UNDERSTANDING AMYLOID FIBRIL GROWTH THROUGH THEORY AND SIMULATION

by

ALEX BEUGELSDIJK

B.S. Civil and Environmental Engineering, University of Illinois Urbana-Champaign, 2012

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Approved by:

Major Professor
Jianhan Chen

Copyright

Abstract

Proteins are fundamental building blocks of life in an organism, and to function properly, they must adopt an appropriate three-dimensional conformation or conformational ensemble. In protein aggregation diseases, proteins misfold to incorrect structures that allow them to join together and form aggregates. A wide variety of proteins are involved in these aggregation diseases and there are multiple theories of their disease mechanism. However, a common theme is that they aggregate into filamentous structures. Therapies that target the process by which the aggregating proteins assemble into these similar fibril-like structures may by effective at countering aggregation diseases. This requires models that can accurately describe the assembly process of the fibrils. An analytical theory was recently described where fibrils grow by the templating of peptides onto an existing amyloid core and the kinetics of the templating process is modeled as a random walk in the backbone hydrogen bonding space. In this thesis, I present my work integrating molecular simulation with this analytical model to investigate the dependence of fibril growth kinetics on peptide sequence and other molecular details. Using the $A\beta_{16\text{-}22}$ peptide as a model system, we first calculate the rate matrix of transitions among all possible hydrogen bonding microscopic states using numerous short-time simulations. These rates were then used to construct a kinetic Monte Carlo model for simulations of long-timescale fibril growth. The results demonstrate the feasibility of using such a theory/simulation framework for bridging the significant gap between fibril growth and simulation timescales. At the same time, the study also reveals some limits of describing the fibril growth as a templating process in the backbone hydrogen bonding space alone. In particular, we found that dynamics in nonspecifically bound states must also be considered. Possible solutions to this deficiency are discussed at the end.

Table of Contents

List of Figures

List of Tables

Acknowledgements

For the indispensible guidance and support on this project, I would like to thank my advisor, Jianhan Chen; my committee members, Jeremy Schmit and Michael Zolkiewski; my fellow lab members, Weihong Zhang, Debabani Ganguly, Michael Mohan, Kuo Hao Lee, and Zhiguang Jia; and my family. I would also like to thank the Biochemistry and Molecular Biophysics Graduate Group and Kansas State University for providing me with this opportunity.

Dedication

I dedicate this thesis to my family.

Chapter 1 - Introduction

1.1 Protein Aggregation Diseases

1.1.1 Introduction

Proteins are an essential and diverse group of molecules in an organism with many different purposes, and the proper functioning of these molecules is essential for life. The three-dimensional conformational properties of a protein define its activity and are encoded by the sequence of its amino acids. Occasionally, proteins may misfold to improper structures, and misfolded proteins may further aggregate into large fibrils.[1-3] Intracellular and extracellular fibrillar aggregates in tissues are a characteristic feature of many protein misfolding diseases, which include the most common types of neurodegenerative diseases such as Alzheimer's and Parkinson's diseases.[1-3] The historical term amyloid is typically used to describe the extracellular deposits, however it occasionally refers to intra-cellular deposits as well.

1.1.2 Pathogenesis

Although aggregates appear in many protein conformational diseases, it is unclear whether they are the source of the disease themselves. The confusion stems from the fact that aggregates can be found in clinically healthy people and that there is limited correlation between the amount of aggregate deposits and the severity of disease symptoms.[4-6] Recent investigation suggests that the deposits are a mere side effect of aggregation disorders and that the actual cause of disease is the misfolding of proteins.[7] The evidence for this idea is inherited cases of aggregation diseases whose fibrillar proteins have an inherently higher degree of mutation,[8-12] which may destabilize the protein and cause it to misfold. Moreover, when human pathogenic, fibrillar proteins are expressed in animal models, disease symptoms are observed,[13-18] although there may be no detection of aggregates in some animals.[19, 20] These observations support the possibility that small aggregates or oligomers of misfolded proteins may be the driving force behind protein misfolding diseases.

