\sigma^2}\right)\mathbf{r}_{ij}. \qquad (10)$$

After each event between partners i and j, the collision time list is updated for particles i and j and any other particles that would have interacted with either i or j. [22, 50]

The equations described thus far are applicable to hard sphere particles undergoing a core collision. Real intermolecular potentials, however, include not just repulsive interactions (like the hard sphere case) but also attractive interactions. The square-well potential is one such potential that has a repulsive interaction at short distances and an attractive interaction at intermediate distances. The square-well potential is mathematically described as,

$$U(\mathbf{r}_{ij}) = \begin{cases} \infty & (r < \sigma_1) \\ -\varepsilon & (\sigma_1 \le r \le \sigma_2) \\ 0 & (r > \sigma_2) \end{cases} . \qquad (11)$$

The equations described above for hard sphere particles can easily be extended to the case of the square-well potential [22].

In the square-well potential model, two basic types of collisions or events can occur [22]. Two particles either undergo (1) a core collision as described for the hard sphere system at $\mathbf{r} = \sigma_1$, or (2) an attractive collision at $\mathbf{r} = \sigma_2$, which can either be a capture, dissociation, or bounce. A capture event occurs when the attractive wells of particles i and j collide, a dissociation event occurs when two particles already inside the well become separated, and a bounce event occurs when two particles inside a squarewell fail to separate due to insufficient kinetic energy. The general procedure is the same as

that shown in Figure 2.4. The collision times are functions of the separation σ_1 or σ_2, resulting in slightly different formulas for calculating the collision times and the velocity changes. Alder and Wainwright provide a detailed flow-chart that can be used to determine which type of collision (event) will take place between two interacting particles, i and j. This is reproduced in Figure 2.5 [22]. Additionally, in DMD we have pseudo-events for bookkeeping purposes. These pseudo-events include implementing the thermostat, implementing efficiency techniques, and data collection.

Although we have only discussed systems of spheres thus far, the equations that we have introduced can easily be adapted to systems of chain-like molecules (peptides or polymers). DMD on chain-like molecules is implemented using the "bead-string" algorithm of Rapaport [24-25] and later modified by Bellemans [26]. In Bellemans's algorithm, adjacent spheres along a chain are bonded together by short, invisible strings whose length ensures that the bond length between spheres varies freely between $(1+\delta)l$ (bond extension) and $(1-\delta)l$ (bond contraction), where l is the ideal bond length and δ is a tolerance for bond length fluctuations ($\delta << l$). This means that bonded spheres along the chain are partially decoupled from each other, moving freely along linear trajectories between bond stretch (extension) collisions and core (contraction) collisions. The collision times and post-collision velocities for a bond contraction are the same as for the hard sphere case. However, the bond extension time is given by

$$t_{ij}^{Bond} = \frac{-b_{ij} + \sqrt{b_{ij}^{\ 2} - v_{ij}^{\ 2}(r_{ij}^{\ 2} - (1+\delta)^2 l^2)}}{v_{ij}^2} \qquad (12)$$

where $(1+\delta)l$ represents the maximum bond length.

The execution speed of DMD is proportional to N^2 but there are several efficiency techniques developed by Smith et al. that can be utilized to reduce the execution speed to be proportional to N. See Smith *et al.* (1997) for a more detailed discussion.

2.2.2c Intermediate Resolution Protein Models – PRIME

Inspired by the early reduced representation model of Takada [51], our group developed an intermediate-resolution protein model for simulations of protein folding and aggregation [52-54]. In this model, which we now call PRIME (for Protein Intermediate-Resolution Model) the protein backbone is represented by three united atom spheres, one for the amide group (NH), one for the carbonyl group (CO) and one for the alpha-carbon and its hydrogen ($C_\alpha H$). The side chains are modeled with a single sphere of variable size. All backbone bond lengths and angles are set to their ideal values. As mentioned earlier for DMD on chain-like molecules, the covalent bonds are maintained with a hard sphere interaction occurring when the bond lengths move outside of the range $(1+\delta)\ l$ to $(1-\delta)\ l$, where l is the ideal bond length and δ is the tolerance for acceptable fluctuation in bond lengths which is set at 2.375% [26]. The covalent bond lengths in this model are given in Table 2.1. Ideal backbone bond angles, C_α-C_α distances and residue L-isomerization are fixed through a series of pseudobonds which are also allowed to

fluctuate within 2.375% of their given length. A depiction of the protein model for alanine showing the backbone united atoms, NH, $C_\alpha H$, and C=O, and side chain united atom, CH_3, along with the covalent bonds and pseudobonds is given in Figure 2.6. The values of bond angles and pseudobond lengths are also given in Table 2.1. Local interactions between united atoms separated along the protein backbone by three or fewer onds are modeled in a similar manner to nonlocal interactions but with different bead diameters. Takada et al. found that it was more appropriate to describe the local interactions with the real atomic diameters (N, C_α, C) rather than the united atom diameters given in Table 2.1. In order to account for the interactions between atoms connected by three or fewer bonds, we allow 25% overlap of their united atom bead diameters. This treatment of local interactions successfully limits the motion of the phi and psi dihedral angles, yielding reasonable Ramachandran plots [51, 53]. Additionally in PRIME, the side chain can either be represented with a single sphere, as in the case of alanine, or by several spheres, as in the case of glutamine; glycine is incorporated into the alpha carbon united atom and therefore not represented as a sphere.

Hydrogen bonding between the peptide backbone amide groups and carbonyl groups on the same or neighboring chain is represented by a directionally-dependent square well attraction between the NH and C=O united atoms of depth ε_{HB} whenever: (1) the virtual hydrogen and oxygen atoms (whose location can be calculated at any time) are separated by 4.2Å (the sum of the NH and C=O well widths), (2) the nitrogen-hydrogen and carbon-oxygen vectors point towards each other within a fairly generous tolerance,

and (3) neither the NH nor the C=O are already involved in a hydrogen bond with a different partner. In order to ensure that criteria 1 and 2 are satisfied we require that the four atom pairs $N_i - C_{\alpha j}$, $N_i - N_{j+1}$, $C_j - C_{\alpha i}$, $C_j - C_{i-1}$ shown connected by thick dashed lines in Figure 2.7 (hereafter referred to as auxiliary pairs), be separated by a distance greater than d_{ij} which is chosen to maintain the hydrogen bond angle constraints [55-56]; their values are given in Table 2.2. Upon the formation of a bond between N_i and C_j, these auxiliary pairs temporarily interact via a square-shoulder potential:

$$U(\mathbf{r}_{ij}) = \begin{cases} \infty & (r < \sigma) \\ \varepsilon_{HB} & (\sigma_1 \le r \le d) \\ 0 & (r > d) \end{cases} \qquad (13)$$

where r is the distance between spheres i and j; σ is the sphere diameter; ε_{HB} is the shoulder height (equal to the well depth of the hydrogen bond between N_i and C_j) and d is the square-shoulder width. These auxiliary pairs return to their original interactions when the hydrogen bond is broken.

The solvent molecules are modeled implicitly via a square well attraction (potential of mean force) between two hydrophobic residues, and a hard sphere repulsion between two polar residues or between a polar and a hydrophobic residue. The depth of the square well attraction, ε_{HP}, between two hydrophobic residues is scaled relative to ε_{HB} by a factor R, which describes the solvent characteristics of the system [55]; R is defined as $\varepsilon_{HP}/\ \varepsilon_{HB}$. All system parameters are scaled by ε_{HB}, so that the system reduced temperature is $T^* = k_B T/\ \varepsilon_{HB}$.

2.2.2d Simulation results on fibril formation and structure using DMD

We performed simulations on a single model polyalanine sequence, Ac-$KA_{14}K$-NH_2 [55]. Polyalanine was chosen for study because Ac-$KA_{14}K$-NH_2 forms fibrils in vitro as shown by Blondelle and coworkers [45]. We began by exploring how the temperature and hydrophobic interaction strength, as modeled through the parameter R, affected the conformational conversion of the isolated chain. At low temperatures and low hydrophobicity ($0 < R < 1/10$) we observed a transition from an α-helix to random coils. However, as the hydrophobic interaction strength was increased ($1/4 < R < 1/2$), a third conformational transition from an α-helix to a β-sheet structure was observed. In that case as the temperature increased, the isolated peptide adopted an α-helix, then a β-hairpin or β-sheet and finally a random coil configuration. Finally, at high hydrophobic interaction strength ($R > 1/2$), the model polyalanine formed only random coils. Since there is little evidence to support a three state transition (α-helix → β-hairpin → random coils), a value of R was chosen low enough to avoid the three state transition but yet high enough so that the most stable state at a low temperature is an α-helix.

We next used PRIME and DMD to explore the phenomena of protein aggregation with large systems comprised of up to 96 16-mers with the $KA_{14}K$ sequence. Starting from a random conformation, we were able to simulate the spontaneous formation of a protofilament or small fibril [18]. Snapshots of the simulation are shown in Figure 2.8. The protofilament was composed of β-sheets with residues positioned in an in-register parallel arrangement. By performing many simulations at different temperatures,

concentrations, and solvent strengths, we determined that as temperature and concentration were increased the number of α-helices decreased and the number of extended ordered structures increased. The protofilament or fibril was found to be stable at temperatures higher than the folding temperature.

We went on to describe the dependence of peptide aggregation on peptide concentration and temperature by conducting equilibrium simulations using the replica-exchange technique [57] on a system containing 96 16-mers of Ac-KA14K-NH_2. A phase diagram in the temperature--concentration space was mapped out, illustrating which structures were stable at each condition [58]. The α-helical region was stable at low temperatures and low concentration. The non-fibrillar β-sheet region was stable at intermediate temperatures and relatively low concentrations and expanded to higher temperatures as concentration was increased. The fibril region was primarily stable at intermediate temperatures and intermediate concentrations and expanded to lower temperatures as the peptide concentration was increased. Finally, the random coil region was stable at high temperatures at all concentrations. Interestingly, we were able to observe the formation of small fibrils (protofilaments) for systems containing 96 peptides within 160 h on an AMD Athlon MP2200+ workstation.

We also investigated the kinetics of fibrillization by simulating systems of 48 to 96 peptides of sequence $KA_{14}K$ [59] using the intermediate resolution protein model, PRIME. We performed both seeded and unseeded simulations to determine if fibril

formation was a nucleation dependent event. In the presence of a seed (a preformed aggregate) the lag time, which is the time to form a fibril or fibril component, disappeared. The lag time decreased with increasing temperature and concentration. Fibril formation proceeded in the following way: small amorphous aggregates associated, rearranged into ordered β-sheet structures and ultimately formed a "nucleus", which rapidly grew into a small fibril or protofilament. We observed two growth mechanisms: lateral addition in which a β-sheet was added to the side of the fibril, and end-to-end growth in which individual peptides were attached to the end of each β-sheet (this mechanism accounts for the indeterminate length of the fibril). Once the fibrillar structure reached a critical number of β-sheets, the monomeric peptides tended to attach to an already formed β-sheet rather than to form a new isolated β-sheet. The number of critical β-sheets was a function of system size. A 12 peptide system formed a fibril with 2-3 β-sheets, a 24 peptide system formed a fibril with 3-4 β-sheets, a 48 peptide system formed a fibril with 3-6 β-sheets and a 96 peptide system formed a fibril with 4-6 β-sheets.

We have extended the PRIME model to polyglutamine to study the aggregation of polyglutamine-containing proteins [60]. The backbone of a polyglutamine residue is modeled with the same level of detail as a polyalanine residue (a single sphere for the carbonyl group, amide group, and alpha-carbon group). The difference between the polyglutamine and polyalanine models is in the description of the side chain. The polyglutamine side chain is represented with four united atom residues, two for the

hydrophobic methyl groups (CH_2, in blue), one for the carbonyl group (CO, in red) and one for the amine group (NH, in green) as shown in Figure 2.9. The carbonyl and amine groups along the side chain allow the side chain to participate in hydrogen bonding either with the backbone amide or carbonyl groups or with other side chains. Simulations were conducted on a system of 24-16 mers (Q_{16}) at a concentration of 5 mM over a range of temperatures starting from a random initial configuration. Amorphous aggregates were formed at low temperatures, annular structures composed of β-sheets were formed at intermediate temperatures, and random coil configurations were found at high temperatures. Figure 2.10 shows a snapshot of one of the annular structures. Similar annular structures were observed by Wacker and coworkers on two cleavage products of the huntingtin protein with 20 and 53 polyglutamine repeats, respectively [61] using atomic force microscopy and were also predicted by Perutz from x-ray scattering data on $D_2Q_{15}K_2$ [62].

The PRIME model is now being extended to the description of heteroproteins. We have performed DMD simulations on the Mouse prion peptide (111-120) of sequence, VAGAAAGAV (Wagoner and Hall, unpublished). Preliminary results from simulations of the isolated peptide reveal that it is too short to adopt a discernible secondary structure. It forms one to two beta-hydrogen bonds at temperatures between $T^* = 0.07$ and 0.11 and it exists as a random coil at temperatures greater than $T^* = 0.11$. Aggregation studies have been performed on a system of 48 peptides of VAGAAAAGAV at a concentration, c = 5 mM, a range of reduced temperatures,

T*=0.07 -- 0.14, and a hydrophobic interaction strength, R=1/10, starting from a configuration of random coils. Amorphous aggregation is observed at or below T*=0.11. Ordered aggregates are formed at temperatures greater than T*=0.11 but less than T*=0.13. The ordered aggregates are composed of stacks of four to six β-sheets. The peptides remain in random coil configurations at temperatures greater than or equal to T*= 0.13. We have also conducted simulations at concentrations, c = 1 mM, 2.5 mM and 10 mM at reduced temperature, T* = 0.12. The peptides formed β-sheets at concentrations greater than c = 1 mM but less than c = 5 mM, and formed ordered aggregates (a protofilament) at concentration, c = 10 mM. Figure 11 is a snapshot of an ordered aggregate formed by a 48 peptide system of VAGAAAAGAV at concentration, c = 1 mM and reduced temperature, T* = 0.12 where peptides which form a β-sheet are colored the same. For example, in Figure 2.11 there are two β-sheets formed, green and blue, respectively, which then associate to form an ordered aggregate.

The great speed of DMD simulations with reduced representation protein models has inspired other groups to develop protein models based upon hard sphere and square well interactions. Dokholyan and coworkers have applied DMD to the study of a two-sphere per residue coarse-grained protein model [63-64] with a Go-type potential used to describe the intermolecular interactions between the side chains. Ding et al. [64] used this two-sphere model to study the aggregation of a system containing eight copies of the Src SH3 protein. Starting from a system of random coils, they observed a possible pathway for amyloidogenesis that included the formation of four dimers → two bundles

of four proteins → one aggregate. Peng et al. [65] applied the two-sphere model to study a large system of β-amyloid proteins, $A\beta_{1-40}$. They started with 28 copies of the peptide arranged randomly, with each peptide in a α-helical conformation as determined by the NMR measurements of Coles et al. [66], at concentration c = 6 mM. They observed the formation of amorphous aggregates at temperatures below the melting temperature of a single peptide, T*=0.4 and dissociation of all structures at temperatures greater than T* = 1.10. Multilayer β-sheet structures formed over a range of simulation temperatures between T* = 0.55 and T* = 1.10.

Recently, Urbanc et al. [67-68] used a model similar to the four-sphere intermediate resolution model introduced by Smith and Hall [53-54] along with DMD to study the oligomerization of amyloid β-protein, $A\beta_{1-40}$ and $A\beta_{1-42}$. They observed an α-helix to β-strand structural transition at intermediate simulation temperatures followed by another transition from β-strand to random coil at relatively high simulation temperatures. They also observed a β-turn between residues D23 and K28. Although a turn is observed experimentally between residues D23 and K28, it is not a true β-turn (which would involve hydrogen bonding between residues D23 and K28) but rather an electrostatically driven salt-bridge which leaves the backbone amide and carbonyl groups free to hydrogen bond with another $A\beta_{1-42}$ peptide [42]. Urbanc et al. [67] went on to study the aggregation of 32 copies of $A\beta_{1-40}$ and $A\beta_{1-42}$ starting from a mostly α-helical configuration at concentration 3.4 mM and simulation temperature T* =0.15. They found that $A\beta_{1-40}$ preferred to form a dimer and $A\beta_{1-42}$ preferred to form a trimer and a

pentamer. The cores of both oligomers were comprised primarily of hydrophobic residues. The observation of the pentamer is interesting because experiments suggest that the composition of protofibrils or paranuclei is predominantly pentameric and hexameric [69-70]. While they were able to see the formation of oligomers, their simulation time was too short to observe the formation of the fibril structure reported by Petkova et al. [71].

2.3 Conclusion

Great progress has been made over the past five years in advancing the study of fibril structure and formation through the use of computer simulations. We have discussed two simulation approaches: atomic resolution models and intermediate-resolution protein models. Each gives us an insightful but incomplete view of the fundamentals associated with fibrillization. Which approach is the most fruitful avenue to pursue? In trying to answer this question the authors usually wind up thinking about the fable of the six blind men who went to "see" an elephant. Each touches a different part of the elephant – side, tusk, trunk, knee, ear, tail – and concludes that an elephant is like – a wall, a spear, a snake, a tree, a fan, a rope. Mistaking the parts for the whole, the blind men then argue passionately about the nature of the elephant, each convinced that his view is correct. Our community is, of course, wiser than the blind men (at least when it comes to science) because we know that the different computational tools (as well as

experimental tools) that we use afford us views of different aspects of fibril formation. It is only by sharing this information and piecing together our various observations that we will be able to assemble a good comprehensive picture of the nature of fibril formation and structure.

We thank Dr. Hung D. Nguyen (UC Irvine) and Dr. Alexander J. Marchut (Abbott Laboratories, Chicago, IL) for insightful discussions. This work was supported by National Institutes of Health Grant GM-56766.

2.4 References

1. Koo, E.H., *The β-amyloid precursor protein (app) and alzheimer's disease: Does the tail wag the dog?* Traffic, 2002. **3**: p. 763-770.

2. Bucciantini, M., et al., *Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases.* Nature, 2002. **416**: p. 507.

3. Dobson, C.M., *The structural basis of protein folding and its links with human disease.* Phil. Trans. R. Soc. Lond. B., 2001. **356**: p. 133.

4. Kelly, J.W., *The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways.* Curr. Opin. Struct. Biol, 1998. **8**: p. 101.

5. Prusiner, S.B., *Prion diseases and the bse crisis.* Science, 1997. **278**: p. 245-251.

6. Sunde, M. and C.C.F. Blake, *The structure of amyloid fibrils by electron microscopy and x-ray diffraction.* Adv. Protein Chem., 1997. **50**: p. 123.

7. Sunde, M., et al., *Common core structure of amyloid fibrils by synchotron x-ray diffraction.* J. Mol. Biol., 1997. **273**: p. 729.

8. Blake, C. and L.C. Serpell, *Synchotron x-ray studies suggest that the core of transthyretin amyloid fibril is a continuous β-sheet helix.* Structure, 1996. **4**: p. 989.

9. Serpell, L., et al., *The protofilament substructure of amyloid fibrils.* J. Mol. Bio., 2000. **300**: p. 1033-1039.

10. Serpell, L.C., *Alzheimers amyloid fibrils: Structure and assembly.* Biochimica et Biophysica Acta, 2000. **1502**: p. 16.

11. Chiti, F., et al., *Designing conditions for in vitro formation of amyloid protofilaments and fibrils.* Proc. Natl. Acad. Sci. USA, 1999. **96**: p. 3590.

12. Chiti, F., et al., *Solution conditions can promote formation of either amyloid protofilaments or native fibrils from the hypf n-terminal domain.* Prot.Sci., 2001. **10**: p. 2542.

13. Caughey, B. and P.T. Lansbury, *Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders.* Annu. Rev. Neurosci., 2003. **26**: p. 267-298.

14. Kelly, J., *Amyloid fibril formation and protein misassembly: A structural quest for insights into amyloid and prion diseases.* Structure, 1997. **5**: p. 595-600.

15. Rochet, J.C. and P.T. Lansbury, *Amyloid fibrilogenesis: Themes and variations.* Curr. Opin. Struct. Biol., 2000. **10**: p. 60.

16. Hou, L. and M.G. Zagorski, *Sorting out the driving forces for parallel and antiparallel alignment in the aβ peptide fibril structure.* Biophys J, 2004. **86**: p. 1-2.

17. Ma, B. and R. Nussinov, *Molecular dynamics simulations of alanine rich β-sheet oligomers: Insight into amyloid formation.* Prot. Sci., 2002. **11**: p. 2335-2350.

18. Nguyen, H.D. and C.K. Hall, *Molecular dynamics simulations of spontaneous fibril formation by rando-soil peptides.* Proc Natl Acad Sci USA, 2004. **101**(46): p. 16180-16185.

19. Frenkel, D. and B. Smit, *Understanding molecular simulation: From algorithms to applications*. 2nd ed. 1996, Chestnut Hill, MA: Academic Press.

20. Allen, M.P. and D.J. Tildesley, *Computer simulation of liquids*. 1987, New York: Oxford University Press.

21. Leach, A.R., *Molecular modelling: Principles and applications*. 2nd ed. 2001: Pearson Education Limited.

22. Alder, B.J. and T.E. Wainwright, *Studies in molecular dynamics, i: General method.* J Chem Phys, 1959. **31**: p. 459-466.

23. Smith, S.W., C.K. Hall, and B.D. Freeman, *Molecular dynamics for polymeric fluids using discontinuous potentials*. Journal of Computational Physics, 1997. **134**: p. 16.

24. Rapaport, D.C., *Molecular dynamics study of polymer chains.* J. Chem. Phys., 1979. **71**: p. 3299.

25. Rapaport, D.C., *Molecular dynamics simulation of polymer chains with excluded volume.* J. Phys. A, 1978. **11**: p. L213.

26. Bellemans, A., J. Orbans, and D.V. Belle, *Molecular dynamics of rigid and non-rigid necklaces of hard disks.* Mol. Phys., 1980. **39**: p. 781-782.

27. Andersen, H.C., *Molecular dynamics simulation at constant temperature and / or pressure.* J. Chem. Phys., 1980. **72**: p. 2384.

28. Brooks, B.R., et al., *Charmm: A program for macromolecular energy minimization and dynamics calculation.* J. Comp. Chem., 1983. **4**: p. 187.

29. Weiner, S.J., et al., *An all atom force field for simulations of proteins and nucleic acids.* J. Comp. Chem., 1986. **7**: p. 230.

30. Dauber-Osguthorpe, P., et al., *Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase trimethoprim, a drug receptor system.* Proteins: Struct. Funct. and Genet., 1988. **4**: p. 31.

31. Levitt, M., *Molecular dynamics of native protein: I. Computer simulation trajectories.* J. Mol. Bio., 1983. **168**: p. 595.

32. Zimmerman, S., et al., *Conformational analysis of the 20 naturally occurring amino acid residues using ecepp.* Macromolecules, 1977. **10**(1): p. 1-9.

33. Duan, Y. and P.A. Kollman, *Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution.* Science, 1998. **282**: p. 740.

34. Khurana, R., et al., *Partially-folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates.* Biochemistry, 2001. **40**: p. 3525.

35. Molecular Simulations, I., *Charmm principles*. 1999, San Diego: Inc., Molecular Simulations.

36. Gsponer, J., U. Haberthur, and A. Caflisch, *The role of side-chain interactions in the early steps of aggregation: Molecular dynamics simulations of an amyloid-forming peptide from the yeast prion sup35.* Proc Natl Acad Sci USA, 2003. **100**(9): p. 5154-5159.

37. Ferrara, P., J. Apostolakis, and A. Caflisch, *Evaluations of a fast implicit solvent model for molecular dynamics simulations.* Proteins: Struct. Funct. and Genet., 2002. **46**: p. 24-33.