1.1.3 Features of Amyloidogenic Proteins

Misfolded protein aggregates share similar structural features[21-23] even though the native conformation of their constituents may vary widely. Most aggregates in protein conformation disorders have a high amount of β-sheet secondary structure with the strands aligned perpendicular to the long axis of the fibril.[22, 24] To attain the β-sheet structure, there may be substantial conformational changes required of the native protein. From what is known, it is probable that the native protein experiences a slight change to its conformation, which exposes hydrophobic elements that are unstable in an aqueous environment.[2, 3] If two or more such proteins encounter one another, their hydrophobic parts may associate and form the β-sheet structures typical of an amyloid fibril.

The amyloidogenic Aβ peptide of Alzheimer's disease is one of the best-characterized proteins among those involved in the different aggregation diseases,[25] and the amyloid fibrils it forms *in vitro* are equivalent in a variety of experimental tests to plaques removed from Alzheimer's disease tissue.[26-29] A hydrophobic fragment of the Aβ peptide containing amino acid residues 17 through 21 has been to shown to be an important contributor to the initial phase of misfolding and oligomerization.[27, 29, 30] Moreover, the addition of hydrophobic amino acids to the C-terminus of the Aβ peptide has been shown to increase aggregation.[28] These two observations support the notion that amyloid formation is partially driven by hydrophobic interactions.

Polyglutamine-containing segments in proteins may also be a driving force towards amyloid formation.[31] As has been seen of the amyloidogenic protein in Huntington's disease, the tendency for the peptide to aggregate may be subject to the length of the polyglutamine repeat.[32] The polar amide group of the amino acid glutamine has both a hydrogen bond donor and acceptor and the capability to form hydrogen bonds. It is possible, therefore, that glutamine-rich peptides tend to aggregate because the side chain amide groups are forming a network of hydrogen bonds with each other, which has an overall stabilizing effect on the fibril that forms.[33] Thus, two different and somewhat opposite types of interactions can drive the formation of amyloid fibrils: the association of hydrophobic segments on peptides or the interconnection of polar hydrogen bonds among amino acid side chains.

1.1.4 Mechanisms of Misfolding and Aggregation

The occurrence of aggregation diseases is sporadic and one's susceptibility can be linked to both environmental and genetic causes. Genetic makeup that consists of many mutations in a gene encoding an amyloidogenic protein is associated with an increased chance of having a protein misfolding disease. The mutations destabilize the protein, which allows it to misfold more easily and then aggregate. There are several environmental factors that are thought to contribute to the misfolding of proteins,[3, 25] many of which are prevalent in old age, a circumstance supported by the typically late onset of aggregation diseases.[34]

The formation of large aggregates has been kinetically shown to follow an initial, slow nucleation event in which unfavorable interactions among amyloid molecules lead to an oligomeric structure that then undergoes rapid growth to long fibrils.[28, 35, 36] Adding pre-formed nuclei eliminates the slow nucleation phase. At least two intermediates are recognized in the build up to a fully developed amyloid fibril (Figure 1). The first is a soluble oligomer with transient, undefined structure composed of about two to ten monomers.[37-40] The second intermediate is a flexible, short rod-like structure less than 100 nm in length. These so-called protofibrils are in equilibrium with the oligomer intermediates,[41] and they can form by the coalescence of smaller protofibrils at a rate dependent on monomer concentration as well as solution properties like temperature, pH, and ionic strength.[42] Protofibrils have a high degree of β-sheet structure, which facilitates their development to the subsequent fibril state.[41] The intermediate structures, monomeric proteins, and amyloid fibrils all exist in dynamic equilibrium.[25, 39]

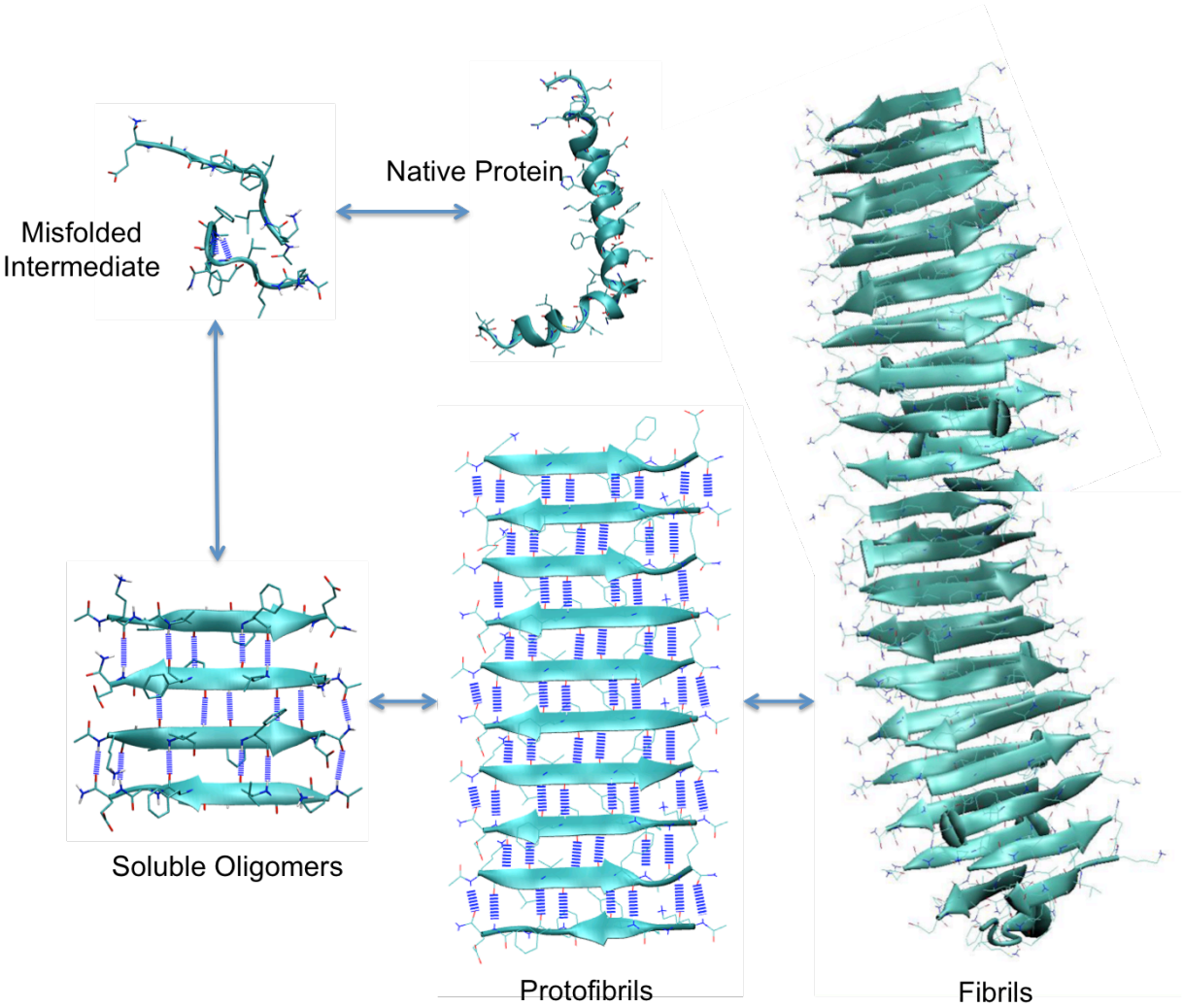


Figure 1 A Model of Protein Misfolding and Aggregation. The native protein assumes a random coil conformation with exposed hydrophobic elements. The misfolded intermediates are stabilized when they aggregate together in an oligomeric structure. Further monomers add to the complex to create protofibrils and then fibrils.