38. Roux, B. and T. Simonson, *Implicit solvent models.* Biophysical Chemistry, 1999. **78**: p. 1-20.

39. Hasel, W., T. Hendrickson, and W. Still, *A rapid approximation to the solvent accessible surface areas of atoms.* Tetrahedron Computational Methodology, 1988. **1**: p. 103-116.

40. Ma, B. and R. Nussinov, *Stabilities and conformations of alzhemier's β-amyloid peptide oligomers (aβ16-22, aβ16-35, and aβ10-35):Sequence effects.* Proc Natl Acad Sci USA, 2002. **99**: p. 14126.

41. Haspel, N., et al., *A comparative study of amyloid fibril formation by residues 15-19 of the human calcitonin hormone: A single β-sheet model with a small hydrophobic core.* JMB, 2004. **345**: p. 1213-1227.

42. Petkova, A.T., et al., *A structural model for alzheimer's β-amyloid fibrils based on experimental constraints from solid state nmr.* Proc Natl Acad Sci USA, 2002. **99**: p. 16742.

43. Santini, S., N. Mousseau, and P. Derreumaux, *In silico assembly of alzheimer's aβ16-22 peptide into β-sheets.* J Am Chem Soc, 2004. **126**: p. 11509-11516.

44. Tarus, B., J.E. Straub, and D. Thirumalai, *Probing the initial stage of aggregation of the aβ10-35-protein: Assessing the propensity for peptide dimerization.* J Mol Biol, 2005. **345**: p. 1141-1156.

45. Blondelle, S.E., et al., *Polyalanine-based peptides as models for self-associated β-pleated-sheet complexes.* Biochemistry, 1997. **36**: p. 8393.

46. Forood, B., et al., *Structural characterization and 5'-mononucleotide binding of polyalanine β-sheet complexes.* J. Mol. Recognit., 1996. **9**: p. 488.

47. Williams, A.D., et al., *Mapping abeta amyloid fibril secondary structure using scanning proline mutagenesis.* J Mol Biol, 2004. **335**(3): p. 833-842.

48. Williams, A.D., S. Shivaprasad, and R. Wetzel, *Alanine scanning mutagenesis of abeta(1-40) amyloid fibril stability.* J Mol Biol, 2006. **357**(4): p. 1283-1294.

49. Smith, S.W., C.K. Hall, and B.D. Freeman, *Large scale molecular dynamics study of entangled hard-chain fluids.* Phys. Rev. Lett., 1995. **75**: p. 1316.

50. Smith, S.W., C.K. Hall, and B.D. Freeman, *Molecular dynamics study of transport coefficients for hard-chain fluids.* J Chem Phys, 1995. **102**: p. 1057-1073.

51. Takada, S., Z. Luthey-Schulten, and P.G. Wolynes, *Folding dynamics with non-additive forces: A simulation study of a designed helical protein and a random heteropolymer.* J. Chem. Phys., 1999. **110**: p. 11616.

52. Smith, A.V. and C.K. Hall, *Protein refolding versus aggregation: Computer simulations on an intermediate resolution model.* J. Mol. Biol., 2001. **312**: p. 187.

53. Smith, A.V. and C.K. Hall, *A-helix formation: Discontinuous molecular dynamics on an intermediate resolution model.* Protein: Structure, Function and Genetic, 2001. **44**: p. 344.

54. Smith, A.V. and C.K. Hall, *Assembly of a tetrameric α-helical bundle: Computer simulations on an intermediate-resolution protein model.* Proteins: Structure, Function and Genetics, 2001. **44**: p. 376.

55. Nguyen, H.D., A.J. Marchut, and C.K. Hall, *Solvent effects on the conformational transition of a model polyalanine peptide.* Prot. Sci., 2004. **13**(11): p. 2909-2924.

56. Ding, F., et al., *Mechanism for the alpha-helix to beta-hairpin transition.* Proteins: Struct. Funct. and Genet., 2003. **53**: p. 220-8.

57. Sugita, Y. and Y. Okamota, *Replica exchange molecular dynamics method for protein folding.* Chem. Phys. Letts., 1999. **314**: p. 141.

58. Nguyen, H.D. and C.K. Hall, *Phase diagrams describing fibrillization by polyalanine peptides.* Biophys J, 2004. **87**(6): p. 4122-4134.

59. Nguyen, H.D. and C.K. Hall, *Kinetics of fibril formation by polyalanine peptides.* J Biol Chem, 2004. **in press**.

60. Marchut, A.J. and C.K. Hall, *Spontaneous formation of annular structures observed in molecular dynamics simulations of polyglutamine peptides.* Comput Biol Chem, 2006. **30**(3): p. 215-8.

61. Wacker, J.L., et al., *Hsp70 and hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer.* Nat. Struct. Mol. Biol., 2004. **11**: p. 1215.

62. Perutz, M.F., et al., *Amyloid fibers are water-filled nanotubes.* Proc Natl Acad Sci USA, 2002. **99**: p. 5591.

63. Ding, F., et al., *Direct molecular dynamics observation of protein folding transition state ensemble.* Biophys J, 2002. **83**: p. 3525.

64. Ding, F., et al., *Molecular dynamics simulation of the sh3 domain aggregation suggests a generic amyloidogenesis mechanism.* J Mol Biol, 2002. **324**: p. 851.

65. Peng, S., et al., *Discrete molecular dynamics simulations of peptide aggregation.* Physical Review E, 2004. **69**: p. 041908-1-041908-7.

66. Coles, M., et al., Biochem, 1998. **37**: p. 11064.

67. Urbanc, B., et al., *In silico study of amyloid β-protein folding and oligomerization.* Proc Natl Acad Sci USA, 2004. **101**(50): p. 17345-17350.

68. Urbanc, B., et al., *Molecular dynamics simulation of amyloid β-dimer formation.* Biophys J, 2004. **87**: p. 1-12.

69. Bitan, G., S.S. Vollers, and D.B. Teplow, *Elucidation of primary structure elements controlling early amyloid beta-protein oligomerization.* J Biol Chem, 2003. **278**(37): p. 34882-34889.

70. Bitan, G., et al., *Amyloid beta-protein (abeta) assembly: Abeta 40 and abeta 42 oligomerize through distinct pathways.* Proc Natl Acad Sci USA, 2003. **100**: p. 330-335.

71. Petkova, A.T., et al., *Self-propagating, molecular-level polymorphism in alzheimer's β-amyloid fibrils.* Science, 2005. **307**: p. 262-265.

72. Inc., M.S. *Charmm principles*. [webpage] 1999 [cited 2005; Available from: www.sinica.edu.tw/~scimath/msi.

73. Marchut, A.J. and C.K. Hall, *Spontaneous formation of annular structures observed in molecular dynamics simulations of polyglutamine peptides*. Computational Biology and Chemistry, 2006. **in press**.

2.5 List of Tables

2.6 List of Figures

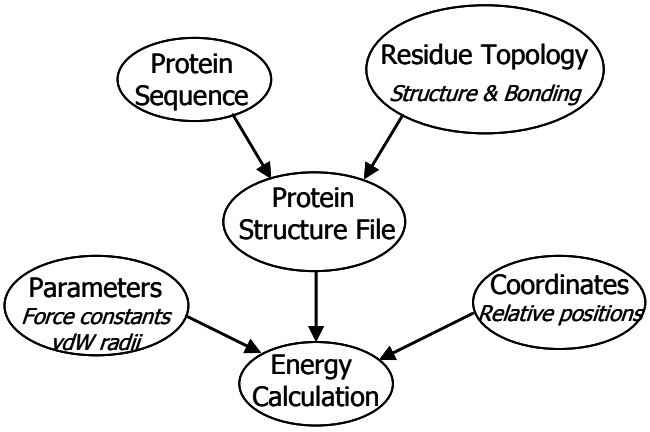

Figure 2. 1 Summary of the information required by CHARMM at the beginning of a simulation [72].

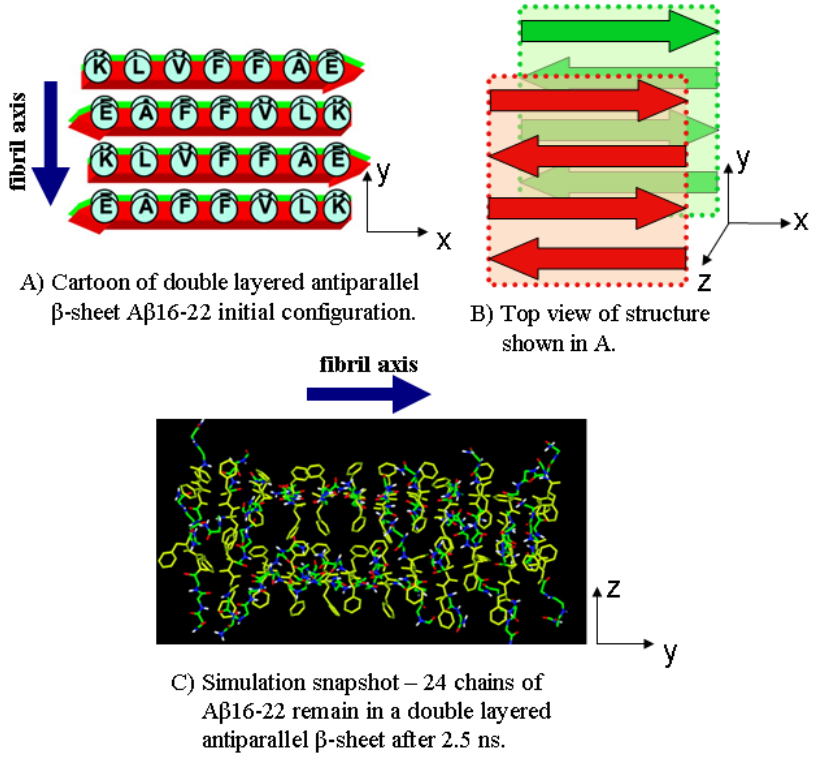

Figure 2. 2 Illustration of the double-layered β-sheet formed by Aβ16-22 and a snapshot from the CHARMM simulation [40].

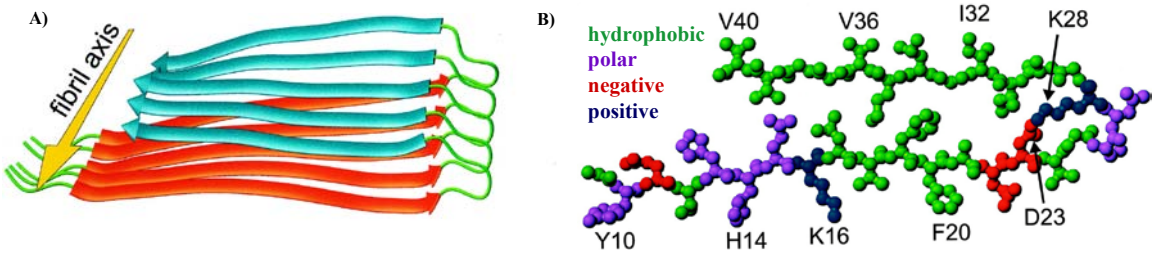

Figure 2. 3 A) Ribbon diagram depicting the arrangement of the 5 copies of Aβ(10-40). B) Aβ(10-40) as depicted by CHARMM all-atom simulations package with residues colored according to type. [42].

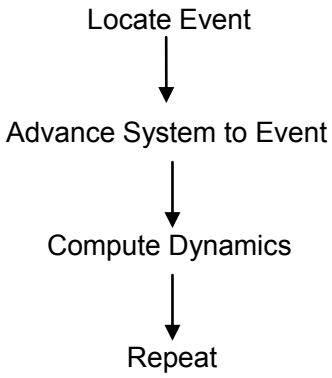

Figure 2. 4 A general procedure for DMD

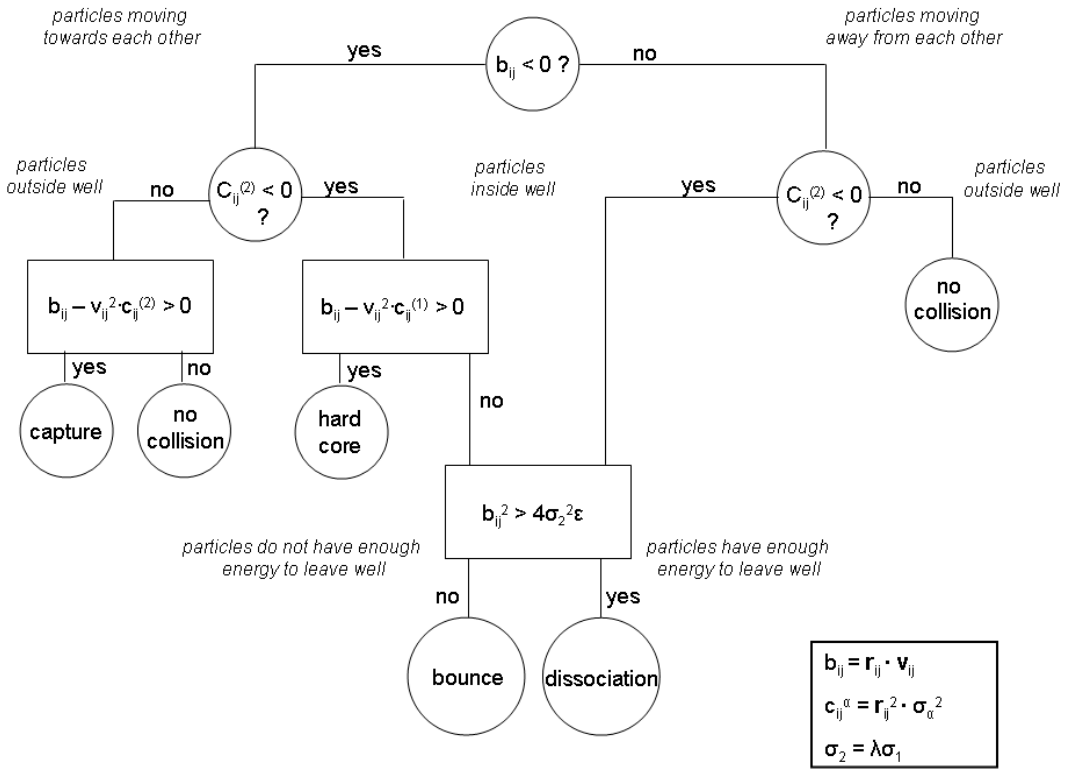

Figure 2. 5 Flowchart for square-well dynamics. Adapted from Alder and Wainwright (1959).

Table 2. 1 DMD Simulation Parameters

Table 1. Simulation Parameters

United atom diameters, σ	(Å)
NH	3.300
$C_\alpha H$	3.700
CO	4.000
Well diameters, λσ	(Å)
NH	4.200
CO	4.200
Bond lengths, *l*	(Å)
$N_i H– C_{\alpha,i} H$	1.460
$C_{\alpha,i} H – C_i O$	1.510
$C_i O – N_{i+1} H$	1.330
Pseudobond lengths, *l*	(Å)
$N_i H – C_i O$	2.45
$C_{\alpha,i} H– N_{i+1} H$	2.41
$C_i O – C_{\alpha,i+1} H$	2.45
$C_{\alpha,i} H – C_{\alpha,i+1} H$	3.80
Bond angles	(°)
$<N_i H – C_{\alpha,i} H – C_i O$	111.0
$<C_{\alpha,i} H – C_i O – N_{i+1} H$	116.0
$<C_i O– N_{i+1} H – C_{\alpha,i+1} H$	122.0

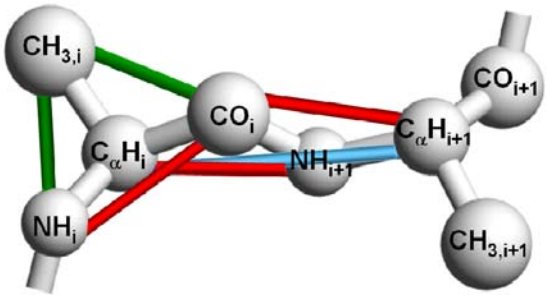

Figure 2. 6 Geometry of inter-mediate resolution protein model, PRIME, for alanine.

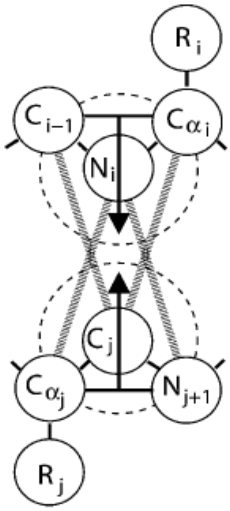

Figure 2. 7 Backbone hydrogen bonding where the dashed circle represents the attractive squarewell of N_i and C_j.

Table 2. 2 Auxiliary pair parameters for hydrogen bond potential.

Table 2. Parameter d_{ij}

Pairs	d_{ij} (Å)
$N_i - C_{\alpha j}$	5.00
$N_i - N_{j+1}$	4.74
$C_j - C_{\alpha i}$	4.86
$C_j - C_{i-1}$	4.83

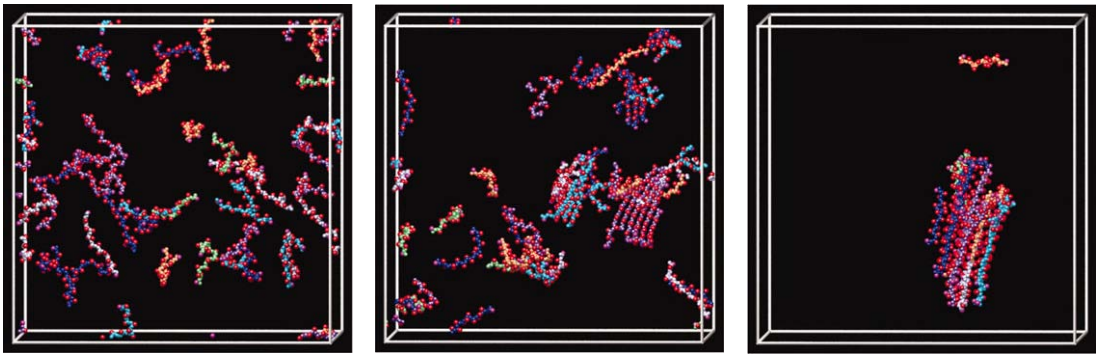

Figure 2. 8 Snapshots of 48 peptide system at various reduced times, t*. The simulation proceeds from a random initial configuration at concentration c=10mM and temperature T*=0.14 until the formation of a protofilament at t*=205.9 [18].

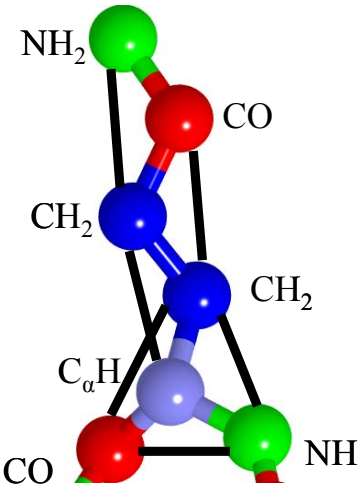

Figure 2. 9 Geometry of intermediate resolution protein model, PRIME, for glutamine.

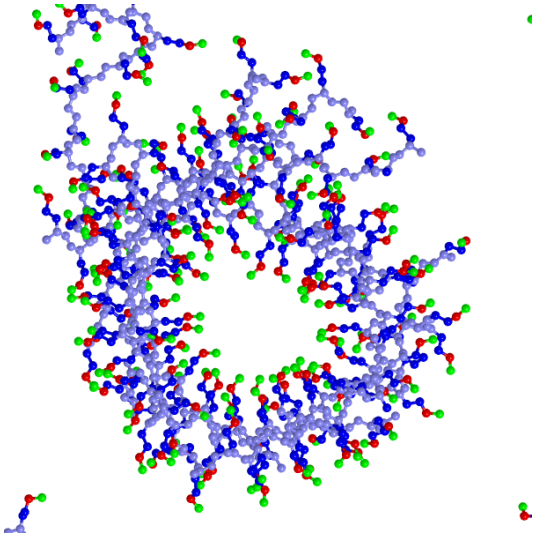

Figure 2. 10 Tube formed during simulation of 24 polyglutamine 16mers [73].

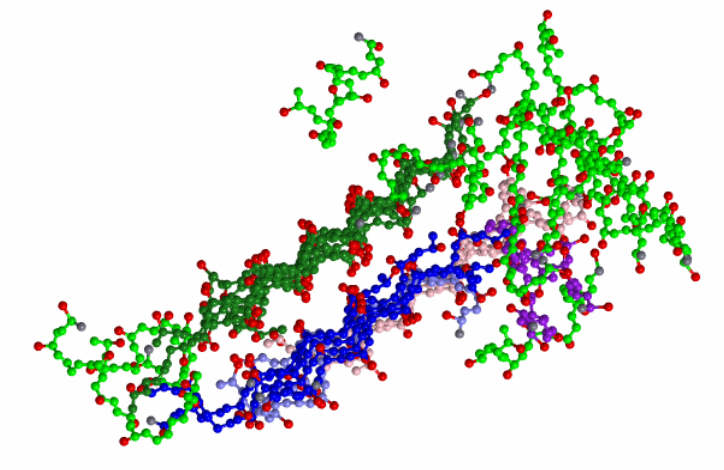

Figure 2. 11 Snapshot of a 48-peptide ordered aggregate obtained from the c=1mM simulation at T*=0.12 (Wagoner and Hall unpublished).

CHAPTER 3

Computer Simulation Study of Amyloid Fibril Formation by Palindromic Sequences in Prion Peptides

3. 1 Introduction

The transmissible spongiform encephalopathies, better known as prion diseases, are fatal neurodegenerative disorders characterized by the misfolding of a normal "prion" protein, PrP^{C} (cellular protein) into a beta-sheet rich conformation, PrP^{Sc} (scrapie protein) that subsequently aggregates to form amyloid structures. Examples include human Creutzfeldt-Jakob disease, scrapie disease in sheep and goats, kuru which is associated with cannibalism, and Mad Cow disease [1-5]. A common feature of the prion protein sequence across the various species is the presence of regions of short oligopeptide repeat sequences that are sometimes palindromic in nature and commonly between four and ten residues in length. These regions have high hydrophobic residue content and are believed to drive aggregation into fibrils [6-14], although the molecular details of this process are not well understood. The short palindromic prion fragments can also form amyloid fibrils in the absence of the flanking residues in the full-length prion protein, suggesting that study of these peptides may help elucidate the mechanism of fibril formation [6-7, 13, 15-16]. In this work, we perform computer simulations of the self-assembly of two

simple palindromic aliphatic sequences, one from the Syrian hamster prion protein, residues 113-120 (AGAAAAGA), and one from the mouse prion protein, residues 111-120 (VAGAAAAGAV), and variations of these two sequences. The goal of this work is to provide insight into the molecular motions that underlie the early steps in the assembly of these short palindromic sequences into amyloid fibrils.

Since it was first postulated in 1991 that the scrapie isoform of the prion protein was a key component in neuronal degeneration [17], interest in how this protein changes conformation and subsequently aggregates has surged. Early experimental studies focused on the polypeptide PrP 27-30, defined as residues 90-145 of PrP^{Sc}, the protein implicated in prion disease pathogenesis [17-19]. PrP 27-30 was isolated from scrapie-infected Syrian hamster brains and made synthetically. Nguyen et al. were interested in identifying the part of the sequence necessary to drive the conformation change from α-helix to β-strand, the α → β transition, and the three-dimensional organization of the β-sheets. They isolated PrP 109-122, PrP 113-120 (A8A) and PrP 90-145 and used electron microscopy and x-ray diffraction to determine which sequences self-assembled into fibrils. All three sequences formed β-sheets under certain conditions. In particular, they found that the octamer A8A formed fibrils composed of β-sheets staggered in the H-bonding direction but with a smaller inter-sheet spacing than had previously been reported for other amyloid fibrils [19]. Kirschner, Cohen, and Prusiner believed this was a result of the packing of the small side chains between the β-sheets [19-20].