1.1.5 Therapy Considerations

In neurodegenerative aggregation diseases, the characteristic signs of neuronal death, synaptic modifications, and inflammation occur in widely varying parts of the brain, which affects what symptoms manifest for the disease.[43] Despite the disparity in affected locations among the diseases, protein misfolding and aggregation are key events in each disorder. Therefore, a potential, comprehensive therapy should target the causal protein misfolding events in the initiation of these diseases. A theoretical model that can accurately characterize amyloid

growth in accordance with experimental data would be a powerful tool in developing therapies. The research presented in this thesis aims to devise such a model.

1.2 Molecular Dynamics

1.2.1 Introduction

Molecular dynamics (MD) is an important theoretical tool for a wide range of applications, including the study of amyloid molecules. MD simulations are generally based on molecular mechanics (MM) models, which use empirical potential-energy functions to describe the interactions in a molecular system. Atoms are usually represented as spheres and bonds as springs. Hooke's Law can then describe bonded interactions. Nonbonded interactions may be described with a Leonard-Jones potential, while Coulomb's Law describes the interaction among charged atoms. The classical equations of motion can be applied to such a system and are solved numerically to calculate its trajectory.[44]

The trajectories produced by MD simulations can be used to extract information of conformational changes on time and spatial scales that are otherwise inaccessible experimentally. Computational experiment also allows the manipulation of the system in nonphysical ways to obtain insights on certain aspects of the system.[45] An example is switching the energy function describing one system to that of another during a simulation, which is an important method for free-energy calculations.[46]

1.2.2 Protein Dynamics

Proteins undergo a wide array of dynamic processes with time scales that range from femtoseconds to hours. These processes also have a range of physical and energy scales, and many are important for the proteins' biochemical function.[47] Fast and microscopic motions are typically prominent in enzymatic reactions, while large and slow motions on the scale of entire proteins are seen in allosteric coupling and folding transitions. Protein association takes place on even larger time and distance scales.

A primary purpose of MD simulation is to study time-dependent features of proteins and other molecules in general to better understand their molecular structure and function. However, the atomic position, velocity, and energy data of a simulation trajectory alone provides limited

useful information. To obtain the desired macroscopic information of the system, statistical mechanics is used to connect microscopic statistics and macroscopic properties, particularly thermodynamic quantities. Determination of these macroscopic properties from the microscopic MD data can then facilitate the study of a variety of molecular structure, kinetic, and thermodynamic aspects. A greater level of exploration of the sampling space by a system corresponds to a more thorough navigation of the free energy landscape and consequently more accurate macroscopic observables to be calculated.

1.2.3 Force Fields

The interparticle interactions within the classical mechanical framework are referred to as a force field, which is required as an input to the MD simulator. The force field provides parameters for all of the possible interactions present in the system and the accuracy of its representation may vary depending on the given parameters and system. The set of parameters in a force field is typically determined from experimental or quantum mechanical assessment of small molecules, and the resulting values are assumed to be valid for larger molecules. The low computational demands of a force field with explicit parameters to describe all interactions make it possible to perform the sizable calculations required to simulate large systems such as proteins, membranes, and other biological structures. The drawback of explicit force fields is that they cannot simulate the breaking or formation of bonds.

1.2.4 Implicit Solvation

For systems consisting of a solute in a bulk solution, the computational cost of calculating the interactions between many explicit solvent molecules can be prohibitive in simulating long time-scales or observing large conformational transitions.[48] Most of the computation is spent on calculating the trajectory of the solvent molecules even though the solute is the focus of study. So-called implicit solvation can reduce the computational load by capturing the solvent effects with the energy function through a potential of mean force term. Explicit and implicit solvents both have their strengths and weaknesses; explicit solvents can more accurately describe detailed water-protein interaction, while implicit solvents provide a good balance between accuracy and efficiency for simulating larger systems and longer timescales. One

method of modeling the solvent implicitly assumes that the mean solvation term is proportional to the solvent accessible surface area (SASA) of the solute and is discussed more in-depth later.[49]

1.2.5 MD Simulation of Amyloidogenic Proteins

MD has extensively been used to study processes related to protein aggregation. For example, coarse-grained simulations of polyalanine peptides in implicit solvent showed that they spontaneously formed fibrils above a concentration-dependent critical temperature above the peptide's folding temperature, which is consistent with experiment results.[50] In the study, amorphous structures formed initially before an ordered nucleus developed onto which incoming peptides formed β strands on either end of the growing fibril.[50] The fibril structure from simulation was found to be similar to that observed experimentally.[50] Yeast prion proteins were found to aggregate in in-register, parallel β strands through MD simulation, consistent with x-ray diffraction data.[51] A mutated sequence of the protein that did not aggregate in an experimental study was correspondingly observed not to oligomerize in simulation as well.[51]

There have already been many studies that used traditional MD simulations to investigate the Aβ peptide. MD simulation of the $A\beta_{16-22}$ peptide fragment introduced as the model system later in this thesis found that the peptide followed an assembly mechanism in which it passed from a random coil to an alpha helix to a beta strand in its path to fibril incorporation, a result that agrees with experimental studies of longer constructs of the same Aβ peptide.[52] In another study, different constructs of the Aβ peptide were simulated in explicit water to determine the relative stabilities of variously arranged β-sheet clusters.[53] The simulations showed that the $A\beta_{16-22}$ construct was most stable in an anti-parallel conformation, a finding in agreement with experimental studies.[54-57] The dock-lock mechanism, proposed by Thirumalai and coworkers, suggests a two-step process by which an amyloidogenic peptide assembles onto a fibril.[58] The first dock step occurs when the disordered monomer encounters the aggregate complex and rapidly experiences a dramatic increase in its β-strand content from an initial low value.[58] In the second lock step, the β-strand content continues to rise at a much slower rate until the peptide eventually assumes its optimal structure.[58] The analysis for this study was taken from simulations of confined volumes of explicit water in which randomly oriented, disordered $A\beta_{16-22}$ peptides were allowed to aggregate into antiparallel β-sheet structures.[58]

Although MD simulations are useful in studying microscopic, short-term behavior of amyloid systems, the current simulation time scale limit of approximately 1 millisecond is significantly smaller than the times at which large aggregate structures form, with symptoms of protein aggregation disease patients developing over the course of years. Hence, short time-scale MD cannot be used to directly study the prolonged growth of amyloid fibrils with today's computational capabilities. In response, this thesis presents an approach of combining MD simulations of an amyloid system with a simplified theory of amyloid fibril growth to predict long-term fibril formation kinetics and mechanism.

Chapter 2 - Integrating Theory and Computation to Study Amyloid Kinetics

2.1 Kinetic Theory of Amyloid Fibril Templating

Schmit recently proposed that the ordered β-sheet structure of a growing amyloid fibril, typically with in-register and parallel strands as observed from experimentally determined structures, could be described by the competition between amino acid sidechain specificity and backbone hydrogen bond periodicity.[59] Only the addition of monomeric proteins to a fibril (as opposed to the coalescence of oligomers) is considered in the theory since low physiologic protein concentrations make the former process far more likely; furthermore, the addition of single peptides to a fibril is simpler to model. Mis-registered or anti-parallel peptides would have a less favorable energy compared to the optimal conformation and would slow the formation of the highly ordered fibril structure. The time that a peptide is bound to the amyloid core can be described by a random walk of breaking and forming backbone hydrogen bonds. The theory predicts that the timescale of peptide release from the end of the fibril, regardless of optimal or sub-optimal conformation, is much greater than the time it takes for a peptide to fully bind or immediately break from the end of fibril given at least one contact. As such, the growth rate of the fibril can be approximated as

$$k_{grow} \simeq \frac{P_+}{(2L - 1)(t_{wait} + t_{diff})} \qquad (1)$$

where P_+ is the probability that an incoming, correctly aligned peptide becomes incorporated into the fibril in its optimal conformation, t_{wait} is the time that passes before an incorrectly bound

peptide releases, and t_{diff} is the diffusion time of the peptide. When compared to experimental results, this calculated growth rate is able to qualitatively capture the effects of concentration, temperature, and denaturants on fibril formation. A key limitation of this theory is that it cannot predict the sequence dependence of fibril growth, which requires information about the microscopic formation and breakage rates of backbone hydrogen bonds of specific residue types. These microscopic rates can be readily calculated from MD simulations. The purpose of the work presented in this thesis is to investigate if this theoretical framework can be combined with molecular simulations to predict and rationalize the dependence of amyloid fibril growth on amino acid sequence and other molecular details.