Although, the studies by the Prusiner, Cohen and Kirschner labs answered many questions, we are left asking how the aggregation proceeds after the α → β transition, what the molecular arrangement of β-sheets in the fibrils is, and which part of the sequence (truncated or otherwise) is most directly associated with the disease. The hydrophobic core found in many disease-related amyloidogenic peptides surely plays a role –it is thought to drive their aggregation into fibrils [21]— but whether or not this part of the aggregation process is a source of neurotoxicity is still under discussion. Recent hydrogen-deuterium exchange studies on the full PrP fibril structure have confirmed the hypothesis that its hydrophobic core is a β-sheet rich region and hence is likely to play a role in fibril formation [16, 22]. Many researchers have investigated the role played by the hydrophobic palindromic sequence, PrP 113-120, (AGAAAAGA) which is common to all mammalian prion peptides (human, sheep, mouse, etc.), in their subsequent toxicity [6, 9, 23-24]. The PrP 106-126 region of the protein is known to be necessary for β-sheet formation. Jobling and coworkers showed that presence of MoPrP 106-126 caused a 50% decrease in the number of wild-type neuronal cells, but that by substituting hydrophilic residues for hydrophobic residues, the peptide's ability to form β-sheets was disrupted and the loss of neuronal cells was significantly reduced [6]. Interestingly Aceto and coworkers found that PrP 106-126 was completely insoluble in water and had a high binding affinity to Thioflavin T (ThT), indicating fibrillar formation but that when alanines were substituted for Gly114 and Gly119 , PrP 106-126 became soluble in water with dramatically decreased ThT binding, no change in its β-

sheet content and increased neurotoxicity. This indicated that the final fibril may not be the neurotoxic agent [24]. It is apparent that some structure along the PrP 106-126 fibril formation pathway is causing cell death and that understanding this pathway could shed light on this transient structure.

In recent years, molecular-level computer simulation has been used to supplement the information gained from experimental studies. The focus has been on understanding how prion peptides assemble into amyloid fibrils and predicting the fibril's molecular structure. Most of these simulations have been performed on truncated sequences like Syrian hamster prion SHaPrP 113-120 [7, 25], yeast prion Sup-35(GNNQQNY) [26-27], and mouse prion MoPrP 106-126 [16] because these short peptides are themselves amyloidogenic. The use of atomistic molecular simulations to study the arrangement of β-strands in fibrils was pioneered by Ma and Nussinov. In order to learn more about the nucleation events believed to initiate amyloid formation and growth they performed high temperature (up to 350K) molecular dynamics simulations on two short sequences, A8 (AAAAAAAA) and SHaPrP 113-120 (AGAAAAGA). Examination of various ordered starting configurations containing 3, 4, 6 and 8 peptides showed that as the number of peptides increased, the stability of the conformation increased. They explored how the arrangement of the peptides relative to each other within and between sheets affected conformational stability by simulating octamers composed of two stacked β-sheet layers with each layer composed of four A8 or SHaPrP 113-120 peptides. Their analysis was based on the assumption that an oligomer that does not disassociate at high temperature

must be the nucleus that "seeds" fibril formation. The most stable arrangement of eight peptides was an octamer with antiparallel β-strands within the β-sheet and parallel β-strands between the octamer layers. The A8 octamers were much more stable than the SHaPrP 113-120 octamers. [25] Roder and coworkers combined experiment and simulation to study MoPrP 106-126 which contains the palindromic sequence VAGAAAAGAV [16]. Their H/D exchange experiments on MoPrP 106-126 fibrils showed, as expected, that the hydrophobic core was protected and therefore likely played a role in the structure of the prion fibrils. In order to obtain a more detailed description of the molecular arrangement within the fibrils, they performed room temperature molecular dynamic simulations on six different arrangements of eight MoPrP 106-126 peptides arranged in a β-sheet. The most stable arrangement was two parallel β-sheet layers composed of parallel β-strands, which is different from the antiparallel β-strand within a β-sheet result of Ma and Nussinov for the hydrophobic sequence SHaPrP 113-120. Many researchers believe that the anti-parallel arrangement of peptides within β-sheets is favored for shorter sequences but the role of sequence in peptide orientation within β-sheets is still unclear [25, 28-29]. In an attempt to elucidate the arrangement of β-strands in PrP amyloid fibrils, Lee and coworkers used experimental data to guide their molecular dynamics simulations of SHaPrP 109-122. Their initial configuration was an octamer composed of two stacked parallel β-sheet layers with each layer composed of four antiparallel β-strands. This configuration remained stable over 2ns [7] and is the same as the stable conformation observed by Ma and Nussinov for SHaPrP 113-120.

One point of general interest is that all of the atomistic simulations agree that the stable fibrillar structure is composed of two stacked β-sheets.

In this work, we attempt to characterize the prion fibrillization pathway by simulating the spontaneous assembly of prion fibrils formed by the truncated prion sequences SHaPrP 113-120 (AGAAAAGA), MoPrP 111-120 (VAGAAAAGAV) and eight variations of these sequences including: GA_6G (a longer uninterrupted alanine stretch flanked by glycine), $(AG)_4$ (a complete disruption of hydrophobic residues), GAAAGAAA (a mimic of Aβ29-36), A8, VAVAAAAVAV (less flexible than MoPrP 111-120), GAVAAAAVAG (uninterrupted hydrophobic sequence), A10 and V10. We combine PRIME20, an intermediate-resolution description of the protein geometry for all twenty amino acids [30], with discontinuous molecular dynamics (DMD) [31-39]. Simulations are performed on 48-peptide homogenous systems containing the eight previously listed peptides starting from random configurations of random coils at high temperatures. The systems are cooled slowly to the desired simulation temperature and concentration, c = 5mM. We explored how changing sequence and temperature affects the pathway of aggregation by monitoring the formation of different structures such as α-helices, β-sheets (dimers, trimers, tetramers, pentamers, hexamers and larger), amorphous aggregates, and fibrils as a function of time.

Highlights of our results are the following. We are able to observe spontaneous fibril formation for most of the sequences studied but the propensity to form fibrils depends on the sequence and temperature. General trends observed were that as the chain length and the length of the stretch of hydrophobic residues increase, the ability to form fibrils increases. However as the hydrophobicity of the sequence increases, the ability to form well-ordered structures decreases. Thus, long hydrophobic sequences like VAVAAAAVAV and V10, form slightly disordered aggregates that are partially fibrillar and partially amorphous. Subtle changes in sequence result in slightly different fibril structures. For example, although the prion sequences, AGAAAAGA and VAGAAAAGAV, predominantly form fibrils with two stacked β-sheets, some sequences like A8, A10, GA_6G, and GAVAAAAVAG prefer three and even four sheet fibrils. The presence of more than two valines per sequence led to structures with both fibrillar and amorphous content, at least for the conditions explored in this paper. The strands show no preference for parallel or antiparallel arrangements (either within the sheets or between the sheets) until the chain length exceeds eight residues, at which point a parallel arrangement for these simple aliphatic sequences seems to be preferred. Also a change in sequence does seem to affect the intrastrand distance. Fibril growth is also affected by sequence. When valine is present fibrils grow by first forming amorphous aggregates and then β-sheets, but when valine is absent fibrils grow by forming β-sheets that then associate into fibrils. For the sequences containing valine (the largest side chain considered) once amorphous aggregates form they are very slow to diminish in size, even

after β-sheet formation and fibril association begins. This suggests that there is a competition between the formation of amorphous aggregates and β-sheets in the aggregation of the peptides with valine that may reduce their ability to form ordered aggregates.

This paper is organized as follows. In the next section, we describe the peptide model and the simulation method. In the following section, we present the results obtained from simulations of multi-peptide systems at various conditions. The last section is a discussion of our results.

3. 2 Methods

Model Peptide and Forces

The truncated prion peptides sequences considered here are: SHaPrP 113-120, AGAAAAGA and MoPrP 111-120, VAGAAAAGAV, and eight variations, GAAAAAAG, AGAGAGAG, GAAAGAAA, A8, VAVAAAAVAV, GAVAAAAVAG, A10 and V10. Our rationale for choosing the eight variations was the following. Our interest in short peptides not interrupted by glycine like A8 and A10 stemmed from previous theoretical [15] and experimental studies [11, 40-41] indicating these sequences easily formed amyloid fibrils. The sequence GAAAGAAA is a mimic of a segment of the Aβ peptide, Aβ29-36, GAIIGLMV, but is slightly more hydrophobic than GAAAGAAA. Poly-valine stretches are known to aggregate and are present in lung

surfactant C protein which can, in certain diseases, accumulate in the lung and inhibit normal respiration [42-44]. The remaining four peptides were either inversions of the placement of the glycine residues relative to the parent sequences like GAAAAAAG and GAVAAAAVAG or the addition (AG_4) or subtraction (VAVAAAAVAV) of glycine from the sequence to help ascertain how the flexibility imparted by glycine impacts the aggregation process.

In this work we apply a new implicit-solvent force field PRIME20 to the description of the geometry and energetics of the aliphatic heteropeptide sequences in the prion protein. PRIME20 was recently introduced by Cheon et al. [30] as an extension of PRIME, an implicit solvent intermediate-resolution protein model previously used in simulations of the aggregation of polyalanine and polyglutamine. PRIME was originally developed by Smith and Hall [31-32] and later improved by Nguyen et al. [36]. More recently the PRIME model was extended to the study of polyglutamine peptides [35] illustrating its versatility. In PRIME, the protein backbone is represented by three united atom spheres, one for the amide group (NH), one for the carbonyl group (CO), and one for the alpha-carbon and its hydrogen ($C_\alpha H$). In the original version of PRIME, each side chain was represented by a single sphere for polyalanine and by a chain of four spheres for polyglutamine. In PRIME20 the twenty possible side chains are modeled as single spheres of unique size, atomic mass and C_α—R bond length. In the truncated prion peptides considered here, the only side chains that appear are alanine, valine and glycine.

All backbone bond lengths and angles are set to their ideal values. In order to maintain the *trans*-configuration we fix the consecutive C_α—C_α distance. The side chains are positioned relative to the protein backbone so that all residues are L-isomers. The solvent molecules in our system are modeled implicitly.

All forces between the united atom spheres are modeled with discontinuous potentials, e.g. hard-sphere and square-well interactions. The excluded volume of each of the peptide's four united atom spheres is modeled using a hard sphere interaction. The covalent bond lengths are maintained using a hard sphere interaction that prevents them from moving outside of the range $(1+\delta)l$ to $(1-\delta)l$, where l is the ideal bond length and δ is the tolerance, which is set at 2.375% [45]. Ideal backbone bond angles, C_α—C_α distance, and residue L-isomerization are maintained by imposing a series of pseudobonds whose lengths are also allowed to fluctuate by 2.375%.

In addition to preserving proper protein geometry, we represent one of the most important energetic contributions to protein folding and aggregation- which is hydrogen bonding. For the simple palindromic aliphatic sequences studied in this paper, hydrogen bonding is represented as a square well attraction of depth ε_{HB} and width 4.5Å between the backbone amide and carbonyl groups. Hydrogen bonds are anisotropic in nature so we must constrain their formation to only occur when the NH united atom vector and the CO united atom vector point towards each other and the angle between those vectors is restricted between 120° and 180° [31]. In order to accomplish this we have set the

following criteria. A hydrogen bond only forms when: 1) the virtual hydrogen and oxygen atoms (whose location can be calculated at any time) are separated by a distance of 4.5Å, 2) the nitrogen-hydrogen and carbon-oxygen vectors point toward each other within a fairly generous tolerance, 3) neither the NH nor the CO are already involved in a hydrogen bond with a different partner, and 4) the NH and CO are separated by at least three intervening residues along the chain. Further details on the hydrogen bonding model can be found in our earlier work. [34, 36, 38]

The interactions in PRIME20 are all modeled as square well interactions between the spherical units on each amino acid with strength (well depth) and range determined individually for each pair. Since solvent is modeled implicitly these are all effective interactions or potentials of mean force. In PRIME20, the energy parameters that describe the side chain/side chain interactions, the hydrogen bonding interactions between backbone NH and CO and between side chain and side chain are derived in the following way. Briefly, the twenty possible amino acids are classified into 14 groups: [LVI] [F] [Y] [W[[M[[A] [C] [ED] [KR] [P] [ST] [NQ] [H] [G], according to their side chain size, charge, potential for disulfide bond formation, hydrophobicity, and possibility of side chain hydrogen bonding. The matrix of energy parameters between the side chains in the different groups was found by Cheon et al. [30] using a perceptron- learning algorithm and a modified stochastic learning algorithm that optimizes the energy gap between 711 known native states from the PDB and decoy structures generated by gapless threading. The number of independent pair-interaction parameters is chosen to

be small enough to be physically meaningful yet large enough to give reasonably accurate results in discriminating decoys from native structures. A total of nineteen interaction parameters were used to describe the side chain energetics. In this work we focus on the amino acids: glycine, alanine and valine. Each is modeled using a three sphere backbone (for NH, CO and $C_\alpha H$) plus a single sphere side chain; the side chain for glycine is quite small. A summary of the geometric and energetic parameters for glycine, alanine and valine is provided in Table 3.1.

The system temperature is scaled by the hydrogen bonding energy between the backbone NH and CO, ε_{HB}, so that the reduced temperature is $T^* = k_B T/\varepsilon_{HB}$.

Discontinuous Molecular Dynamics

Discontinuous molecular dynamics (DMD) is a variant on standard molecular dynamics that is applicable to systems of molecules interacting via discontinuous potentials (e.g., hard sphere and square-well potentials). Unlike soft potentials such as the Lennard-Jones potential, discontinuous potentials exert forces only when particles collide, enabling the exact (as opposed to numerical) solution of the collision dynamics. This imparts great speed to the algorithm, allowing sampling of longer time scales and larger systems than traditional molecular dynamics. The particle trajectories are followed by analytically integrating Newton's equations of motion, locating the time between collisions and then advancing the simulation to the next collision (event) [46-47]. DMD on chain-like molecules is generally implemented using the "bead string" algorithm

introduced by Rapaport [48-49] and later modified by Bellemans et al. [45]. Chains of square-well spheres can be accommodated in this algorithm by introducing well-capture, well-bounce, and well-dissociation "collisions" when a sphere enters, attempts to leave, or leaves the square well of another sphere. In this paper, DMD simulations are performed in the canonical ensemble (NVT) with the initial velocities chosen randomly from a Maxwell-Boltzmann distribution about the desired system temperature. The initial positions of the particles or spheres are chosen randomly while still ensuring that no geometrical constraints are violated.

In DMD simulations of protein aggregation, we determine the number of particles in the system by specifying the concentration which is given by $c = N/L^3$, where N is the number of molecules in the box and L is the simulation box length. Periodic boundary conditions are employed. Since we are simulating large systems at high concentrations in random initial configuration the box length, L, must be large enough to prevent the macromolecules from interacting with themselves but still allow them to interact with their periodic image. We set L = 252Å in this study. The simulation proceeds according to the following schedule: identify the first event (e.g., a collision), move forward in time until that event occurs, calculate new velocities for the pair of spheres involved in the event and calculate any changes in system energy resulting from hydrogen bond events or hydrophobic interactions, find the second event, and so on. Types of events include excluded volume events, bond events, and square-well hydrogen

bond and hydrophobic interaction events. An excluded volume event occurs when the surfaces of two hard spheres collide and repel each other. A bond (or pseudobond) event occurs when two adjacent spheres attempt to move outside of their assigned bond length and the two particles feel an infinite repulsion that forces them back into their assigned bond length. Square-well events include well-capture, well-bounce, and well-dissociation ‘‘collisions’’ when a sphere enters, attempts to leave, or leaves the square well of another sphere. For more details on DMD simulations with square-well potentials, see articles by Alder and Wainwright [46] and Smith et al. [47].

A total of ten sequences are studied in this work; all contain 48 peptides at concentrations c = 5mM. Each simulation is started at high temperature to ensure a random initial configuration and then slow-cooled to the temperature of interest to minimize kinetic trapping. Slow-cooling is achieved by decreasing the temperature in discrete steps starting from a high temperature until the desired simulation temperature is reached. The simulation temperature is maintained using the Andersen thermostat [50]; in this method all the particles undergo random infrequent “events” or “collisions” with a ghost particle that reassigns the particle’s velocity randomly from a Maxwell-Boltzmann distribution centered at the simulation temperature. Five simulations are run for each sequence at the given temperature and concentration (state). Error bars are taken to be the standard deviation at each state. All simulations are run for an average of 50-100 billion collisions, depending on simulation conditions, sequence, temperature and concentration. We can perform 500 million collisions per hour on a 2.4 GHz AMD Opteron™

workstation, which means that our simulations require approximately a 200 hours.

In this study, we monitor and analyze the formation of β-strands, β-sheets, amorphous aggregates and fibrils. We also check to see if the β-strands in a β-sheet are arranged in a parallel or anti-parallel configuration. The criteria for assigning the types of structures formed are the following. If each peptide in a group of peptides has at least two inter-peptide hydrogen bonds or hydrophobic interactions with a neighboring peptide in the same group, then that group is classified as an aggregate. Aggregates can be either ordered or amorphous. If an aggregate contains β-sheets or fibrils, we classify it as an ordered aggregate. If each peptide in a group of peptides has at least n/2, where n = chain length, inter-peptide β-hydrogen bonds to a particular neighboring peptide in the group, we classify this group as a β-sheet. (A β-hydrogen bond is a hydrogen bond between two residues whose backbone angles are in the β-region of the Ramachandran plot.) If at least two β-sheet structures form inter-sheet hydrophobic interactions (at least four hydrophobic interactions per peptide per β-sheet), we classify this as a fibril; otherwise, we classify this and isolated β-sheets as non-fibrillar β-sheet structures. If an aggregate does not contain β-sheets but the peptides in the aggregate have any hydrophobic contacts, then the aggregate is considered amorphous. If an aggregate contains peptides with less than n/2 inter-peptide β-hydrogen bonds between neighboring chains, then this is also considered to be an amorphous aggregate.

3. 3 Results

DMD simulations were performed on 48-peptide systems containing SHaPrP 113-120 (AGAAAAGA), MoPrP 111-120 (VAGAAAAGAV) and eight variations on these sequences: $(AG)_4$, GA_6G, GAAAGAAA, A8, GAVAAAAVAG, VAVAAAAVAV, A10, V10 to determine if they form ordered structures (fibrils), what types of intermediate structures (β-sheets and/or amorphous aggregates) occur along the fibrillization pathway, and how changes in sequence influence the ability to form fibrils.

Ordered vs. Disordered Structures

Table 3.2 is a summary of the types of structures formed by each sequence at temperatures below the random coil temperature but above temperatures where only amorphous aggregates would be seen. Since the variations in sequence considered here are small the optimum temperature for fibril formation differs only slightly from peptide to peptide, not enough to indicate a trend. The optimum temperature range for fibril formation of the SHaPrP 113-120 and similar variations is between $0.14 \leq T^* \leq 0.16$. The longer sequences related to MoPrP 111-120 have a slightly higher optimum temperature range for fibril formation between $0.16 \leq T^* \leq 0.18$. At temperatures above the upper limit for each sequence the peptides primarily exist as free monomers. Table 3.2 is organized so that the Syrian hamster prion AGAAAAGA and its variations are listed on the top half and the mouse prion VAGAAAAGAV and its variations are listed

in the bottom half. Within each half, the sequences are listed from least likely to most likely to form ordered structures.

The percentage of fibrillar content in Table 3.2 for each sequence correlates with both the length of the hydrophobic stretch as well as the overall chain length of the peptide. We begin our discussion of this by first focusing on the sequences in the top half of the table. Sequences with a run of less than four hydrophobic residues ($(AG)_4$ and GAAAGAAA) were either unable to form fibrils or formed a small amount of fibrils during our simulations. Two possible reasons that $(AG)_4$ it is not a strong fibril former are, (1) that there are not enough side chains available to energetically drive the stacking of the β-sheets (fibrillization), and (2) that the increased flexibility provided by the glycine residues make it less likely that the peptides will zip up into multi-mer structures. It is harder to understand why the sequence GAAAGAAA is a weak fibril former (although it readily forms β-sheets) because it is a mimic of Aβ 29-36 (GAIIGLMV) which is believed to aggregate into amyloid fibrils. Two possible reasons for this are: (1) the simulation has not reached a stable equilibrium and that the β-sheets would aggregate into amyloid fibrils if given enough time, and (2) the reduced hydrophobicity of GAAAGAAA compared to GAIIGLMV could delay fibrillization or the overall ability to form fibrils. For AGAAAAGA at least half of the peptides were in fibrils with the remaining peptides were in β-sheets indicating an ordered structure. It is interesting to note that just the addition of a single alanine residue to the core of the sequence (compare AGAAAAGA to GAAAGAAA) leads to a higher percentage of peptides in fibrils and a

faster fibril formation time. In fact, as the hydrophobic stretch increases from four (AGAAAAGA) to six (GAAAAAAG) residues the percentage of peptides in a fibril increases. As the number of hydrophobic residues increases to eight (A8) the sequence continued to form fibrils but there were still many free monomer chains available, indicating that the system may not have yet reached a stable equilibrium.

We next consider the trends displayed in the second half of Table3.2 MoPrP 111-120 peptides (VAGAAAAGAV) and its variations. MoPrP 111-120 (VAGAAAAGAV) is very similar to SHaPrP 113-120 (AGAAAAGA) but has two additional flanking valine residues. Although VAGAAAAGAV formed fibrils, it also formed amorphous aggregates, suggesting that the flanking valine residues destabilize the highly ordered hydrophobic core but are strong enough to still cause peptide association. Moving down the table to GAVAAAAVAG, (where the glycine and valine residues in VAGAAAAGAV have been exchanged) we observed an increase in fibrillar content and a decrease in amorphous aggregates. In fact, GAVAAAAVAG formed the most-well-ordered structures of all the peptides simulated. Moving down the table further to VAVAAAAVA (where glycine residues in GAVAAAAVAG have been replaced by valines) resulted in an overall decrease in order; in fact, the peptides were as likely to be in amorphous aggregates as β-sheets or fibrils. In this case the freedom of the peptide backbone to adopt certain types of conformations has been reduced due to the increased bulkiness and hydrophobicity of the side chains.

Figure 3.1 shows snapshots of the simulation of GAVAAAAVAG at T*=0.17 and VAVAAAAVAV at T*=0.17 illustrating the various degrees of order. For ease of viewing we have portrayed the backbone as ribbons and the side chains in ball and stick formats. It is easy to see that GAVAAAAVAG readily forms β-sheets that then stack to form fibrils whereas the simple substitution of valine for glycine on the end caps of the sequence causes the increased formation of amorphous aggregates with some β-sheet content. It is possible that the structure formed by VAVAAAAVAV is only transient on the pathway to fibril formation or perhaps we are witnessing the small spherical oligomers believed to be off pathway of the fibrillization process. What is clear here is that simple substitutions in primary sequence can result in changes in structure not readily observable in experiment.

Molecular arrangement of Ordered Structures

Table 3.3 summarizes the molecular arrangements observed for those sequences that were able to form well-ordered structures. The table lists the number of fibrils produced during the simulation, the number of β-sheets per fibril, the number of β-strands per β-sheet, the ratio of parallel to anti-parallel strands within each β-sheet, and the average intrastrand distance between peptides within a β-sheet. This type of information can be difficult to ascertain from experiment alone. Given the small system size considered here, 48 peptides, it is not surprising that only one fibril is formed. Recent simulations in our group have indicated that tripling the system size to 192 chains allows

observation of multiple fibrils. For the purposes of the study, we are interested in the arrangement of a single fibril. As we increase the hydrophobic stretch of residues we see the number of beta sheets in a fibril increase from two to three. In fact only two sequences prefer stacks of two β-sheets, the SHaPrP 113-120 and MoPrP 111-120, which is consistent with recent experimental and simulation studies of these two peptides. We also calculated the average number of strands per sheet for each sequence. In general, if fewer sheets/fibril are present, then the higher the number of strands per sheet.

There has been significant debate over how the peptide strands arrange within a β-sheet. Do short strands show a preference for anti-parallel versus parallel? We measured the number of strands either arranged in parallel or anti-parallel and then calculated the (parallel:anti-parallel) ratio. What we see for the shorter sequences related to SHaPrP 113-120 is an equal preference for parallel and anti-parallel arrangements; this is likely due to the amphiphillic and palindromic nature of the sequences, i.e. there is no physical reason why these short, un-charged peptides in a "neutral" environment should prefer either orientation. As the chain length increases we see a slight shift towards preference for parallel arrangement of strands within the sheet, particularly for sequences with longer stretches of uninterrupted hydrophobic residues.

The distance between the β-strands within each sheet, the intra-strand distance, is of interest because it can be measured using X-ray diffraction. For the short peptides containing only alanine and glycine we observe distances of approximately 4.8Å which is slightly higher than the expected 4.7 Å observed for the cross-β diffraction pattern of the

model cross-β x-ray amyloid structure. For sequences containing valine, the intra-strand distance increases slightly, again suggesting that at least for the MoPrP 111-120 sequence there is still a lot of fluctuation in the backbone of the peptide. This may be due to the disturbance posed by the flanking valine residues which disrupt the tight close-packed β-sheets observed experimentally.

For the sequences that form ordered structures, such as SHaPrP 113-120, the arrangement of side chains at the interface the β-sheets is of particular interest. Snapshots from our simulations of SHaPrP 113-120 (AGAAAAGA) have enabled us to determine the side chain arrangement for these peptides. In Figure 3.2, we show a simulation snapshot of the fibril formed by AGAAAAGA including a closeup of the fibril that shows how the strands of one sheet alternate or stagger relative to the strands in the neighboring sheet. This staggered conformation is consistent with experimental data. Figure 3.3 shows a rotated version of the zoomed simulation snapshot; the side chains on the peptide form a steric zipper between the β-sheets, holding the sheets together and forming the cross-β structure. The backbone has been depicted as a ribbon for ease of viewing. The alanine side chains are shown in red. Due to the palindromic nature of the sequence, the four center alanine residues stick up and down from the plane of the peptide backbone to interact with the neighboring sheets alanine side chains. We verified this structure by calculating the phi-psi angles for each residue. The phi-psi angles were primarily in the β-region of the Ramachandran plot, with the degree of flexibility expected for the presence of glycine (data not shown). The fibril formed in our

simulations of the AGAAAAGA peptide are generally consistent with fibrils formed experimentally [19] and predicted using molecular dynamics [15]. To our knowledge this is the first time that the spontaneous formation of AGAAAAGA fibrils has been observed in a simulation.

Kinetics along the Fibril Formation Pathway

Early events in the fibril growth pathway are difficult to identify because the soluble disordered intermediate species that populate this pathway are difficult to isolate with current experimental techniques. One of the great benefits of computer simulations is the ability to monitor the various steps along the fibril formation pathway from an initial configuration of random coils to their final fibrillar structure. Figure 3.4 provides us with a detailed picture of how AGAAAAGA self-assembles into fibrils. Figures 3.5 and 3.6 show population data on the number of monomers, β-sheets, amorphous aggregates and fibrils in our simulation for the peptides AGAAAAGA and VAGAAAAGAV, respectively. At t=0, all of the peptides are free monomers (blue curve). Figure 3.4 shows that at t = 400, small β-sheets are starting to form; this corresponds to the red curve in Figure 3.5. Very few if any amorphous aggregates form before β-sheet formation occurs. Figure 3.4 shows that at t =1000, fibrils begin to appear; this corresponds to the purple curve in Figure 3.5. The fraction of β-sheets starts to diminish as they associate to form fibrils. At the end of the simulation in Figure 3.4 at t=1700, we see that all the free β-sheets are now in a fibril structure.

Comparing and contrasting the kinetic pathways toward the more-ordered AGAAAAGA and less-ordered VAGAAAAGAV structures may help us to understand the why the structures are different and how sequence influences that process. The number of monomers of AGAAAAGA (blue line) shown in Figure 3.5, slowly decrease by associating into dimers (data not shown) then into larger β-sheets (red line), and eventually fibrils at t=1000 (purple line). The assembly mechanism for the VAGAAAAGAV peptides shown in Figure 3.6, is slightly different. The number of monomers of VAGAAAAGAV (blue line) shown in Figure 3.6 decreases in favor of association into amorphous aggregates (green line) at t=250. This could be because the valine end groups are pulling the peptides together but the energy is not strong enough to overcome the entropic loss associated with the formation of β-sheets. As more amorphous aggregates appear, however, the energy of the system becomes large enough to overcome the entropic barrier and the more ordered β-sheets start to form at t=500 (red line). However, the competition between β-sheets and amorphous aggregates never abates even after the beginning of fibril formation (purple line) at t=750. This is why VAGAAAAGAV can in fact make fibrils but to a lesser degree than other sequences simulated in this paper.

As mentioned earlier, GAVAAAAVAG was one of the strongest fibril forming sequences simulated. For this sequence, the first non-monomer species to appear are the amorphous aggregates and then, after a long lag time, β-sheets start to form. The difference between the kinetics of VAGAAAAGAV and GAVAAAAVAG is that as

soon as GAVAAAAVAG begins to form β-sheets, the amorphous aggregates drop off, whereas for VAGAAAAGAV in Figure 3.6 the amorphous aggregates never completely disappear. A short time after β-sheets start to form for GAVAAAAVAG, fibrils start to grow and then dominate the rest of the kinetic pathway (data not shown) until almost 80% of the peptides in the system are in a fibril, (see Table 3.2).

3.4 Discussion

Spontaneous fibril formation was observed in discontinuous molecular dynamics simulations of 48 peptide systems containing various sequences related to the short palindromic prion peptides, SHaPrP 113-120 and MoPrP 111-120 starting from a configuration of random coils.. This was accomplished using our new implicit-solvent forcefield, PRIME20. The ten different sequences (AGAAAAGA, GAAAGAAA, $(AG)_4$, GAAAGAAA, A8, A10, V10, VAGAAAAGAV, GAVAAAAVAG and VAVAAAAVAV) formed structures with varying degrees of order, depending on their chain length and hydrophobicity over a range of temperatures. At higher temperatures, the peptides remained as random coils and at lower temperatures the peptides just "globed" together into disordered structures because of their short chain lengths.

We were able to follow the kinetic pathway of fibril formation for the two prion sequences. We observed slightly different fibrillization pathways for each sequence. SHaPrP 113-120 first forms small β-sheets that grow by lateral addition and then

associate into fibrils. MoPrP 111-120 forms amorphous aggregates and then after a lag time begins to form β-sheets; the amorphous aggregates and β-sheets compete with one another throughout the simulation until fibrils begin to form. We believe that the persistent presence of the amorphous aggregates for the mouse prion peptide prevents it from forming fibrils that are as nicely structured as the Syrian hamster prion peptide and the GAVAAAAVAG sequence.

We were also able to characterize the structures of the fibrils formed by each sequence. Most of the sequences that were able to form fibrils, ended up with 2-3 sheets. The shorter sequences showed no preference over anti-parallel versus parallel arrangement. We observed an intrastrand distance between 4.7-5Å which is slightly higher than 4.7Å predicted for the cross-β x-ray model. Currently, we do not have a systematic manner to calculate the intersheet distance but a rough estimate from the molecular structure file suggested an intersheet distance of ~9Å which is consistent with earlier estimates [19]. We were also able to observe the steric zipper pattern of the side chains between the β-sheets in the fibril. The differences observed between the very similar Syrian hamster and mouse prion peptides (and their variants) lends further support to the importance of peptide sequence in the fibrillization process.

It is important to point out that although PRIME20 is a very sophisticated intermediate-resolution model, some refinements may need to be considered before we can make definitive conclusions regarding its ability to capture the physics of valine. In

fact, additional simulations are underway to determine how well PRIME20 accounts for the subtle differences between the twenty amino acid residues especially in regards to the essential physics of fibril formation.

3.5 Acknowledgements

The authors are grateful to Ms. Erin Phelps whose kinetic model of aggregation helped to clarify the fibrillization processes of the sequences studied in this paper, Dr. Mookyung Cheong for his development of PRIME20, and Mr. Johnny Maury-Evertsz for his help with visualization software. All of the simulation snapshots in this paper were generated using Visual Molecular Dynamics developed at the University of Illinois Urbana-Champaign.

3.6 References

1. Kelly, J.W., *The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways.* Curr. Opin. Struct. Biol, 1998. **8**: p. 101.

2. Prusiner, S.B., *Novel proteinaceous infectious particles cause scrapie.* Science, 1982. **216**: p. 136-44.

3. Prusiner, S.B., *Prion diseases and the bse crisis.* Science, 1997. **278**: p. 245-251.

4. Prusiner, S.B., *Prions.* Proc Natl Acad Sci USA, 1998. **95**: p. 13363-13383.

5. Clarke, A.R., G.S. Jackson, and J. Collinge, *The molecular biology of prion propagation.* Phil. Trans. R. Soc. Lond. B, 2001. **356**: p. 185-195.

6. Jobling, M.F., et al., *The hydrophobic core sequence modulates the neurotoxic and secondary structure properties of the prion peptide 106-126.* J Neurochemistry, 1999. **73**(4): p. 1557-1565.

7. Lee, S.-W., et al., *Steric zipper of the amyloid fibrils formed by residues 109-122 of the syrian hamster prion protein.* Journal of Molecular Biology, 2008. **378**: p. 1142-1154.

8. Grosset, A., et al., *Rapid presymptomatic detection of prpsc via conformationally responsive palindromic prp peptides.* Peptides (New York, NY, United States), 2005. **26**(11): p. 2193-2200.

9. Gasset, M., et al., *Predicted alpha-helical regions of the prion protein when synthesized as peptides form amyloid.* Proc. Natl. Acad. Sci. USA., 1992. **89**: p. 10940.

10. Ziegler, J., et al., *Putative aggregation initiation sites in prion protein.* FEBS Lett., 2006. **580**: p. 2033-2040.

11. Shinchuk, L.M., et al., *Poly-(l-alanine) expansions form core beta-sheets that nucleate amyloid assembly.* Proteins: Structure, Function and Bioinformatics, 2005. **61**: p. 579-589.

12. Tessier, P.M. and S.L. Lindquist, *Unraveling infectious structures, strain variants and species barriers for the yeast prion [psi+].* Nature Structural & Molecular Biology, 2009. **16**(6): p. 598-605.

13. Walsh, P., K. Simonetti, and S. Sharpe, *Core structure of amyloid firbrils formed by residues 106-126 of the human prion protein.* Structure, 2009. **17**: p. 417-426.

14. Walsh, P., et al., *Morphology and secondary structure of stable beta-oligomers fromed by amyloid peptide prp(106-126).* Biochem, 2009. **48**: p. 5779-5781.

15. Ma, B. and R. Nussinov, *Molecular dynamics simulations of alanine rich b-sheet oligomers: Insight into amyloid formation.* Prot. Sci., 2002. **11**: p. 2335-2350.

16. Kuwata, K., et al., *Nmr-detected hydrogen exchange and molecular dynamics simulations provide structural insight into fibril formation of prion protein fragment 106-126.* Proc Natl Acad Sci USA, 2003. **100**(25): p. 14790-14795.

17. Prusiner, S.B., *Molecular biology of prion diseases.* Science, 1991. **252**(5012): p. 1515-22.

18. Inouye, H. and D.A. Kirschner, *Polypeptide chain folding in the hydrophobic core of hamster scrapie prion: Analysis by x-ray diffraction.* J Struct Biol, 1998. **122**(1-2): p. 247-55.

19. Nguyen, J.T., et al., *X-ray diffraction of scrapie prion rods and prp peptides.* J Mol Biol, 1995. **252**: p. 412-422.

20. Shinchuk, L.M., et al., *Poly-(l-alanine) expansions form core beta-shhets that nucleate amyloid assembly.* Proteins: Struct. Funct. and Genet., 2005. **61**: p. 579-589.

21. Teplow, D.B., *Structural and kinetic features of amyloid b-protein fibrilogenesis.* Int. J. Exp. Clin. Invest., 1998. **5**: p. 121.

22. Lu, X., P.L. Wintrode, and W.K. Surewicz, *Beta-sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange.* Proc Natl Acad Sci U S A, 2007. **104**(5): p. 1510-5.

23. Kupfer, L., W. Hinrichs, and M.H. Groschup, *Prion protein misfolding.* Current Molecular Medicine, 2009. **9**: p. 826-835.

24. Florio, T., et al., *Contribution of two conserved glycine residues to fibrillogenesis of the 106-126 prion protein fragment. Evidence that the soluble variant of the 106-126 peptide is neurotoxic.* Journal of Neurochemistry, 2003. **85**: p. 62-72.

25. Ma, B. and R. Nussinov, *Stabilities and conformations of alzhemier's b-amyloid peptide oligomers (ab16-22, ab16-35, and ab10-35):Sequence effects.* Proc Natl Acad Sci USA, 2002. **99**: p. 14126.

26. Zheng, J., et al., *Structural stability and dynamics of an amyloid-forming peptide gnnqqny from the yeast prion sup-35.* Biophys J, 2006. **91**: p. 824-833.

27. Balbirnie, M., R. Grothe, and D.S. Eisenberg, *An amyloid-forming peptide from the yeast prion sup35 reveals a dehydrated beta-sheet structure for amyloid.* Proc Natl Acad Sci U S A, 2001. **98**(5): p. 2375-80.

28. Antzutkin, O.N., et al., *Multiple quantum solid-state nmr indicates a parallel, not antiparallel, organization of beta-sheets in alzheimer's beta-amyloid fibrils.* Proc. Natl. Acad. Sci. USA, 2000. **97**: p. 13045.

29. Benzinger, T.L., et al., *Two-dimensional structure of beta-amyloid(10-35) fibrils.* Biochem., 2000. **39**: p. 3491.

30. Cheon, M., I. Chang, and C.K. Hall, *Extending the prime model for protein aggregation of all twenty amino acids.* PRoteins, 2010. **in print**.

31. Smith, A.V. and C.K. Hall, *A-helix formation: Discontinuous molecular dynamics on an intermediate resolution model.* Protein: Structure, Function and Genetic, 2001. **44**: p. 344.

32. Smith, A.V. and C.K. Hall, *Protein refolding versus aggregation: Computer simulations on an intermediate resolution model.* J. Mol. Biol., 2001. **312**: p. 187.

33. Marchut, A.J. and C.K. Hall, *Effects of chain length on the aggregation of model polyglutamine peptides: Molecular dynamics simulations.* Proteins, 2007. **66**(1): p. 96-109.

34. Marchut, A.J. and C.K. Hall, *Spontaneous formation of annular structures observed in molecular dynamics simulations of polyglutamine peptides.* Comput Biol Chem, 2006. **30**(3): p. 215-8.

35. Marchut, A.J. and C.K. Hall, *Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations.* Biophys J, 2006. **90**(12): p. 4574-84.

36. Nguyen, H.D., A.J. Marchut, and C.K. Hall, *Solvent effects on the conformational transition of a model polyalanine peptide.* Prot. Sci., 2004. **13**(11): p. 2909-2924.

37. Nguyen, H.D. and C.K. Hall, *Phase diagrams describing fibrillization by polyalanine peptides.* Biophys J, 2004. **87**(6): p. 4122-4134.

38. Nguyen, H.D. and C.K. Hall, *Molecular dynamics simulations of spontaneous fibril formation by rando-soil peptides.* Proc Natl Acad Sci USA, 2004. **101**(46): p. 16180-16185.

39. Nguyen, H.D. and C.K. Hall, *Kinetics of fibril formation by polyalanine peptides.* J Biol Chem, 2004. **in press**.

40. Blondelle, S.E., et al., *Polyalanine-based peptides as models for self-associated b-pleated-sheet complexes.* Biochemistry, 1997. **36**: p. 8393.

41. Forood, B., et al., *Structural characterization and 5'-mononucleotide binding of polyalanine b-sheet complexes.* J. Mol. Recognit., 1996. **9**: p. 488.

42. Gustafsson, M., et al., *The palmitoyl groups of lung surfactant protein c reduce unfolding into a fibrillogenic intermediate.* Journal of Molecular Biology, 2001. **310**: p. 937-950.

43. Hosia, W., J. Johansson, and W.J. Griffiths, *Hydrogen/deuterium exchange and aggregation of a polyvaline and an polyleucine alpha-helix investigated by matrix-assisted laser desorption ionization mass spectrometry.* Molecular and Cellular Proteomics, 2002. **1.8**: p. 592-597.

44. Szyperski, T., et al., *Pumonary surfactant-associated polypeptide c in a mixed organic solvent transforms from a monomeric alpha-helical state into insoluble beta-sheet aggregates.* Protein Science, 1998. **7**: p. 2533-2540.

45. Bellemans, A., J. Orbans, and D.V. Belle, *Molecular dynamics of rigid and non-rigid necklaces of hard disks.* Mol. Phys., 1980. **39**: p. 781-782.

46. Alder, B.J. and T.E. Wainwright, *Studies in molecular dynamics, i: General method.* J Chem Phys, 1959. **31**: p. 459-466.

47. Smith, S.W., C.K. Hall, and B.D. Freeman, *Molecular dynamics for polymeric fluids using discontinuous potentials.* Journal of Computational Physics, 1997. **134**: p. 16.

48. Rapaport, D.C., *Molecular dynamics study of polymer chains.* J. Chem. Phys., 1979. **71**: p. 3299.

49. Rapaport, D.C., *Molecular dynamics simulation of polymer chains with excluded volume.* J. Phys. A, 1978. **11**: p. L213.

50. Andersen, H.C., *Molecular dynamics simulation at constant temperature and / or pressure.* J. Chem. Phys., 1980. **72**: p. 2384.

51. Haspel, N., et al., *A comparative study of amyloid fibril formation by residues 15-19 of the human calcitonin hormone: A single b-sheet model with a small hydrophobic core.* JMB, 2204. **345**: p. 1213-1227.

3.7 List of Tables

3.8 List of Figures

Table 3. 1 PRIME20 Geometric and Energetic Parameters for glycine, alanine, and valine.

Identity	Abbr.	R-C_α bond	σ	m	ε_G	ε_A	ε_V
Glycine	G	0.654	3.10	0.34	0.034	0.034	0.034
Alanine	A	1.600	2.70	1.00	0.034	0.086	0.141
Valine	V	2.002	3.30	2.87	0.034	0.141	0.198

Table 3. 2 Types of structures formed by each sequence at inter-mediate temperatures. The percentages of peptides that are monomers, beta-sheets, amorphous aggregates and fibrils are indicated.

	T*	%Mono	%Beta	%Amor	%Fibril	Structure
AGAGAGAG	0.14-0.15	22.6	74.6	2.8	0.0	Disordered
GAAAGAAA	0.14-0.15	2.4	64.5	0.7	32.4	Slightly Ordered
AGAAAAGA	0.14-0.15	2.0	45.4	0.7	52.0	Ordered
GAAAAAAG	0.15-0.16	3.7	37.8	0.5	58.0	Ordered
AAAAAAAA	0.15-0.16	18.3†	29.1	3.4	49.2	Ordered†
VAGAAAAGAV	0.17-0.18	9.7	29.5	17.4	43.4	Slightly Ordered†
GAVAAAAVAG	0.16-0.17	6.3	15.1	4.1	74.5	Ordered
AAAAAAAAAA	0.16-0.17	22.1†	14.1	3.5	60.4	Ordered†
VAVAAAAVAV	0.17-0.18	4.4	31.3	31.0‡	33.3	Slightly Disordered‡
VVVVVVVVVV	0.17-0.18	0.9	20.5	44.5‡	34.1	Slightly Disordered‡

Hydrophobicity and Chain length

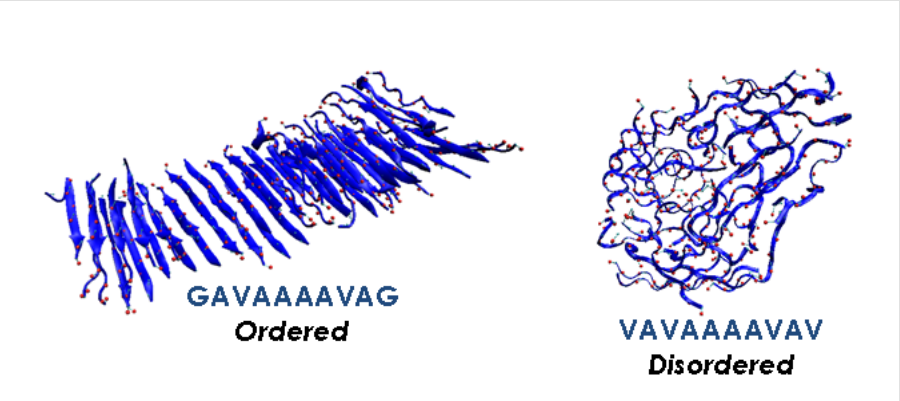

Figure 3. 1 Simulation snapshots of GAVAAAAVAG (left) and VAGAAAAGAV (right) at T*=0.17. Snapshots rendered using VMD.

Table 3. 3 Physical characteristics of fibrils formed by AGAAAAGA, GAAAAAAG, A8, VAGAAAAAGAV, GAVAAAAGAV, and A10.

	# Fibrils	# Sheets /Fibril	# Strands /Sheet	Parallel / Antiparallel	Intrastrand Distance
AGAAAAGA	1	2	16±7	0.90±0.2	4.80±0.4
GAAAAAAG	1	2-3	14±4	1.2±0.3	4.76±0.6
AAAAAAAA	1	2-3	15±3	0.99±0.2	4.83±0.4
VAGAAAAGAV	1	2	12±1	1.3±0.8	5.03±0.6
GAVAAAAVAG	1	2-3	13±2	1.9±0.2	4.83±0.5
AAAAAAAAAA	1	2-4	9±1	1.8±0.5	4.95±0.4

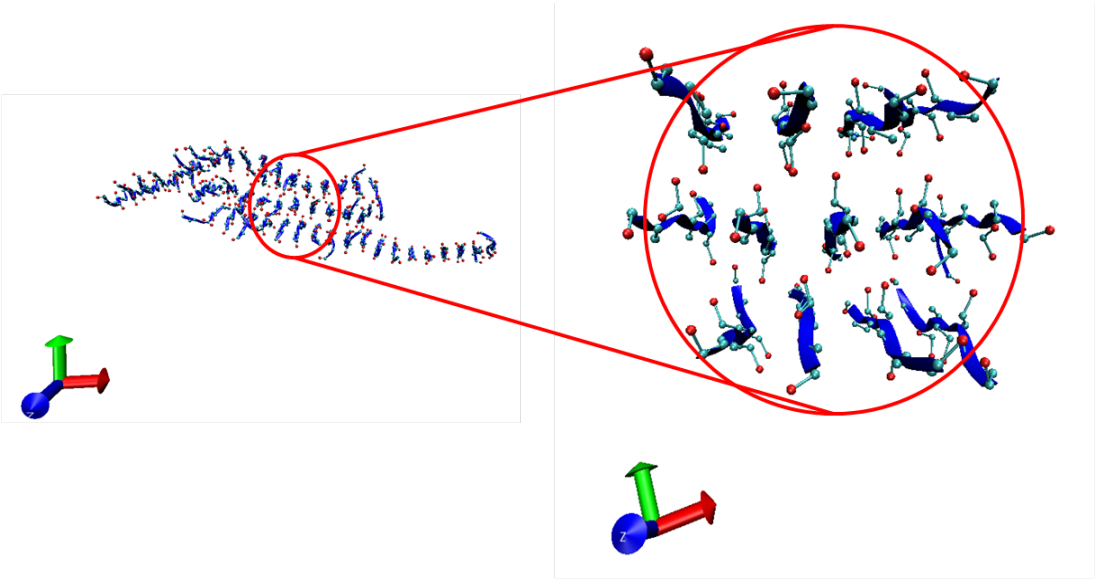

Figure 3. 2 Simulation snapshot of the SHaPrP 113-120 fibril 3.1 and close up of the arrangement of side chains at the interface between the sheets.