2.2 Computational Strategy

2.2.1 Overview

Inspired by the kinetic theory, our overall strategy consists of two main steps. In the first step, all possible hydrogen bonding registry states were enumerated and many short (~50 ns) MD simulations were performed, from which the mean first passage time of all possible conformational transitions was derived. In the second step, a Markov state model was employed to perform kinetic Monte Carlo simulations in determining the fibril growth kinetics.

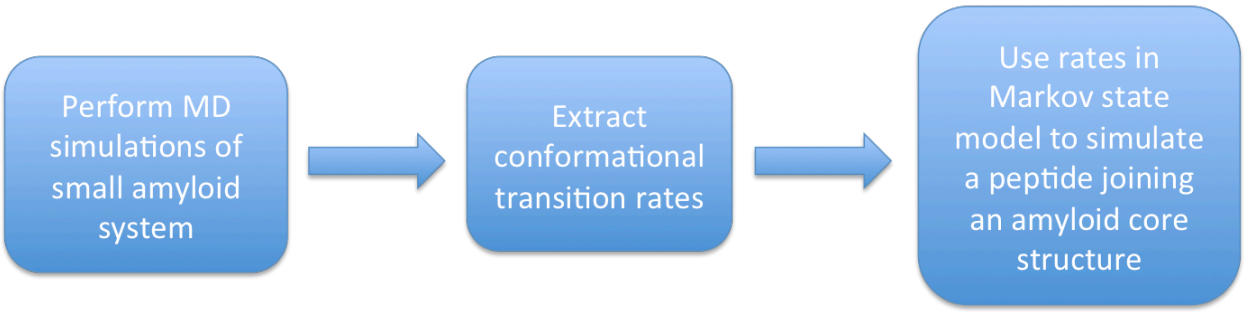


Figure 2 Computational Strategy Flowchart

2.2.2 Multi-Scale MD Simulations

In order to achieve adequate sampling of the many possible conformational transitions expected of a peptide on the end of an amyloid fibril, many short time-scale simulations were performed of a model system consisting of a small amyloid core with two incoming peptides on either side. Both incoming peptides were examined in subsequent analysis to effectively double

the sampling measured. In order to minimize computational time, the smallest stable system of a bilayer was used as the amyloid core onto which the two incoming peptides joined.

An incoming strand may form a β-strand on this bilayer in a number of conformations: in-register, mis-registered, parallel, or anti-parallel. Two end peptides on either side of the bilayer and on opposing sheets were manipulated to imitate an incoming strand joining to the amyloid core in all of these possible β-sheet conformations, or *configurations* as I have chosen to term each specific alignment that was simulated. Only anti-parallel β-strand configurations were considered since the model peptide (to be described in the next chapter) was observed to form anti-parallel fibrils in the experimental study,[60] so these configurations would be the most physiologically relevant. Also, it was reasoned that parallel β-sheet configurations would experience similar conformational transitions that could be understood and estimated by the data from anti-parallel simulations. In addition, only mis-registered configurations that had at least two backbone hydrogen bond pairs were considered to be simulated since any configuration with less than two pairs was assumed to be a transient conformation in the course of fibril growth and does not contribute significantly to t_{wait}.

To generate conformations with a particular configuration, a peptide on the end of one of the core β-sheets is first translated and rotated so that it makes the backbone hydrogen bond contacts appropriate for that configuration. Then, restraints are added to backbone atoms of the core peptides and only the backbone atoms on the incoming peptide that are involved in its first contact to form with the core. Finally, the system is heated to 1000K; the restraints retain the existing core structure, however, the unrestrained portion of the end peptides not involved in the first contact undergo rapid fluctuation and become disordered. The resulting structure represents the case where it appears two disordered peptides in solution have just made their first contact with the amyloid core in the desired backbone hydrogen bond registry. This procedure is repeated to generate multiple initial structures of all the configurations to be simulated. See Appendix A for representative structures for each configuration. At this point, these bilayer structures are simulated for a period of time suitable to observe many transitions that occur between the incoming peptides and the core.