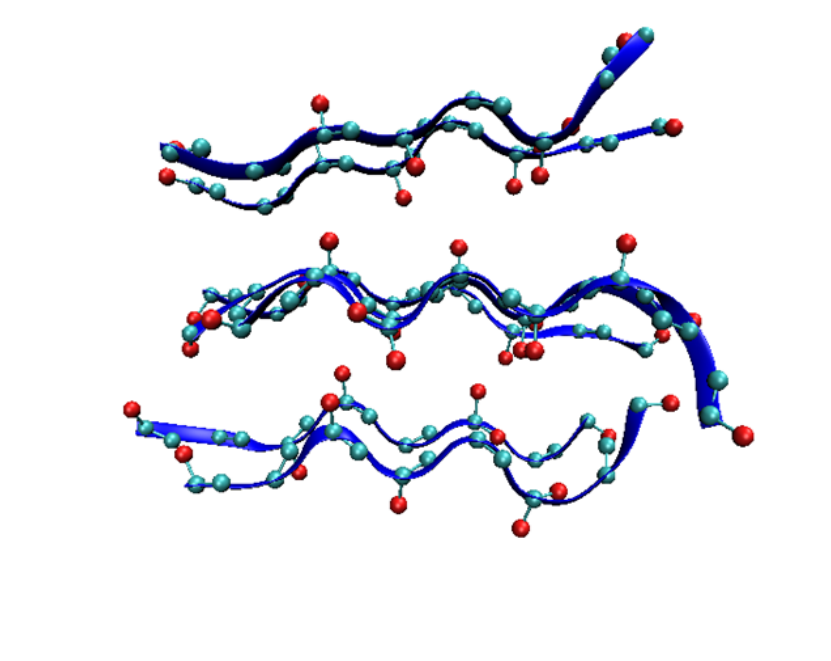

Figure 3. 3 Snapshot of the AGAAAAGA, rotated image of Figure 3.2.

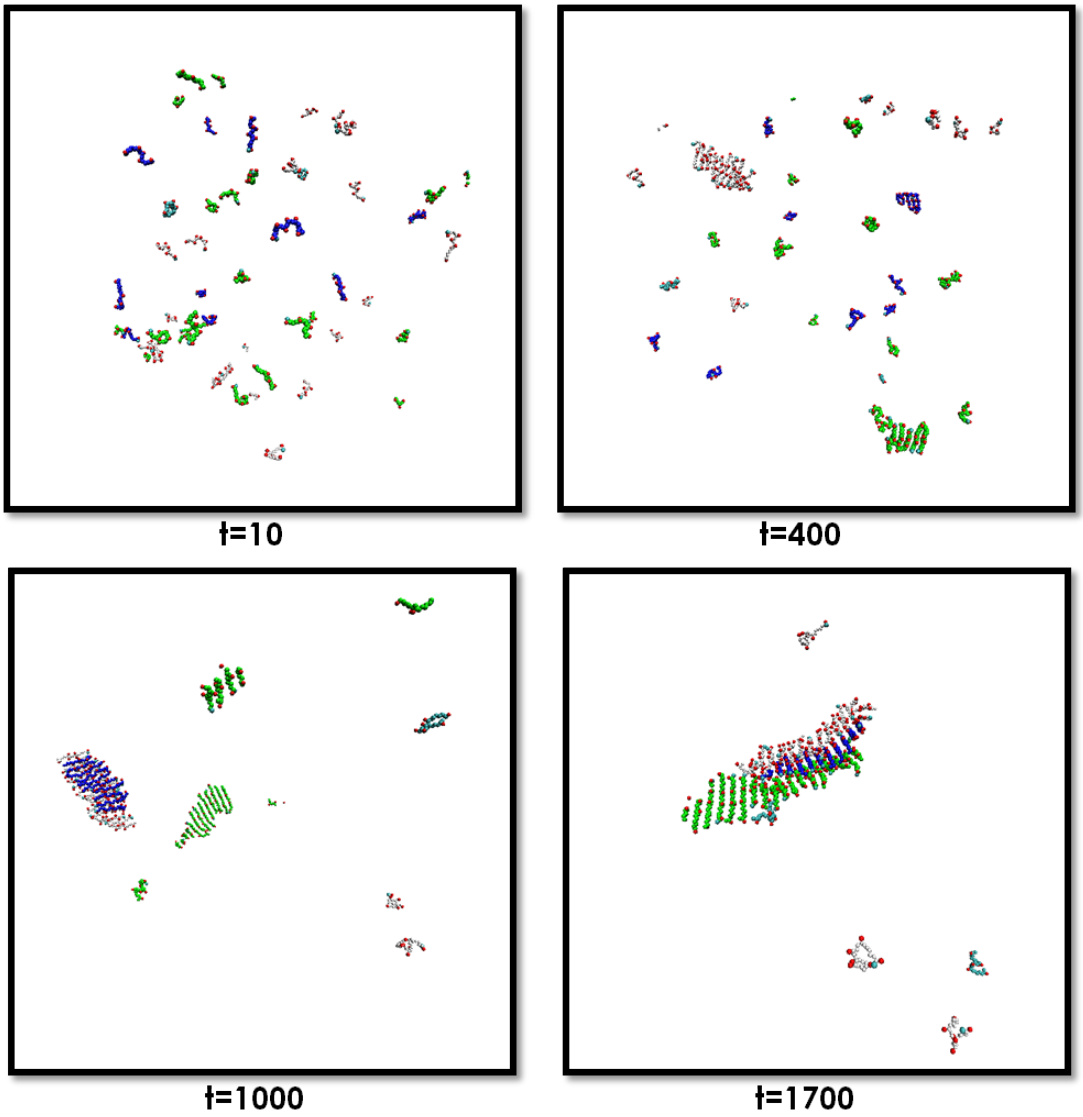

Figure 3. 4 Simulation snapshots for SHaPrP 113-120 at T*=0.15, shown at time, t=10 (very early), t=400, t=1000 and at the end of the simulation, t=1700.

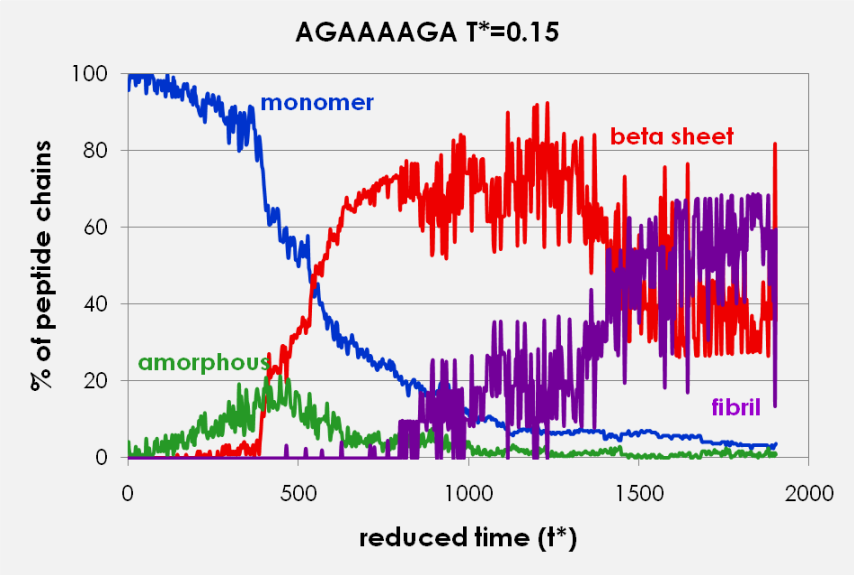

Figure 3. 5 Population graph for the various species (monomers, β-sheets, amorphous and fibrils for AGAAAAGA at T*=0.15.

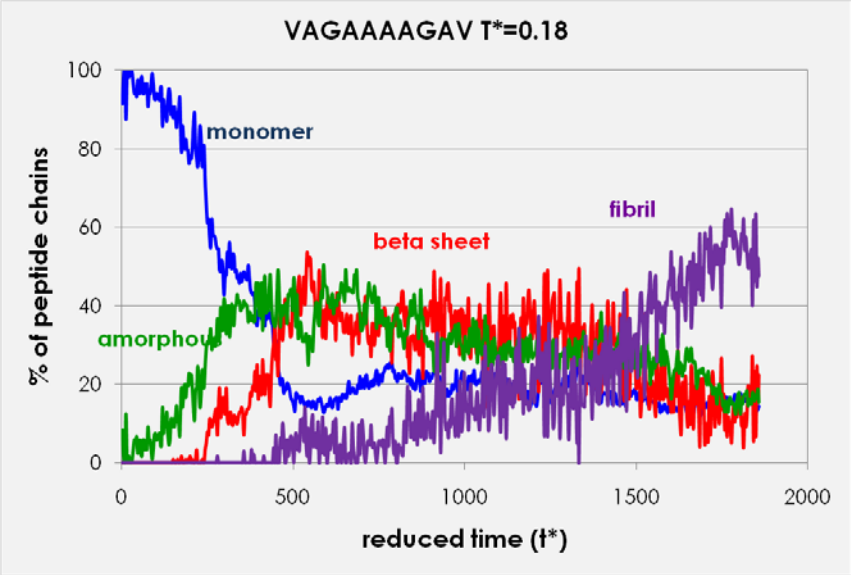

Figure 3. 6 Population graph for the various species (monomers, β-sheets, amorphous and fibrils for VAGAAAAGAV at T*=0.18.

CHAPTER 4

Understanding the Molecular Assembly of Small Fibril-forming Peptides: A Computer Simulation Study

4.1 Introduction

There are more than twenty known protein deposition diseases or "amyloidoses." These diseases affect multiple organs and are characterized by the abnormal deposition of ordered aggregates of proteins, called fibrils. Notable amyloidoses are Alzheimer's disease, caused by the abnormal deposition of the normally-present β-amyloid protein in the brain, Mad Cow disease,caused by aggregation of prion proteins in the central nervous system, and chronic inflammatory diseases, caused by an immune response triggered by the deposition of precursor Protein A in the liver. Although a different protein and a different organ system are involved in each of these diseases, the fibrils associated with all of them share a common basic structural feature , the so called "cross beta structure," which consists of layers of beta sheets running parallel to the fibril axis containing strands that run perpendicular to the fibril axis. This paper focuses on

understanding how variations in the sequence affect the likelihood that a particular sequence will form a fibril, the structure of the fully formed fibril, and the kinetic events that occur along the fibrillization pathway. Since the likely toxic species in the various diseases is believed to occur either on or off the fibrillization pathway, it is of interest to understand the connection between the protein sequence and the structure of the final product, the fibril.

Based on high-resolution x-ray diffraction studies on a range of synthetic and natural amyloid fibrils, Sunde et al. identified a general 'cross-β' structure in all fibrils, with β strands perpendicular to the fibril axis and backbone hydrogen bonds forming β-sheets parallel to the fibril axis with a 15° axial twist, a hydrogen bonding distance of 4.7-4.8 Å intra-sheet and an inter-sheet spacing of 10 Å [1]. All amyloid fibrils are ordered, insoluble structures 70-120 Å in diameter and any length [1-5]. These observations, along with work in Dobson's laboratory showing that non-disease-related proteins could also form fibrils [6], prompted the suggestion that fibril or protofibril [7] formation was an inherent property of proteins, regardless of the sequence [8-11] . The idea here was that all protein backbones can form hydrogen bonds and that backbone-backbone hydrogen bonds are the glue that holds proteins together in a fibril. Recent evidence suggests however that the primary sequence of a protein also plays a role in determining its ability to adopt the cross-β structure necessary for fibril formation [12-15] and that the side chains influence the unique molecular arrangement of amyloid peptides in a fibril.

We have known for some time that short, truncated sequence of amyloid peptides can in fact form fibrils [16-17] For example, Aβ, the peptide associated with Alzheimer's Disease, has several fragments that can independently aggregate into β-sheets [18-20]. These short peptide sequences are believed to play a role in determining whether or not a peptide will form fibrils [21]; in fact, the identity and position of specific side chains in these short peptides regulates fibril formation. Furthermore, side chain interactions at the molecular-level may enable the peptide to adopt the cross-β structure and also help explain the differences observed in the aggregation kinetics of different peptides.

The factors described above lead to the hypothesis that short segments of fibril forming peptides may be the key to amyloid formation, acting as the "Velcro" that holds these structures together. Accordingly, the goal of this work is to determine: which sequences will form fibrils, what types of side chains disrupt assembly, and how sequence affects molecular arrangement. Knowing the answers to these questions will help us to improve our understanding of amyloid fibril structure and the events leading to its formation.

Recent advances in x-ray microcrystal diffraction and solid-state NMR [22] have facilitated determination of the molecular structure of amyloid fibrils. Sawaya et al. have used high-resolution x-ray diffraction to determine the structure of one of the short prion sequences, GNNQQNY [13] They observed a pair of β-sheets whose side chains inter-

digitate to form a steric zipper, a so-called cross-β spine. The β-strands within a sheet are parallel and in-register; the strands of one sheet are exactly antiparallel to the strands in the next sheet, composing the cross-β spine. The interface between the two paired β-sheets is thought to be a dry interface (meaning no water molecules are present). The side chains (N2, Q4, and N6) tightly pack against the same three side chains on the neighboring sheet. This detailed structure provides further proof that primary sequence has a significant influence on the formation and structure of fibrils. Serpell and coworkers have also used high resolution x-ray diffraction to elucidate the molecular arrangement of peptides within a fibril [23]. They designed a 12 residue peptide, KFFEAAAKKFFE, to have sequence similarity with peptides associated with amyloidoses. Their x-ray data suggested that the β-strands within each sheet are antiparallel and that at least two sheets are held together in a staggered conformation (off-centered to allow the side chains to inter-digitate) due to stacking of the phenylalanine residues (π-π interactions).

Recently, Eisenberg and coworkers have identified several fibril-forming segments from disease-linked proteins and characterized their structure using x-ray diffraction on microcrystals formed by each short peptide [14]. They grouped the types of cross-β spines into eight classes but have only observed five to date. The classes are defined according to whether: 1) the β-strands within each sheet are parallel or antiparallel, 2) the sheets pack face-to-face (as in GNNQQNY) or face-to-back, and 3) the sheets are stacked parallel (up-up) or antiparallel (up-down). They found six different sequences, including GNNQQNY, which have parallel β-strands, sheets packed face-to-

face and stacked in an antiparallel conformation (Class 1). They also determined the molecular arrangement of short sequences related to the human prion protein (SNQNNF), amyloid-β (GGVVIA) and the insulin associated pre-cursor protein (VEALYL). Highlights of their results are the following. The prion protein (Class 2) has parallel β-strands within sheets packed face-to-back, with the β-sheets arranged up-up or parallel between the layers, i.e. F5(on sheet 2) fits into the pocket created by the S1, N2 and Q3 residues on the neighboring sheet (sheet1). Amyloid-β (Class 4) has parallel β-strands within sheets packed face-to-back, i.e. I5(on sheet 1), V4(on sheet 2) and V3(on sheet 1) form the core with the sheets stacked anti-parallel. The insulin pre-cursor protein (Class 7) has anti-parallel β-strands within sheets packed so that the residues of the core are the interior L4 of each peptide (face= back) and the sheets are stacked in parallel.

As we can see, primary sequence plays a significant role in determining the packing of the amyloid fibril structure but often times sequence alone is not enough to predict molecular arrangement. In order to tease out the details linking sequence to structure, Lopez de la Paz and Serrano performed a large-scale scanning mutagenesis experiment on the small designed amyloid peptide STVIIE. Each residue was replaced by nineteen of the twenty naturally occurring amino acids [15, 24] to determine which of the one hundred and fourteen combinations were still able to form fibrils. The identity of the interior residues played the most significant role in determining whether a sequence was able to form fibrils. Using the knowledge gained from this study, Lopez de la Paz and Serrano were able to determine which sequences would be good fibril formers and

which would not. Due to the sheer number of possible sequences considered, they did not examine the pathway for fibril assembly or the molecular arrangement of the strands within all the predicted fibrils.

Although great strides have been made in understanding the molecular details of fibril structure and how individual side chains may influence that arrangement, the mechanism for fibril formation is still not well understood. It is here that computational methods can provide insights. Nussinov and coworkers have studied many truncated amyloid sequences including GNNQQNY [25] and Aβ17-42 which contains the sequence GGVVIA [26]. Zheng et al. performed stability simulations at 330K on a single GNNQQNY β-sheet with 2, 3, or 4 strands arranged in parallel, or on a stack of two β-sheets with the β-strands within the sheet parallel and the orientation of the sheets either parallel or anti-parallel. They observed that at least 3 strands per sheet was necessary to retain the β-sheet structure and that two sheets were more stable than one sheet, better retaining native contacts and hydrogen bonds. Interestingly the tetramer composed of parallel β-strands with the sheets oriented in parallel dissolved, contradicting the experimental observation by Nelson, et al. [13]. Nussinov also observed that breaking up the hydrophobic core of GNNQQNY, by substituting the residues N2, Q4 and N6 with an A can significantly decrease the tetramer's stability. In more recent work, Zheng et al. studied the β-sheet association of $A\beta_{17-42}$ [26], incorporating the U-loop structure predicted by the Tycko model [12]. They created a double-layered model of eight Aβ peptides (four stacked β-sheets containing the salt bridge induced U-loop of Aβ). The

strands within each β-sheet were arranged in parallel and the β-sheets were arranged in an antiparallel conformation relative to one another (up-down) with either C-terminal to C-terminal interfaces or N-terminal to N-terminal interfaces. They found that C-terminus – to-C-terminus packing was more energetically favorable than the N-terminus—to—N-terminus packing. $A\beta_{17-42}$ contains the sequence GGVVIA (the C-terminal end of Aβ peptide) which has previously been shown in the x-ray diffraction study by Sawaya et al.[14] to form parallel β-sheets, with the sheets aligning in an up-down arrangement similar to the preferred structure predicted by Nussinov and coworkers.

Despite the advances made towards understanding how peptide sequence affects the ability of different peptides to adopt a universal fibril structure (cross-β structure with steric zipper core) there is still a lack of understanding as to which side chains might prevent fibril formation and the role of the side chain in the fibril structure. In an attempt to go beyond their earlier experimental work, Lopez de la Paz and Serrano, performed MD simulation on sixteen different sequences (STVIIE, SGVIIE, STVIGE, STVIFE, SAVIIE, and STVIAE) arranged in six-stranded β-sheets to determine which side chain mutations would destabilize the β-sheet [27]. Four different peptide conformations within each β-sheet were considered: 1) flat parallel (FP), 2) flat antiparallel (FA), 3) twist parallel, and 4) twist antiparallel. For most of the sequences considered, the FP conformation was slightly more stable than its twisted counterpart. Of course, as the authors point out, since there was a positive charge on the N-terminus amide group; charge-charge repulsion would prevent the parallel arrangement of strands. They went

on to test the conclusion from their earlier paper that position five of STVIIE was most likely to disrupt fibrillar structure by showing that the sequence STVIGE disrupted the stability of the antiparallel β-sheet most likely by disrupting the hydrophobic core of the peptide. Substituting phenylalanine at position five (STVIFE) resulted in a preference for a twisted antiparallel conformation of β-strands within the sheet. This is interesting because STVIFE is one of only three mutations at position five where they observed fibril growth *in vitro*, the other two being the parent sequence STVIIE and STVIYE. Two additional sequences that did not form fibrils *in vitro*, STVIGE and STVIAE, were somewhat able to remain in their initial β-sheet complex [27].

Since the Sawaya et al. [14] publication of the measured atomic structures of the fibrils formed by the thirteen amyloid-related peptides by, many researchers have used atomistic molecular dynamics simulations to answer questions about assembly kinetics, details of fibril structure, preference for a particular arrangement for these peptides. Others have tried to determine if the predicted crystal structure of these peptides is the free energy minimum (thermodynamically most stable), and how peptide sequence contributes to its stability. Park et al. [28] simulated five different peptide sequences: 1) GNNQQNY, 2) NNQQ, 3) VEALYL, 4) KLVFFAE, and 5) STVIIE, initially placed into all possible arrangements of β-sheet bilayers; the free energy of binding for each arrangement was calculated using the generalized Born solvation model. Briefly, the most stable structures for GNNQQNY and NNQQ were parallel β-sheets and for VEALYL, KLVFFAE, and STVIIE were antiparallel β-sheets. They also found that as

previously reported by Nussinov's group, Eisenberg's x-ray structure for GNNQQNY is the most thermodynamically stable conformation but multiple minimum energy conformations were observed for NNQQ. VEALYL (Ainv P2) preferred staggered parallel interfaces between the β-sheet bilayer with side chains arranged front-to-front and KLVFFAE (Areg FF) preferred staggered parallel interfaces between the β-sheet layers with side chains arranged front-to-back; the latter is consistent with Ma and Nussinov's earlier findings [29] Finally, the most stable configuration for STVIIE was the same as that for VEALYL where the antiparallel β-sheets form a bilayer of staggered β-sheets with the side chains in a front-to-front arrangement. Interestingly, the greatest energetic contribution to the binding energies seemed to come from the free energies of the non-bonded interactions, i.e. side chains. Based upon this work, Park and coworkers concluded that the steric zipper forms most easily when the β-sheets are parallel because this allows similar or identical side chains to interact at the interface. In addition, since the backbone hydrogen bonds accounted for almost one-quarter of the total non-bonded interaction energy in their simulations, they concluded that the long-held idea that hydrogen bonding plays a significant role in the formation of the universal fibril structure is correct. [28].

Several other groups have used molecular dynamics simulations to study the short, truncated amyloid-forming sequences identified by Sawaya et al. De Simone et al. performed GROMACS-based all atom, explicit solvent, 20 ns molecular dynamics on a pair of β-sheets composed of either two, three, four, ten or fifty SNQNNF peptides [30]

associated together into the measured crystal structures. Twenty nanoseconds was not enough time for the fifty SNQNNF peptides system to equilibrate. The smallest system that remained stable over the time course of the simulation was the tetramer bilayer structure. Interestingly, they observed β-sheet twisting with a face-to-back orientation of the side chains between the two sheets and F6 packed tightly into the core. Vitagliano et al. [31] performed 60 ns all-atom molecular dynamics on a bilayer of β-sheets composed of ten strands of SSTSAA and of VQIVYK using the GROMACS package with explicit solvent. Over the time course of the simulation the inter-sheet interactions taken from the measured crystal structure of SSTSAA (face-to-face) completely deteriorated although the β-structure remained, suggesting that SSTSAA cross-β spines are unstable. The VQIVYK bilayer with face-to-face cross-β spine composed of V1, I3 and V4, was more stable than the SSTSAA bilayer and retained much of its β-structure, although it did not preserve the cross-β spine motif. Ho and coworkers have also used all-atom molecular dynamics simulations with DISCOVER 2.9.8 to the study β-sheets of GGVVIA [32] and VEALYL [33]. For GGVVIA their starting configurations were single-layered parallel β-sheets with two, three, four, or five strands and bilayers of two, three, four or five parallel β-strands arranged in antiparallel β-sheets face-to-back as predicted by the crystal structure of Sawaya et al.. Over the 10 ns simulated, the oligomers of GGVVIA remained relatively stable, indicating that the hydrophobic contacts I5, V4 and V3 play a significant role in holding the cross-β spine together. Lin et al. performed 10 ns stability simulations using DISCOVER 3 on the crystal structure of VEALYL, with the β-strands

anti-parallel within a sheet and sheets stacked in parallel, face-to-back[33]. Over the course of the simulation, the β-sheet bilayer remained relatively stable, again suggesting the importance of the hydrophobic residues at the interface.

Although the atomistic simulations described above provide insights on the stability of particular fibrillar structures, they do not provide information about the pathway to fibril formation. In this work, simulations are used instead to characterize the kinetics along the pathway of fibril formation. This is accomplished by looking at the spontaneous assembly of several short, truncated amyloid and amyloid-like peptides. We simulate five of the peptides whose fibril crystal structures were measured by Sawaya et al. VEALYL, MVGGVV, SSTSAA, SNQNNF, GGVVIA, [14] and the *de novo* designed peptide of Lopez de la Paz and Serrano, STVIIE [15]. We also simulate the assembly of variations of STVIIE: STVIFE, STVIEE, STAIIE, STVIAE, STVIGE, STVIVE, the first two of which were found experimentally to form fibrils and the latter five of which were not, to test the robustness of our protein model. We combine PRIME20, a new intermediate–resolution description of the protein geometry for all twenty amino acids [34] that we have developed , with discontinuous molecular dynamics (DMD) [35-40] to simulate 48-peptide homogenous systems containing the previously listed peptides starting from random configurations of random coils at high temperatures. Our goals are to test the ability of PRIME20 to distinguish the role played by each of the twenty different amino acids in fibril formation, to validate PRIME20's ability to predict each sequence's propensity to form fibrils, and to contribute to the fundamental understanding

of the fibril formation pathway. The systems are cooled slowly to the desired simulation temperatures. We explored how changing sequence and temperature affect the aggregation pathway by monitoring the formation of different structures such as β-sheets (dimers, trimers, tetramers, pentamers, hexamers and other large oligomers), amorphous aggregates and fibrils as a function of time. We also observe the molecular arrangement of β-sheets and determine which side chains are located at the interface between two β-sheets.

Highlights of our results are the following. We are able to observe the spontaneous formation of fibrillar structures in a 48-peptide system starting from a random configuration of random coil conformations, depending upon the sequence and the temperature. The short amyloid peptide fragments SSTSAA and SNQNNF form fibrils at low temperatures; MVGGVV forms fibrils at intermediate temperatures, and GGVVIA forms β-sheets but not fibrils over a range of temperatures at least for the simulation times we have studied thus far. Despite having remarkably different sequences, SSTSAA and SNQNNF aggregate by a process in which monomers are completely depleted through association into β-sheets, followed by β-sheet association into fibrils. SSTSAA shows no preference for a parallel or antiparallel arrangement of β-strands in a β-sheet whereas SNQNNF prefers a parallel arrangement. At higher temperatures MVGGVV experiences a competition between amorphous aggregates and

fibrils and becomes more disordered. The short *de novo* designed peptides form fibrils over a broader range of temperatures than the amyloidogenic ones. STVIIE, STVIFE formed fibrils as expected. STVIVE, STAIIE, STVIAE, and STVIGE were all observed to form fibrils though to a lesser degree than STVIIE and STVIFE; these sequences did not form fibrils *in vitro*. STVIEE did not form fibrils in our simulations nor in experiments. In our simulations, when the fifth position contains a hydrophobic amino acid alanine (STVIAE) or valine (STVIVE) the sequence readily forms fibrils. However, as temperature increases most of the sequences not expected to form fibrils do not, like STAIIE, STVIAE and STVIGE. The best results are found at the highest temperatures where our results on the fibrillization propensity of six of the seven short *de novo* designed peptides agree with experiment. This is consistent with other simulation work which shows that the best fibrils form at a marginal temperature above which the system forms random coils. None of the designed peptides show a preference for parallel or antiparallel strand within their β-sheets or form amorphous intermediates along their fibrillization pathway.

This paper is organized as follows. In the next section, we describe the peptide model and simulation method. In the following section, we present the results obtained from simulation of multi-peptide systems at various conditions. The last section is a discussion of our results.

4.2 Methods

Model Peptide and Forces

In this work we apply a new implicit solvent force field PRIME20 to describe the geometry and energetics of the short segments of heteropeptides sequences considered here. PRIME20 was recently introduced by Cheon et al. [34] as an extension of PRIME, an implicit solvent intermediate-resolution protein model previously used in simulations of the aggregation of polyalanine and polyglutamine. PRIME was originally developed by Smith and Hall [36, 41] and later improved by Nguyen et al. [38]. More recently the PRIME model was extended to the study of polyglutamine peptides [39] illustrating its versatility. In PRIME, the protein backbone is represented by three united atom spheres, one for the amide group (NH), one for the carbonyl group (CO), and one for the alpha-carbon and its hydrogen ($C_\alpha H$). In the original version of PRIME, each side chain was represented by a single sphere for polyalanine and by a chain of four spheres for polyglutamine. In PRIME 20, the twenty possible side chains are modeled as single spheres of unique size, atomic mass and C_α—R bond length. All backbone bond lengths and angles are set to their ideal values. In order to maintain the *trans*-configuration we fix the consecutive C_α—C_α distance. The side chains are positioned relative to the protein backbone so that all residues are L-isomers. The solvent molecules in our system are modeled implicitly.

All forces between the united atom spheres are modeled with discontinuous potentials, e.g. hard sphere and square-well interactions. The excluded volume of each of the peptide's four united atoms is modeled using a hard sphere interaction. The covalent bond lengths are maintained using a hard sphere interaction that prevents them from moving outside of the range $(1+\delta)l$ to $(1-\delta)l$, where l is the ideal bond length and δ is the tolerance, which is set at 2.375% [42]. Ideal backbone bond angles, C_α—C_α distance, and the residue L-isomerization are maintained by imposing a series of pseudobonds whose lengths are also allowed to fluctuate by 2.375%.

Hydrogen bonding is represented in PRIME 20 as a square well attraction of depth ε_{HB} and width 4.5Å between the backbone amide and carbonyl groups. Hydrogen bonds are anisotropic in nature so we must constrain their formation to occur only when the NH united atom vector and the CO united atom vector point towards each other and the angle between those vectors is restricted between 120° and 180°. In order to accomplish this, the following criteria must be met. A hydrogen bond forms only when: 1) the virtual hydrogen and oxygen atoms (whose location can be calculated at any time) are separated by a distance of 4.5Å (the sum of the NH and CO well-widths), 2) the nitrogen-hydrogen and carbon-oxygen vectors point toward each other within a fairly generous tolerance, 3) neither the NH nor the CO are already involved in a hydrogen bond with a different partner, and 4) the NH and CO are separated by at least three intervening residue along the chain. Further details on the hydrogen bonding model can be found in our earlier work. [37-38, 40]

The non-hydrogen-bonding interactions in PRIME20 are all modeled as square well interactions between the spherical units on each amino acid with strength (well depth) and range determined individually for each pair. Since solvent is modeled implicitly these are all effective interactions or potentials of mean force. In PRIME20, the energy parameters that describe the side chain / side chain interactions and the hydrogen bonding interactions between backbone NH and CO, and between side chain and side chain are derived in the following way. Briefly, the twenty possible amino acids are classified into 14 groups: [LVI] [F] [Y] [W[[M[[A] [C] [ED] [KR] [P] [ST] [NQ] [H] [G], according to their side chain size, hydrophobicity, and possibility of hydrogen bonding. The aforementioned energy parameters were determined by Cheon et al. [34] who applied a perceptron-learning algorithm and a modified stochastic learning algorithm to optimize the energy gap between 711 known native states from the PDB and decoy structures generated by gapless threading. The number of independent pair-interaction parameters was chosen to be small enough to be physically meaningful yet large enough to give reasonably accurate results in discriminating decoys from native structures. A total of nineteen interaction parameters were used to describe the side chain energetics. Tables 4.1 and 4.2 provide a summary of the geometric and energetic parameters for each amino acid side chain.

The system temperature is scaled by the hydrogen bonding energy between the backbone NH and CO, ε_{HB}, so that the reduced temperature is $T^* = k_B T/\varepsilon_{HB}$.

Discontinuous Molecular Dynamics

Discontinuous molecular dynamics (DMD) is a variant on standard molecular dynamics that is applicable to systems of molecules interacting via discontinuous potentials (e.g., hard sphere and square-well potentials). Unlike soft potentials such as the Lennard-Jones potential, discontinuous potentials exert forces only when particles collide, enabling the exact (as opposed to numerical) solution of the collision dynamics. This imparts great speed to the algorithm, allowing sampling of longer time scales and larger systems than traditional molecular dynamics. The particle trajectories are followed by locating the time between collisions and then advancing the simulation to the next collision (event) [43-44]. DMD on chain-like molecules is generally implemented using the "bead string" algorithm introduced by Rapaport [45-46] and later modified by Bellemans et al. [42]. Chains of square-well spheres can be accommodated in this algorithm by introducing well-capture, well-bounce, and well-dissociation "collisions" when a sphere enters, attempts to leave, or leaves the square well of the adjacent sphere. In this paper, DMD simulations are performed in the canonical ensemble (NVT) with the initial velocities chosen randomly from a Maxwell-Boltzmann distribution about the desired system temperature. The initial positions of the particles or spheres are chosen randomly while still ensuring that no geometrical constraints are violated.

In DMD simulations of protein aggregation, we often determine the number of particles in the system by specifying the concentration which is given by $c = N/L^3$, where

N is the number of molecules in the box and *L* is the simulation box length. Periodic boundary conditions are imposed so that if a molecule travels outside the simulation box its image reenters the box on the opposite side and continues to participate in the dynamics. Since we are simulating large systems at high concentrations starting from random initial configurations we must ensure that the box length is large enough to prevent the molecules from interacting with themselves but still allow them to interact with their periodic image. We set L = 200Å in this study. The simulation proceeds according to the following schedule: identify the first event (e.g., a collision), move forward in time until that event occurs, calculate new velocities for the pair of spheres involved in the event and calculate any changes in system energy resulting from hydrogen bond events or side chain interactions, find the second event, and so on. Types of events include excluded volume events, bond events, and square-well hydrogen bond and side chain interaction events. An excluded volume event occurs when the surfaces of two hard spheres collide and repel each other. A bond (or pseudobond) event occurs when two adjacent spheres attempt to move outside of their assigned bond (or pseudobond) length and the two particles feel an infinite repulsion that forces them back into their assigned bond length. Square-well events include well-capture, well-bounce, and well-dissociation ‘‘collisions’’ when a sphere enters, attempts to leave, or leaves the square well of another sphere. For more details on DMD simulations with square-well potentials, see articles by Alder and Wainwright [43] and Smith et al. [44].

A total of twelve model systems are studied in this work; all contain 48 peptides at concentrations c = 10mM. The peptides considered are VEALYL, MVGGVV, SSTSAA, SNQNNF, GGVVIA, STVIIE, STAIIE, STVIAE, STVIFE, STVIVE, STVIGE, and STVIEE. Each simulation is started at high temperature to ensure a random initial configuration and then slow-cooled to the temperature of interest to minimize kinetic trapping. Slow-cooling is achieved by decreasing the temperature in discrete steps starting from a high temperature until we reach the desired simulation temperature. The simulation temperature is maintained using the Andersen thermostat [47]; in this method all the particles undergo random infrequent "events" or "collisions" with a ghost particle that reassigns the particle's velocity randomly from a Maxwell-Boltzmann distribution centered at the simulation temperature. Five simulations are run for each sequence at the given temperature and concentration (state). Error bars are taken to be the standard deviation at each state. All simulations are run for an average of 100-200 billion collisions depending on simulation conditions, sequence, temperature and concentration. We can perform 500 million collisions per hour on these systems using a 2.4 GHz AMD Opteron™ workstation, which means that our simulations require approximately a 200 hours.

In this study, we monitor and analyze the formation of β-strands, β-sheets, amorphous aggregates and fibrils. We also check to see if the β-strands in a β-sheet are arranged in a parallel or anti-parallel configuration. The criteria for assigning the types of structures formed are the following. If each peptide in a group of peptides has at least

two inter-peptide hydrogen bonds or side chain interactions with a neighboring peptide in the same group, then that group is classified as an aggregate. Aggregates can be either ordered or amorphous. If an aggregate contains β-sheets or fibrils, we classify it as an ordered aggregate. If each peptide in a group of peptides has at least n/2, where n = chain length, inter-peptide β-hydrogen bonds to a particular neighboring peptide in the group, we classify this group as a β-sheet. (A β-hydrogen bond is a hydrogen bond between two residues whose backbone angles are in the β-region of the Ramachandran plot.) If at least two β-sheet structures form inter-sheet side chain interactions (at least four side chain interactions per peptide per β-sheet), we classify this as a fibril; otherwise, we classify this and isolated β-sheets as non-fibrillar β-sheet structures. If an aggregate does not contain β-sheets but the peptides in the aggregate have any side chain contacts, then the aggregate is considered amorphous. If an aggregate contains peptides with less than n/2 inter-peptide β-hydrogen bonds between neighboring chains then this is also considered to be an amorphous aggregate.

4.3 Results

DMD simulations were performed on 48-peptide systems containing fragments of several amyloidogenic peptides: VEALYL (from insulin), SSTSAA (from RNase), SNQNNF (from prion), MVGGVV and GGVVIA (both from Aβ); and *de novo* designed peptides: STVIIE, STAIIE, STVIAE, STVIFE, STVIVE, STVIGE, STVIEE. The goals

were to validate PRIME20's ability to predict each sequence's propensity to form fibrils, to identify and analyze structural aspects of our simulated fibrils that might beexperimentally observable, to learn, how variations in sequence affect the tendency of peptides to form fibrils, and to see if changes at particular positions along the chain prevented fibril formation.

Fibrillization of short fragments of amyloidogenic peptides

Table 4.3 summarizes our simulation results on STVIIE, VEALYL, MVGGVV, SSTSAA, SNQNNF, and GGVVIA. The table is organized into three sections, each containing the results of multiple runs at three temperatures that span the range over which each sequence is most likely to form fibrillar structures. Table 4.3 lists the percentage of peptides at the end of the simulation runs that are monomers, in β-sheets, in amorphous aggregates or in fibrils. The nature of the final structures formed are categorized as "ordered" (greater than 40% fibril and β-sheet), "slightly ordered" (low or no observed fibrils and high β-sheet), "slightly disordered" (high amorphous content but still containing fibrils or β-sheets), and "disordered" (high amount of monomers or amorphous aggregates).

From Table 4.3a, we can see that at low temperature T* = 0.13 our STVIIE and SNQNNF readily form fibrils and that SSTSAA also forms fibrils but to a lesser extent. At T*=0.13, GGVVIA almost completely assembles into β-sheets but no fibrils are observed. As we increase to intermediate temperatures in Table 4.3b, we see that

STVIIE continues to form fibrils and that the other Aβ sequence, MVGGVV, forms an even higher percentage of fibrils at T*=0.15. At T*=0.15, SNQNNF forms less fibrilsthan it did at lower temperatures; it is only slightly more ordered than VEALYL but still readily forms β-sheets. At the intermediate temperatures, SSTSAA and SNQNNF still associate into β-sheets but their ability to form fibrils is drastically reduced. As we increase to the highest temperature, in Table 4.3c only STVIIE and SNQNNF continue to form fibrils but in much lower numbers than at lower temperatures. Even the robust β-sheet producers, SSTSAA and GGVVIA begin to lose β-sheet content. Since β-sheet formation is primarily a function of the peptide backbone it is not surprising to see that most of the six-residue fragments considered were able to form β-sheets regardless of temperature. Interestingly, the sequence VEALYL had difficulty forming β-sheets at all temperatures and barely formed any fibrils at T*=0.15. We are not sure if VEALYL cannot easily form fibrils or if we have not yet found the set of conditions which would allow fibril formation. Another perplexing sequence is SNQNNF at the intermediate temperature, T*=0.15, which forms more β-sheets but less fibrils on average than at the higher temperature, T*=0.16. This suggests that a temperature between $0.15 \leq T^* \leq 0.16$ may be a transition temperature for this sequence and that further exploration of temperature space is required. GGVVIA did not form fibrils in our simulations even though its crystal structure has been observed and a two–layer structure remained stable during simulations by Chang et al. It may be that our time scales are too short or our temperatures are too high to allow the β-sheets to find and associate together into fibrils.

The results from Table 4.3 are not surprising. Most of the six-residue segments were able to form β-sheets and very little amorphous aggregation was observed except for VEALYL. At low temperatures, SNQNNF and STVIIE readily formed fibrils but STVIIE formed fibrils over a wider range of temperatures perhaps because it is not as hydrophobic as SNQNNF. We expected SSTSAA to have a more difficult time forming fibrils because of the absence of hydrophobic residues in the core. GGVVIA should also have a more difficult time forming fibrils because of the presence of glycines, which essentially truncate the segment to four active residues.

Figure 4.1 shows snapshots of the best fibrils formed by STVIIE (Figure 4.1a) at T*=0.16 and by SNQNNF (Figure 4.1b) at T*=0.13. The fibril for STVIIE is a bilayer with a prominent twist; one β-sheet is colored pink and the other is colored light blue. The fibril for SNQNNF is also a bilayer (with one β-sheet colored blue and one β-sheet colored white). The β-sheets in the SNQNNF fibril intersect each other (see ribbon diagram).

We examined the assembly of GGVVIA and MVGGVV at intermediate temperatures. These sequences are of particular interest because together they make-up the eight residues of the Aβ peptide, Aβ 35-42. Fibrils for both of these sequences have been isolated experimentally. In our simulations, however, we observed fibrils for MVGGVV but only β-sheets for GGVVIA at intermediate temperatures. At higher temperatures, MVGGVV forms fewer fibrils and instead forms an equal percentage of β-

sheets and amorphous aggregates. At higher temperatures, GGVVIA is still able to form a high percentage of β-sheets but almost no amorphous aggregates. This would seem to suggest that the end residues of Aβ, isoleucine and alanine do not significantly contribute to the fibrillization of Aβ. It is unclear why we do not observe fibrils for GGVVIA. It is well-understood that glycine increases chain flexibility but both sequences contain the same number of glycines. In fact, the glycines disrupt the hydrophobic core of MVGGVV suggesting that GGVVIA should still have the propensity to form fibrils and MVGGVV should not. Interestingly, the temperatures at which association for GGVVIA occurred were overall much lower than those for MVGGVV (GGVVIA is able to form β-sheets at T*=0.13 but remains in a random-coil configuration at higher temperatures). This might indicate that the identity and position of particular amino acids like glycine may determine which sequences are able to form fibrils (see results of Chapter 3). It may be interesting to substitute the alanine residue on GGVVIA with methionine to see if that substitution would allow GGVVIM(A) to form fibrils.

In order to understand the differences in the structures of these two different sequences we examined their aggregation pathways. Figures 4.2 and 4.3 show the percentage of peptides in each species (monomer, β-sheet, amorphous aggregate, fibril) as a function of simulation time at intermediate temperatures for GGVVIA at T*=0.13 and MVGGVV at T*=0.15)., respectively. At time, t=0, we see that both segments start out as monomers (blue curve) in Figure 4.2 and Figure 4.3. As time increases, both GGVVIA (Figure 4.2) and MVGGVV (Figure 4.3) form some amorphous aggregates

(green curve) and at time, t=500, both systems begin to form β-sheets (red curve). For GGVVIA as β-sheets (red) continue to grow, amorphous aggregates (green) begin to disappear. In contrast for MVGGVV as β-sheets (red) continue to grow, amorphous aggregates (green) grow as well. In Figure 4.2 at time, t=600, there appears to be a competition between β-sheet and amorphous aggregate formation for GGVVIA until finally at t=1000, β-sheet growth takes off but no fibrils are observed. Figure 4.3 shows at time, t=400, that β-sheet growth rapidly increases for MVGGVV and at time, t=600, fibrils (purple curve) begin to appear and amorphous aggregates begin to disappear. The delayed time observed for β-sheet formation by GGVVIA may simply be a result of its lower system temperature. Simulations are currently being performed at the lower temperature T*=0.13 for MVGGVV to determine if fibrils still grow and at the lower temperature T*=0.10 for GGVVIA to see if given enough time the β-sheets come together to form a fibril.

Figures 4.4 and 4.5 show simulation snapshots of the kinetic pathways leading to the formation of the fibrils by GGVVIA at T*=0.13 and for MVGGVV at T*=0.15 respectively. In Figure 4.4 we see that at t= 0, GGVVIA starts off as random coils. By t=600 a β-sheet containing a blue hexamer and a white tetramer has begun to form. At time, t=2000 the blue and white β-sheets have grown into larger oligomers but they have not associated together to form a fibril. In Figure 4.5 we see that at time t=10 MVGGVV starts off in a random coil configuration. At time, t=200, we see both β-sheets (blue) and amorphous aggregates (green, white and dark blue peptides). As time increases to t=700

most of the amorphous aggregates have disappeared and a fibril has begun to form. That fibril continues to grow. The snapshot at time, t=1000, shows a close up of the final fibril.

MVGGVV clearly prefers to arrange into ordered conformations at T* = 0.15 (See Figure 4.5) but undergoes a shift towards amorphous aggregate formation at higher temperatures like T*=0.17 (See Table 4.3c.). This shift towards amorphous aggregate formation is not observed for SSTSAA or SNQNNF. In order to understand why this might be happening for MVGGVV we plot in Figure 4.6 population curves (monomers, β-sheets, amorphous aggregates and fibrils) for MVGGVV at T*=0.17. From the very beginning of the simulation at time, t=0, as monomers decrease (blue curve), amorphous aggregates increase (green curve). After a short lag time, β-sheets (red curve) begin to grow but they never succeed in overtaking the amorphous aggregates. Since the population of β-sheets remains small there are not enough β-sheets to encourage fibril growth (purple curve) and the system remains in the amorphous state.

In order to understand the details of the molecular arrangement of each peptide we calculated the number of strands either in parallel or anti-parallel conformations for the best fibril formed by each sequence. Table 4.4 shows which conformation was preferred by each sequence in our simulations. Row 1 of Table 4.4 is the fraction of peptides in a parallel conformation and row 2 is the fraction of peptides in an anti-parallel conformation. Since all of these peptides are relatively well-structured, the sum of

parallel and anti-parallel peptides is close to 1.0; therefore we report the ratio of parallel to anti-parallel in row 3. STVIFE and STVIIE, which are both *de novo* designed peptides and not associated with any amyloidogenic disease both, show a preference for anti-parallel conformations. Three of the peptides associated with disease, VEALYL, MVGGVV, SSTSAA show a strong preference for parallel conformations. The prion sequence, SNQNNF, did not show a preference for either parallel of anti-parallel. It is likely that the charged residues on the end of STVIFE and STVIIE contribute to their anti-parallel conformational preference because charged residues would prefer to avoid each other when they are arranged in a β-sheet. According to the most recent micro-crystal x-ray structures for VEALYL and MVGGVV, these peptides should prefer an anti-parallel arrangement. Interestingly SSTSAA does show a preference for parallel β-strands within a sheet. In our 48 peptide chain simulations, we also observed a preference for two sheet fibrils which is in agreement with the x-ray data. At this time we are unable to conclusively determine the amino acid residues at the β-sheet interface of the fibrils. Some of the difficulty lies in the imperfect nature of the fibrils we see in our simulations. Further investigation into physical parameter classification is ongoing.

Fibrillization of de novo designed peptides

Before we describe the simulation results, we remind the reader of the experiments by Lopez de la Paz and Serrano [24, 27] that motivate these simulations. Recall that their designed sequence STVIIE readily formed fibrils *in vitro*, and that STVIFE also formed fibrils, but that mutants STVIVE, STVIAE, STAIIE, STVIGE, and STVIEE did not. See column 2 on Table 4.5. The fibrils were grown for 20 days at an initial concentration of 1mM, at pH 7.4 and at room temperature (25°C), and then concentrated to 10mM. Fibrils were observed by CD and EM 10 days later.

Table 4.6 summarizes our simulation results on the *de novo* designed sequences, STVIIE, STAIIE, STVIAE, STVIVE, STVIFE, STVIGE, and STVIEE at the end of the simulations at low, intermediate and high temperatures. The table lists the percentage of peptides at the end of the simulation runs that are monomers, in β-sheets, in amorphous aggregates, or in fibrils. In our simulations, at low temperature, T*=0.13, STVIIE readily forms fibrils and so do its sequence variants STAIIE, STVIAE, STVIVE, STVIFE. STVIGE and STVIEE form fibrils to a lesser extent. Experimentally however, only STVIIE and STVIFE formed fibrils. As we increase temperature to a range of $0.145<T^*<0.155$, STVIIE, STAIIE, STVIAE, STVIVE and STVIFE continue to be able to form fibrils. The number of fibrils formed by STVIGE has been halved and STVIEE no longer forms fibrils. At high temperatures, STVIIE, STVIVE and STVIFE still readily form fibrils and STVIGE has all but lost the ability to form fibrils. Our results on

STVIAE are inconclusive and we are continuing those simulations. STVIEE does not form fibrils at the higher temperatures and its ability to form β-sheets has decreased as well.

The comparison with experiments is summarized in Table 4.5. In order to qualitatively compare our results with the experiments we decided to use the following criterion to determine whether a simulation had formed a fibril or not. If the fibril percentage is less than 40% then we do not consider it a fibril. Using this definition we observed that at low temperatures, STVIIE and STVIFE form fibrils as expected, STVIGE and STVIEE do not form fibrils as expected, and STVIVE, STVIAE and STAIIE forms fibrils in contradiction to the experimental results. Note that STVIAE and STVIIE barely meet our cut-off for fibril formation at low temperatures. As we increase the temperature to intermediate ranges we observe that STVIIE and STVIFE form fibrils as expected, STVIGE and STVIEE do not form fibrils as expected, and STVIVE, STVIAE, and STAIIE form fibrils in contradiction to the experimental results. Note that at intermediate temperatures, we observe an increase in % peptides in fibrils for STVIIE, STVIVE, STVIAE, and STAIIE. At high temperatures, however, we only observed fibrils for STVIIE, STVIFE (as seen experimentally) and for STVIVE which was not observed to form fibrils *in vitro*; STVIAE, STAIIE, STVIGE and STVIEE do not form fibrils. Figure 4.1a shows a snapshot of STVIIE at T*=0.16 and c=10mM and Figure 4.7 show a snapshot of STVIEE at T*=0.16 and c=10mM. Figure 4.7 clearly shows the STVIEE peptides forming β-sheets but these never come close to associating into fibrils,

in agreement with experiment. Figure 4.1a shows a very well-ordered fibril formed by the STVIIE peptides. These snapshots clearly illustrate that sequence influences fibril formation and structure.

In comparing our simulation results to experiments in Table 4.6, it is appropriate to ask which of the three reduced temperatures that we considered are most likely to correspond to the experimental conditions of Lopez de la Paz and Serrano. The agreement between simulation and experiment is best at the highest temperature, in which case our results on all sequences considered agree with experiment except for STVIVE. We think that this correspondence makes sense based on the following. In other simulation work performed in our group on polyalanine [48] and on Aβ(16-22) - we have found that fibrils are most likely to form at a "marginal" temperature, above which the system forms random coils. The marginal temperatures are the highest temperatures listed in the table. Although this interpretation of our results may seem gratuitous, we feel that it has a good scientific basis. For example, Nguyen and Hall found that for a simple polyalanine peptide system there is in fact a range of temperatures where fibrils form for a given concentration and that above that temperature fibrils are not observed. [48].

Our results also make good physical sense based on the types of residues that were substituted at position 5 on STVIIE. Clearly, substituting a bulky aromatic group like F in STVIFE for a charged residue like E in STVIEE should influence the ability of

the sequence to form fibrils. Figure 4.8 is a simulation snapshot of STVIFE at T*=0.16 and c=10mM shown as a ribbon diagram for clarity. This fragment forms an extremely well-ordered fibril which is reminiscent of the first x-ray diffraction models of fibrils. Furthermore, even a single glycine positioned strategically in STVIGE can dramatically reduce fibril formation. All of the *de novo* designed peptides tend to favor β-sheets regardless of ability to form fibrils. It is interesting to note that none of the *de novo* designed sequences form amorphous aggregates or undergo any amorphous transitions in the early stages of the simulation (data not shown).

Finally, it is important to point out that the single-bead geometry of PRIME20 may not be robust enough to distinguish between single residue substitutions of small hydrophobic groups even though alanine-alanine interactions are considered weak in our model. In other words, STVIVE and STVIAE would have a similar probability of forming fibrils as STVIIE.

4.5 Conclusions

Recent improvements in our intermediate-resolution protein model, PRIME20, have enabled us to observe the spontaneous formation of amyloid fibrils starting from an initial configuration of random coils. We were able to observe spontaneous fibril formation for fragments of amyloidogenic peptides like Aβ and for the designed amyloid peptide STVIIE and some of its substitutional mutants. The short amyloid peptide

fragments SSTSAA and SNQNNF and MVGGVV all form fibrils with structures that are in qualitative agreement with experiments, although further investigation into strand arrangement is necessary. For GGVVIA we observed β-sheet formation and for VEALYL we observed amorphous aggregtates suggesting further exploration of temperature is needed to determine if our model will capture fibril formation of these two sequences. At the highest temperatures considered, our results on the fibrillization propensity of six of the seven short *de novo* designed peptides based on STVIIE agree with experiment. Although our results are not a perfect match for the *de novo* designed peptides experimental results we believe that our simulations at higher reduced temperatures on in qualitative agreement.

4.6 Acknowledgements

The authors are grateful to Dr. Mookyung Cheon for his development of PRIME20 and many thoughtful discussions on "real" protein behavior. Special thanks to Miss Erin Phelps whose kinetic model of aggregation helped to clarify the fibrillization process and Mr. Johnny Maury turning on the light when everything was dark. All of the simulation snapshots in this paper were generated using Visual Molecular Dynamics developed at the University of Illinois Urbana-Champaign and Discovery Studio by Accelrys.

4.7 References

1. Sunde, M., et al., *Common core structure of amyloid fibrils by synchotron x-ray diffraction.* J. Mol. Biol., 1997. **273**: p. 729.

2. Sunde, M. and C.C.F. Blake, *The structure of amyloid fibrils by electron microscopy and x-ray diffraction.* Adv. Protein Chem., 1997. **50**: p. 123.

3. Serpell, L.C., *Alzheimers amyloid fibrils: Structure and assembly.* Biochimica et Biophysica Acta, 2000. **1502**: p. 16.

4. Fraser, P.E., et al., *Fibril formation by primate, rodent, and dutch-hemorrhagic analogues of alzheimer amyloid beta-protein.* Biochem, 1992. **31**: p. 10716.

5. Blake, C. and L.C. Serpell, *Synchotron x-ray studies suggest that the core of transthyretin amyloid fibril is a continuous β-sheet helix.* Structure, 1996. **4**: p. 989.

6. Chiti, F., et al., *Solution conditions can promote formation of either amyloid protofilaments or native fibrils from the hypf n-terminal domain.* Prot.Sci., 2001. **10**: p. 2542.

7. Caughey, B. and P.T. Lansbury, *Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders.* Annu. Rev. Neurosci., 2003. **26**: p. 267-298.

8. Kelly, J., *Amyloid fibril formation and protein misassembly: A structural quest for insights into amyloid and prion diseases.* Structure, 1997. **5**: p. 595-600.

9. Kelly, J.W., *The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways.* Curr. Opin. Struct. Biol, 1998. **8**: p. 101.

10. Rochet, J.C. and P.T. Lansbury, *Amyloid fibrilogenesis: Themes and variations.* Curr. Opin. Struct. Biol., 2000. **10**: p. 60.

11. Hou, L. and M.G. Zagorski, *Sorting out the driving forces for parallel and antiparallel alignment in the aβ peptide fibril structure.* Biophys J, 2004. **86**: p. 1-2.

12. Petkova, A.T., W.-M. Yau, and R. Tycko, *Experimental constraints on quaternary structure in alzheimer's beta-amyloid fibrils.* Biochem, 2006. **45**: p. 498-512.

13. Nelson, R., et al., *Structure of the cross-beta spine of amyloid-like fibrils.* Nature, 2005. **435**(7043): p. 773-8.

14. Sawaya, M.R., et al., *Atomics structures of amyloid cross-beta spines reveal varied steric zippers.* Nature, 2007. **447**: p. 453-457.

15. Lopez de la Paz, M. and L. Serrano, *Sequence determinants of amyloid fibril formation.* Proc Natl Acad Sci U S A, 2004. **101**(1): p. 87-92.

16. Jobling, M.F., et al., *The hydrophobic core sequence modulates the neurotoxic and secondary structure properties of the prion peptide 106-126.* J Neurochemistry, 1999. **73**(4): p. 1557-1565.

17. Chiti, F., et al., *Designing conditions for in vitro formation of amyloid protofilaments and fibrils.* Proc. Natl. Acad. Sci. USA, 1999. **96**: p. 3590.

18. Antzutkin, O.N., et al., *Multiple quantum solid-state nmr indicates a parallel, not antiparallel, organization of beta-sheets in alzheimer's beta-amyloid fibrils.* Proc. Natl. Acad. Sci. USA, 2000. **97**: p. 13045.

19. Balbach, J.J., et al., *Amyloid fibril formation by aβ 16-22, a seven-residue fragment of the alzheimer's beta-amyloid peptide, and structural characterization by solid state nmr.* Biochem., 2000. **39**: p. 13748.

20. Benzinger, T.L., et al., *Two-dimensional structure of beta-amyloid(10-35) fibrils.* Biochem., 2000. **39**: p. 3491.

21. Teng, P.K. and D. Eisenberg, *Short protein segments can drive a non-fibrillizing protein into the amyloid state.* Protein Engineering and Design, 2009. **22**(8): p. 531-536.

22. Tycko, R., *Molecular structure of amyloid fibrils: Insights from solid-state nmr.* Quarterly Reviews in Biophysics, 2006: p. 1-55.

23. Makin, O.S., et al., *Molecular basis of amyloid fibril formation and stability.* Proc Natl Acad Sci U S A, 2005. **102**(2): p. 315-320.

24. Lopez de la Paz, M., et al., *De novo designed peptide-based amyloid fibrils.* Proc Natl Acad Sci USA, 2002. **99**: p. 16052.

25. Zheng, J., et al., *Structural stability and dynamics of an amyloid-forming peptide gnnqqny from the yeast prion sup-35.* Biophys J, 2006. **91**: p. 824-833.

26. Zheng, J., et al., *Modeling the alzheimer aβ_{17-42} fibril architecture: Tight intermolecular sheet-sheet association and intramolecular hydrated cavities.* Biophys J, 2007. **93**: p. 3046-3057.

27. Lopez de la Paz, M., et al., *Sequence dependence of amyloid fibril formation: Insights from molecular dynamics simulations.* Journal of Molecular Biology, 2005. **349**: p. 583-596.

28. Park, J., B. Kahng, and W. Hwang, *Thermodynamics selection of steric zipper patterns in the amyloid cross-beta spine.* PLoS Computational Biology, 2009. **5**(9): p. 1-17.

29. Ma, B. and R. Nussinov, *Stabilities and conformations of alzhemier's β-amyloid peptide oligomers (aβ16-22, aβ16-35, and aβ10-35):Sequence effects.* Proc Natl Acad Sci USA, 2002. **99**: p. 14126.

30. De Simone, A., C. Pedone, and L. Vitagliano, *Structure, dynamics and stability of asssemblies of the human prion fragment snqnnf.* Biochemical and Biophysical Research Communications, 2008. **366**: p. 800-806.

31. Vitagliano, L., et al., *Dynamics and stability of amyloid-like steric zipper assemblies with hydrophobic dry interfaces.* Biopolymers, 2009. **91**(12): p. 1161-1171.

32. Chang, L.-K., et al., *Molecular dynamics simulations to investigate the structural stability and aggregation behavior of the ggvvia oligomers derived from amyloid beta peptide.* Journal of Biomolecular Structure and Dynamics, 2009. **26**(6): p. 731-740.

33. Lin, Y.-F., et al., *Structural stability and aggregation behavior of the vealyl peptide derived from human insulin: A molecular dynamics simulation study.* Peptide Science, 2009. **94**(3): p. 269-278.

34. Cheon, M., I. Chang, and C.K. Hall, *Extending the prime model for protein aggregation of all twenty amino acids.* PRoteins, 2010. **in print**.

35. Smith, A.V. and C.K. Hall, *Assembly of a tetrameric α-helical bundle: Computer simulations on an intermediate-resolution protein model.* Proteins: Structure, Function and Genetics, 2001. **44**: p. 376.

36. Smith, A.V. and C.K. Hall, *Protein refolding versus aggregation: Computer simulations on an intermediate resolution model.* J. Mol. Biol., 2001. **312**: p. 187.

37. Nguyen, H.D. and C.K. Hall, *Molecular dynamics simulations of spontaneous fibril formation by rando-soil peptides.* Proc Natl Acad Sci USA, 2004. **101**(46): p. 16180-16185.

38. Nguyen, H.D., A.J. Marchut, and C.K. Hall, *Solvent effects on the conformational transition of a model polyalanine peptide.* Prot. Sci., 2004. **13**(11): p. 2909-2924.

39. Marchut, A.J. and C.K. Hall, *Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations.* Biophys J, 2006. **90**(12): p. 4574-84.

40. Marchut, A.J. and C.K. Hall, *Spontaneous formation of annular structures observed in molecular dynamics simulations of polyglutamine peptides.* Comput Biol Chem, 2006. **30**(3): p. 215-8.

41. Smith, A.V. and C.K. Hall, *A-helix formation: Discontinuous molecular dynamics on an intermediate resolution model.* Protein: Structure, Function and Genetic, 2001. **44**: p. 344.

42. Bellemans, A., J. Orbans, and D.V. Belle, *Molecular dynamics of rigid and non-rigid necklaces of hard disks.* Mol. Phys., 1980. **39**: p. 781-782.

43. Alder, B.J. and T.E. Wainwright, *Studies in molecular dynamics, i: General method.* J Chem Phys, 1959. **31**: p. 459-466.

44. Smith, S.W., C.K. Hall, and B.D. Freeman, *Molecular dynamics for polymeric fluids using discontinuous potentials.* Journal of Computational Physics, 1997. **134**: p. 16.

45. Rapaport, D.C., *Molecular dynamics study of polymer chains.* J. Chem. Phys., 1979. **71**: p. 3299.

46. Rapaport, D.C., *Molecular dynamics simulation of polymer chains with excluded volume.* J. Phys. A, 1978. **11**: p. L213.

47. Andersen, H.C., *Molecular dynamics simulation at constant temperature and / or pressure.* J. Chem. Phys., 1980. **72**: p. 2384.

48. Nguyen, H.D. and C.K. Hall, *Phase diagrams describing fibrillization by polyalanine peptides.* Biophys J, 2004. **87**(6): p. 4122-4134.

4.8 List of Tables

4.9 List of Figures

Table 4. 1 PRIME20 Geometric Parameters for All Twenty Amino Acids

Identity	Abbr.	R-C_α bond	σ	m		Identity	Abbr.	R-C_α bond	σ	m
Glycine	G	0.654	3.10	0.34		Aspartate	D	2.500	3.40	3.86
Alanine	A	1.600	2.70	1.00		Asparagine	N	2.500	3.30	3.86
Valine	V	2.002	3.30	2.87		Glutamate	E	3.180	3.20	4.79
Leucine	L	2.625	3.40	3.80		Glutamine	Q	3.400	3.60	4.80
Isoleucine	I	2.440	3.30	3.80		Lysine	K	3.550	3.50	4.87
Serine	S	1.960	2.50	2.06		Arginine	R	4.700	3.20	6.73
Threonine	T	1.980	2.90	3.00		Histidine	H	3.150	3.40	5.39
Cysteine	C	2.350	2.10	3.13		Phenylalanine	F	3.425	3.30	6.06
Methionine	M	3.400	3.70	5.00		Tyrosine	Y	3.850	3.00	7.13
Proline	P	1.920	3.10	2.80		Tryptophan	W	4.200	3.70	8.66

Table 4. 2 PRIME20 Energetic Parameters for All Twenty Amino Acids

	G	A	V	L	I	S	T	C	M	P	D	N	E	Q	K	R	H	F	Y	W
G	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034	.034
A	.034	.086	.141	.141	.141	.020	.020	.035	.141	.020	.020	.020	.020	.020	.020	.020	.020	.141	.141	.141
V	.034	.141	.198	.198	.198	.027	.027	.035	.198	.027	.027	.027	.027	.027	.027	.027	.027	.207	.207	.207
L	.034	.141	.198	.198	.198	.027	.027	.035	.198	.027	.027	.027	.027	.027	.027	.027	.027	.207	.207	.207
I	.034	.141	.198	.198	.198	.027	.027	.035	.198	.027	.027	.027	.027	.027	.027	.027	.027	.207	.207	.207
S	.034	.020	.027	.027	.027	.055	.055	.152	.152	.020	.055	.055	.055	.055	.055	.055	.055	.027	.055	.055
T	.034	.020	.027	.027	.027	.055	.055	.152	.152	.020	.055	.055	.055	.055	.055	.055	.055	.027	.055	.055
C	.034	.035	.035	.035	.035	.152	.152	.434	.035	.027	.152	.152	.152	.152	.152	.152	.152	.035	.152	.152
M	.034	.141	.198	.198	.198	.152	.152	.035	.198	.027	.027	.152	.027	.152	.152	.152	.152	.207	.282	.282
P	.034	.020	.027	.027	.027	.020	.020	.027	.027	.020	.020	.020	.020	.020	.020	.020	.020	.027	.027	.027
D	.034	.020	.027	.027	.027	.055	.055	.152	.027	.020	.215	.055	.215	.055	.139	.139	.055	.027	.055	.055
N	.034	.020	.027	.027	.027	.055	.055	.152	.152	.020	.055	.075	.055	.075	.055	.055	.075	.027	.055	.055
E	.034	.020	.027	.027	.027	.055	.055	.152	.027	.020	.215	.055	.215	.055	.139	.139	.055	.027	.055	.055
Q	.034	.020	.027	.027	.027	.055	.055	.152	.152	.020	.055	.075	.055	.075	.055	.055	.075	.027	.055	.055
K	.034	.020	.027	.027	.027	.055	.055	.152	.152	.020	.139	.055	.139	.055	.017	.017	.055	.027	.055	.027
R	.034	.020	.027	.027	.027	.055	.055	.152	.152	.020	.139	.055	.139	.055	.017	.017	.055	.027	.055	.027
H	.034	.020	.027	.027	.027	.055	.055	.152	.152	.020	.055	.075	.055	.075	.055	.055	.075	.027	.055	.055
F	.034	.141	.207	.207	.207	.027	.027	.035	.207	.027	.027	.027	.027	.027	.027	.027	.027	.216	.216	.216
Y	.034	.141	.207	.207	.207	.055	.055	.152	.282	.027	.055	.055	.055	.055	.055	.055	.055	.216	.243	.243
W	.034	.141	.207	.207	.207	.055	.055	.152	.282	.027	.055	.055	.055	.055	.027	.027	.055	.216	.243	.216

Table 4. 3a Percentage of each sequence in monomer, β-sheet, amorphous, and fibril at low temperatures.

	T*	%Mono	%Beta	%Amor	%Fibril	Structure
STVIIE	0.13	1.8	48.2	2.0	48.0	Ordered
VEALYL	0.13	---	---	---	---	Still running
MVGGVV	0.13	---	---	---	---	Still running
SSTSAA	0.13	0.0	64.5	0.8	34.8	Ordered
SNQNNF	0.13	0.0	48.8	0.0	51.1	Ordered
GGVVIA	0.10	2.1	94.5	3.4	0.0	Slightly Ordered

Table 4. 3 b Percentage of each sequence in monomer, β-sheet, amorphous, and fibril at intermediate temperatures.

	T*	%Mono	%Beta	%Amor	%Fibril	Structure
STVIIE	0.15	1.0	39.3	1.9	57.8	Ordered
VEALYL	0.15	0.8	18.9	62.0	18.3	Slightly Disordered
MVGGVV	0.15	0.8	29.7	9.1	60.3	Ordered
SSTSAA	0.15	6.5	90.3	2.3	0.9	Slightly Ordered
SNQNNF	0.15	1.3	79.0	0.4	19.3	Slightly Ordered
GGVVIA	0.12	8.5	89.0	2.5	0.0	Slightly Ordered

Table 4. 3 c Percentage of each sequence in monomer, β-sheet, amorphous, and fibril at high temperatures.

	T*	%Mono	%Beta	%Amor	%Fibril	Structure
STVIIE	0.175	42.0	22.6	9.8	25.7	Slightly Ordered
VEALYL	0.18	3.8	16.3	76.2	3.7	Disordered
MVGGVV	0.17	10.5	38.5	41.5	9.5	Slightly Disordered
SSTSAA	0.16	27.8	60.2	10.6	1.4	Slightly Ordered
SNQNNF	0.16	14.9	52.1	2.2	30.9	Slightly Ordered
GGVVIA	0.13	18.3	76.8	4.9	0.1	Slightly Ordered

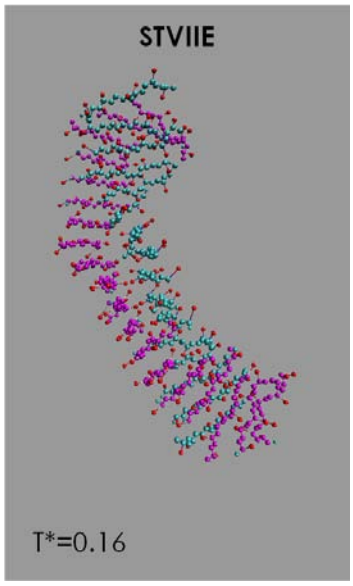

Figure 4. 1a Simulation snapshot of fibril formed by 48 peptides of STVIIE at T*=0.16 and c=10mM.

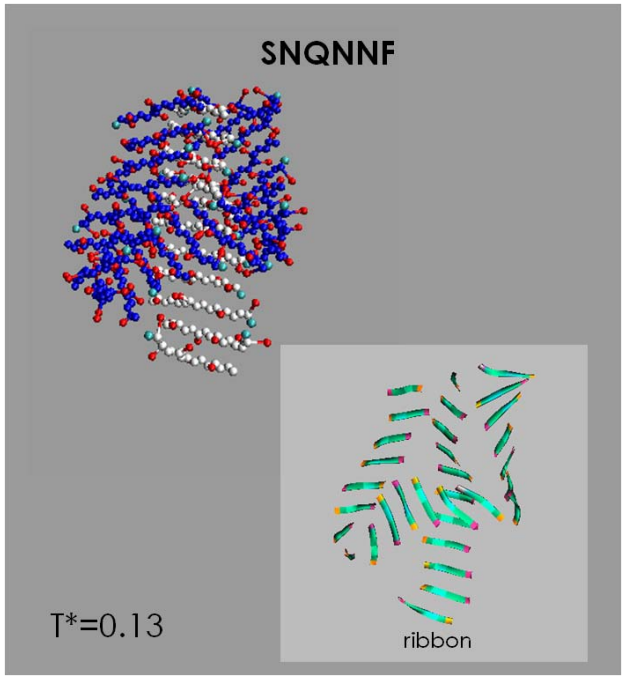

Figure 4. 1 b Simulation snapshot of fibril formed by 48 peptides of STVIIE at T*=0.16 and c=10mM.

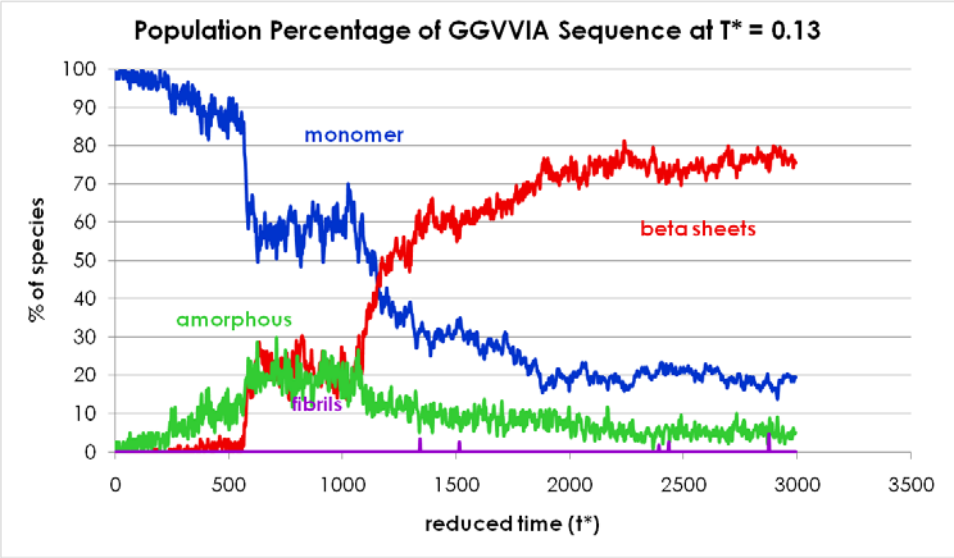

Figure 4. 2 Population of each species: monomer, β-sheet, amorphous, and fibril for GGVVIA at T*=0.13

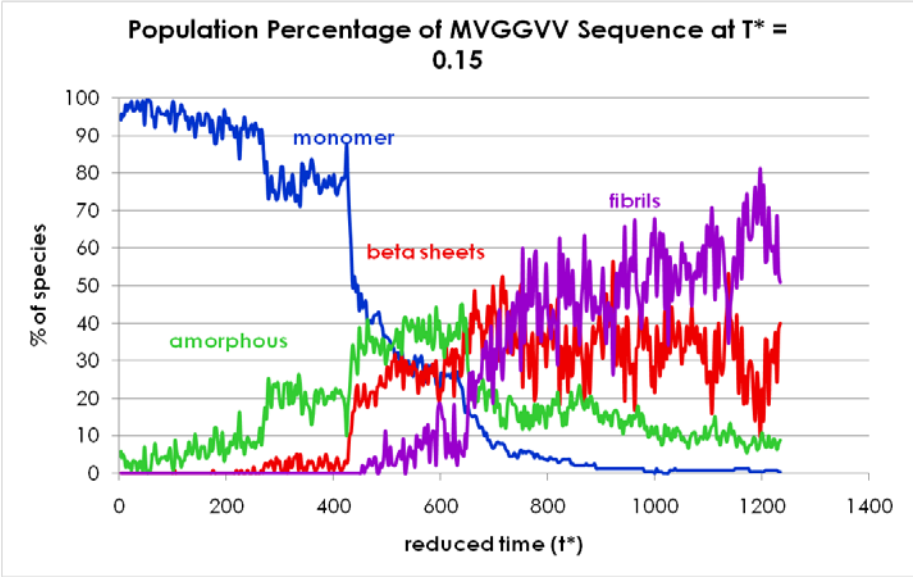

Figure 4. 3 Population of each species: monomer, β-sheet, amorphous, and fibril for MVGGVV at T*=0.15

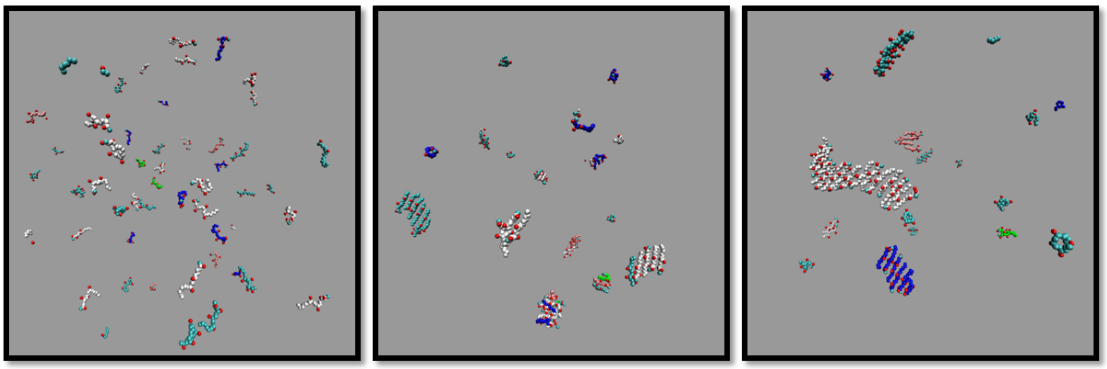

Figure 4. 4 Simulation snapshots of 48 peptides of GGVVIA at T*=0.13 and c=10mM

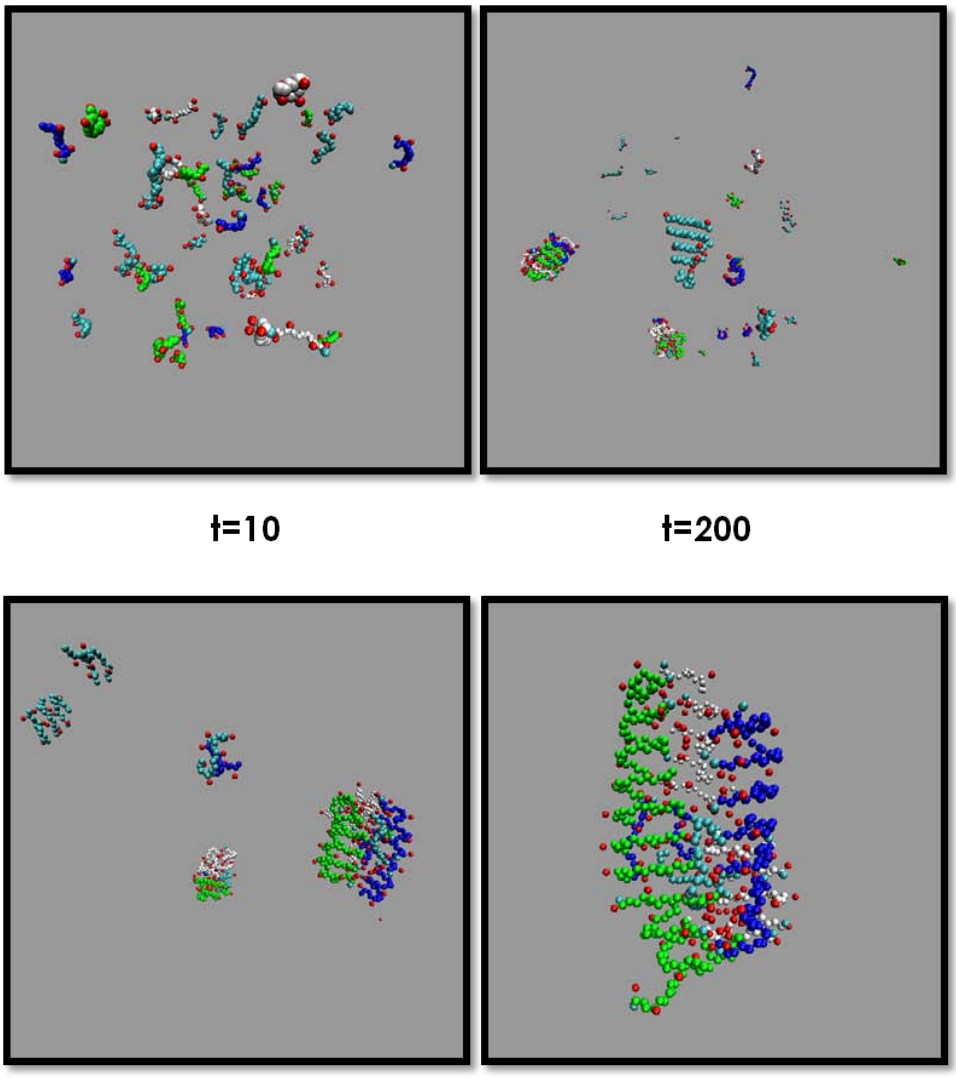

Figure 4. 5 Simulation snapshots of 48 peptides of MVGGVV at T*=0.15 and c=10mM

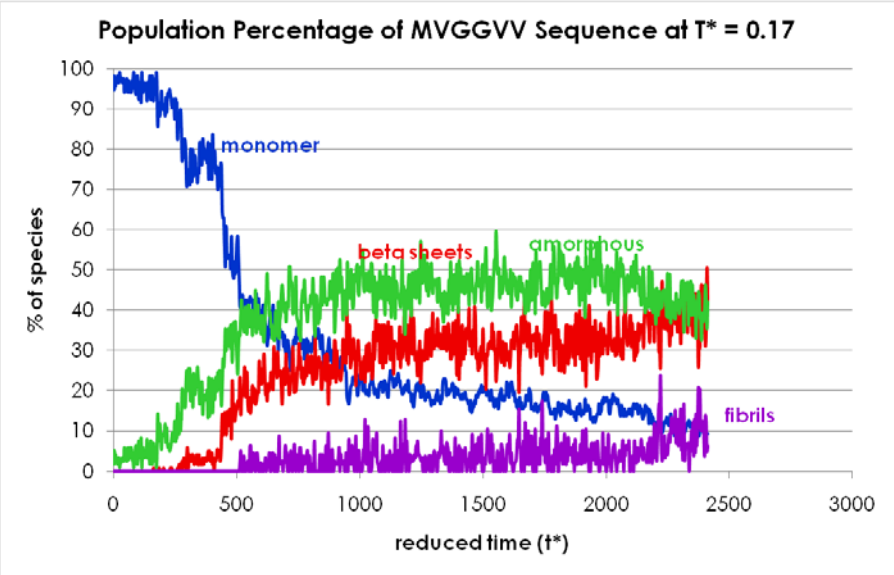

Figure 4. 6 Population of each species: monomer, β-sheet, amorphous, and fibril for MVGGVV at T*=0.17

Table 4. 4 Molecular arrangement of peptides in β-sheets formed by STVIFE, STVIIE, VEALYL, MVGGVV, SSTSAA, and SNQNNF at the temperature corresponding to highest structural order.

	STVIFE	STVIIE	VEALYL	MVGGVV	SSTSAA	SNQNNF
T*	0.16	0.16	0.15	0.15	0.13	0.13
Parallel	0.33	0.42	0.68	0.48	0.56	0.47
Anti-parallel	0.67	0.56	0.23	0.31	0.39	0.52
Ratio (P:A)	0.49	0.75	2.96	1.55	1.44	0.90

Table 4. 5 Table describing the experimental results of Lopez de la Paz and Serrano versus PRIME20 simulations at low, intermediate and high T* for each sequence. The * on at high T* for STVIAE indicates that simulations on that sequence at higher temperatures were not complete at the time of printing.

De Novo Designed Sequence	Observed Fibril Growth *in vitro*	%Fibrils observed at low T* using PRIME20	%Fibrils observed at intermediate T* using PRIME20	%Fibrils observed at high T* using PRIME20
STVIIE	Yes	48.0	63.1	67.6
STVIFE	Yes	49.1	49.7	64.0
STVIVE	No	41.5	77.8	79.4
STVIAE	No	40.4	49.3	*
STAIIE	No	50.4	59.2	19.7
STVIGE	No	24.5	11.1	1.3
STVIEE	No	10.7	0.3	0.0

Table 4. 6a Percentage of each sequence in monomer, β-sheet, amorphous, and fibril at low temperatures.

	T*	%Mono	%Beta	%Amor	%Fibril	Structure
STVIIE	0.13	1.8	48.2	2.0	48.0	Ordered
STAIIE	0.13	0.0	48.8	0.8	50.4	Ordered
STVIAE	0.13	0.7	58.9	0.0	40.4	Ordered
STVIVE	0.13	0.0	56.7	1.8	41.5	Ordered
STVIFE	0.13	0.0	48.6	2.3	49.1	Ordered
STVIGE	0.13	1.2	71.9	2.4	24.5	Slightly Ordered
STVIEE	0.13	1.5	86.8	1.0	10.7	Slightly Ordered

Table 4. 6b Percentage of each sequence in monomer, β-sheet, amorphous and fibril at intermediate temperatures.

	T*	%Mono	%Beta	%Amor	%Fibril	Structure
STVIIE	0.155	1.1	34.2	1.5	63.1	Ordered
STAIIE	0.155	6.5	33.3	1.0	59.2	Ordered
STVIAE	0.155	2.8	47.2	0.7	49.3	Ordered
STVIVE	0.155	1.5	20.2	0.6	77.8	Ordered
STVIFE	0.155	1.2	47.2	2.0	49.7	Ordered
STVIGE	0.145	6.3	81.5	1.0	11.1	Slightly Ordered
STVIEE	0.15	14.7	83.3	1.7	0.3	Slightly Ordered

Table 4. 6c Percentage of each sequence in monomer, β-sheet, amorphous and fibril at high temperatures.

	T*	%Mono	%Beta	%Amor	%Fibril	Structure
STVIIE	0.17	7.8	21.7	2.9	67.6	Ordered
STAIIE	0.165	22.9	50.3	7.1	19.7	Slightly Ordered
STVIAE	0.17	---	---	---	---	still running
STVIVE	0.17	7.0	12.2	1.3	79.4	Ordered
STVIFE	0.175	15.0	18.3	2.7	64.0	Ordered
STVIGE	0.15	16.0	80.3	2.4	1.3	Slightly Ordered
STVIEE	0.16	29.1	66.4	4.5	0.0	Slightly Ordered

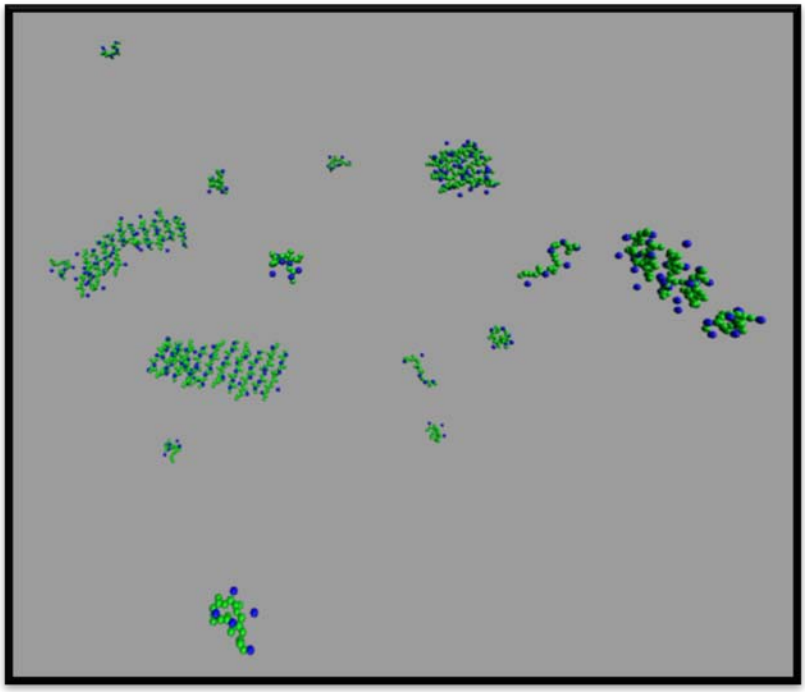

Figure 4. 7 Snapshots of STVIEE at T***=0.16** and c=10mM

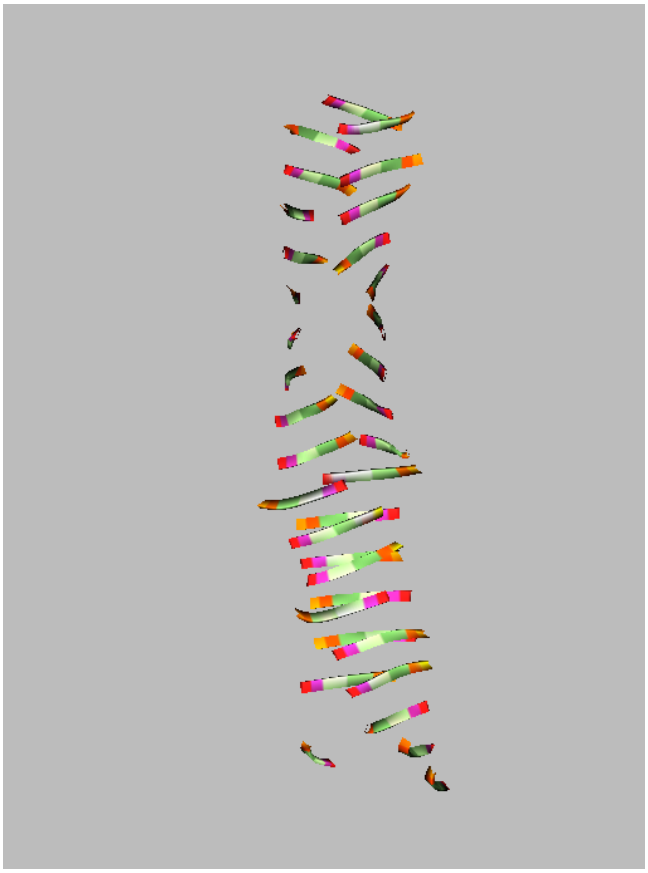

Figure 4. 8 Simulation snapshot at STVIFE at T*=0.16 and c=10mM.

CHAPTER 5

Future Work

The computer simulations presented in this thesis are the first application of our new implicit solvent force field PRIME20 (*Pr*otein *I*ntermediate Resolution *M*od*e*l). PRIME20 is exceptionally powerful in that when combined with discontinuous molecular dynamics (DMD) it allows us to simulate the aggregation and possible fibrillization of large systems of proteins composed of all twenty amino acids starting from an initial configuration of random coils. In most cases, the spontaneous assembly of these peptides has never before been observed in a computer simulation. We have attempted to characterize the early steps of the aggregation pathway for several peptides that are known to form amyloid, to understand the role that primary sequence plays in the amyloid fibril process, and to see if simulations with our simple model can distinguish the differences between sequences.

There are many possibilities for future work on the simulation of fibril-forming peptides including: (1a) improving the representation of the side chain geometry in PRIME20 to better capture the details of each specific side chain's functional group including the ability of certain amino acids to adopt multiple low energy conformations,

(1b) refining the discontinuous potential to better describe charge-charge interactions, (1c) incorporating pH changes into our implicit-solvent model by including the change in ionization that occurs for certain amino acids as the pH is raised or lowered, (2) simulating full-length amyloidogenic peptides like Aβ1-40 and Aβ1-42, and finally (3) simulating the impact of inhibitors on the fibrillization of certain peptides.

5.1a Improving the representation of protein geometry in PRIME20

There are potentially many avenues to take in improving the representation of protein geometry in PRIME20. For example, earlier work in our group focused on a more realistic description of polyglutamine peptides in which four spheres were used to model the glutamine side chain [1]. The advantage of having this level of detail is that we can more accurately capture the physics of the side chain However, increasing the level of detail increases the costs in terms of computational time. There is also some concern that increasing the detail of the side chain might, for example, prohibit the molecule from readily adopting its lowest energy conformation, causing it to become kinetically trapped in a metastable intermediate along the fibrillization pathway. A benefit to continuing with a single or at most a two sphere side chain is that the code would still be computationally fast. Ideally, we would like to distinguish between the rotamers for each side chain, i.e. each side chain's lowest energy conformation [2-3]. Currently in PRIME20 the distance between the model side chain and the peptide

backbone is taken from Protein Data Bank PDB values for the distance between the side chain's center of mass and peptide backbone; the energetic parameters of the system are determined in part by the position of the side chain. A new set of energy parameters would need to be calculated if a new geometrical representation is chosen. (1)

5.1b Refinement of the discontinuous potentials used to describe particularly complex side chains

Currently in PRIME20 we treat all side chains with attractive potentials as square-wells, side chains with repulsive potentials (like two positively charged side chains) as square shoulders and all other side chains as hard spheres. For the charged interactions it would be relatively easy to add the two- or three- step discontinuous potential model [4] developed in our group for the study of colloid particles. This would allow us to better mimic the distance dependence of charge-charge interactions such as screening by ionic solutions, and to better capture the effect of water in our implicit-solvent model.

5.1c Incorporation of solution conditions into PRIME20 by modeling changes in pH

Recent evidence suggests that the net charge of ±1 of the peptide sequence plays an important role in determining whether or not the sequence will aggregate into amyloid

fibrils [5-6]. This indicates that somehow we have to be able to include the effect of pH on our system to study different environments and also to explore the role that overall net charge may play in the fibrillization process. Since the effects of pH on the charge of the side chains of amino acids are well understood, a simple model can easily be put in place so that if the system is at a pH below the pKa for a certain amino acid , then the previously charged atom becomes protonated and no longer participates in charge-charge interactions.

5.2 Application of PRIME20 to full length amyloidogenic peptides

Although recent evidence suggests that it is short sequences within the full-length amyloidogenic peptides that drive the fibrillization process, it is unclear exactly how the other "non-sticky" segments in the full-length peptide affect the aggregation pathway. Further refinements of PRIME20 (discussed above) coupled with experimental knowledge of the fibril structure may allow us to simulate the aggregation of Aβ9-40 and Aβ9-42. We may have to exploit other simulation techniques such as parallel tempering to ensure that these longer peptides do not become trapped in metastable states, preventing us from accessing the fibrillization process. This would be the ultimate test of PRIME20's abilities. It may prove to be difficult even with our simple protein model and DMD due to the limits of current computational resources, but we won't know until we try.

5.3 Understanding the fibrillization process and its applications

As our understanding of the fibrillization process becomes clearer through simulations and comparison with experiment, our next goal would be to use simulation to learn how to inhibit or prevent fibrillization. Currently, the identity of the toxic species is still unclear. We don't know if it is the soluble spherical off-pathway oligomer or the amyloid fibril. We also don't know if sequestering these "misbehaving proteins" into amyloid plaques does more harm than good, but it is well-known that Aβ accumulation in the brain is toxic whether it be the soluble Aβ intermediates or the Aβ fibrils [7-8]. A simple way to test if certain "drugs" could inhibit the aggregation of Aβ would be to introduce known inhibitors into simulations of "amyloidogenic" peptides. We could do this at various stages along the fibrillization pathway since we have access to the early stages of aggregation and to the "final" fibril structure formation. Inhibitors can be several different types of compounds: short peptides, small organic molecules, liposomal delivery vesicles and even DNA therapeutics. This would be an exciting endeavor because currently coworkers in our lab are developing models for lipids and DNA. The easiest first step would to model the inhibitors in the simplest manner as either a single sphere or a small chain of spheres with interaction energies estimated to mimic compounds known to inhibit amyloid fibril growth.

5.4 References

1. Marchut, A.J. and C.K. Hall, *Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations.* Biophys J, 2006. **90**(12): p. 4574-84.

2. Ding, F., S.V. Buldyrev, and N.V. Dokholyan, *Folding trp-cage to nmr resolution native structure using a coarse-grained protein model.* Biophysical Journal, 2005. **88**(1): p. 147-155.

3. Lovell, S.C., et al., *Structure validation by calpha geometry: &phis;,psi and cbeta deviation.* Proteins: Structure, Function, and Genetics, 2003. **50**(3): p. 437-450.

4. Goyal, A., C.K. Hall, and O.D. Velev, *Phase diagram for stimulus-responsive materials containing dipolar colloidal particles.* Physical Review E, 2008. **77**(3): p. 031401.

5. Lopez de la Paz, M., et al., *De novo designed peptide-based amyloid fibrils.* Proc Natl Acad Sci USA, 2002. **99**: p. 16052.

6. Chiti, F., et al., *Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(Suppl 4): p. 16419-16426.

7. Kadowaki, H., et al., *Amyloid [beta] induces neuronal cell death through ros-mediated ask1 activation.* Cell Death Differ, 2005. **12**(1): p. 19-24.

8. Ferreira, S.T., M.N.N. Vieira, and F.G.D. Felice, *Soluble protein oligomers as emerging toxins in alzheimer's and other amyloid diseases.* IUBMB Life, 2007. **59**(4-5): p. 332-345.