2.2.3 Transition Data Collection

After the simulations have completed, the trajectories are analyzed to detect and collect all conformational transitions sampled by the end peptides. A transition is denoted by any formation or breakage of a pair of backbone hydrogen bonds between residues of the end and core peptides. The pair of backbone hydrogen bonds of an amino acid is monitored for transitions as opposed to a single hydrogen bond since the formation or breakage of the second bond in the pair after formation or breakage of the first is likely to be a cooperative and rapid event. Therefore, the hydrogen bond pair is determined to have a certain overall "on" or "off" state only when both bonds of the pair are formed or broken, respectively. Once the hydrogen bond pair enters a given state, the formation or breakage of a single bond in the pair will not signal a new state; only when both bonds form or break will the corresponding new state be assumed. The distance between the acceptor and donor of the hydrogen bond determines whether the bond is formed or broken, and the distance criterion for each is different. The "on" state occurs when the distance of both bonds in the pair is shorter than 2.5 angstroms, while the "off" state occurs when the distance of both bonds is greater than 3.5 angstroms. The purpose of using different bond distance criteria for each state is to remove the possibility of spurious bond formation or breakage events that would occur if bonds distances were hovering around a single distance cutoff value for both states.

Whenever a transition occurs, it is recorded in terms of the initial and final free chain length (FCL), the residues that are interacting between the end and core peptides at the position of transitional hydrogen bond pair, and the time that elapsed since the last transition. The FCL is defined as the number of dissociated, disordered amino acids of the end peptide starting from last or first residue and counting inwards on the peptide until the first residue bound to the amyloid core with backbone hydrogen bonds is encountered. FCL was chosen as a parameter in recording transitions because it was reasoned that the viscous drag of the disordered chain is a key factor that modulates the hydrogen bond kinetics besides the identities of contacting residues.

The position of residues whose identity is recorded in a transition depends on whether the transition is due to the forming or breaking of backbone hydrogen bonds. For transitions in which bonds break, the interacting residues of the end and core peptides are recorded at the position of the existing backbone hydrogen bonds that undergo breaking. For transitions of bond formation, the residues are recorded at the position where the backbone hydrogen bonds form. In

each case, the nature of the residues are noted at positions where the backbone hydrogen bonds undergo change since it is reasoned those residues will be the most important influence on bond formation or breakage in terms of sidechain interaction.

2.2.4 Markov State Model

Next, a Markov state model (MSM) is employed to simulate the long timescale growth of amyloid fibrils. A Markov model simulates a stochastic process in which the fate and dynamics of the current state do not depend on its past. The states of MSM include all the possible combinations of FCL and residue pairs found in the transition rate matrix. After the transitions are recorded from all trajectories, the average time for each type of transition is determined and used to calculate the rate at which a particular state occurs. The Gillespie algorithm creates a trajectory of the stochastic system as follows.[61] Two random numbers, R_1 and R_2, are generated that determine what state will occur and the amount of time that elapses to reach that state. Given the rates $k_1, k_2, \ldots, k_n$ for the possible transitions from the current state and the sum of these rates, k_{tot}, state $i+1$ is the next state to occur in the trajectory such that

$$\frac{k_i}{k_{tot}} < R_1 < \frac{k_i + k_{i+1}}{k_{tot}} \qquad (2)$$

is satisfied. The second random number is set equal to the cumulative distribution function

$$R_2 = \int_0^t P(t')dt' \qquad (3)$$

to determine the time that elapses before the state $i+1$ occurs. Given that the probability function follows a single exponential distribution, the resulting explicit formula for this time is

$$t = -\frac{1}{k_{tot}} ln(1-R_2) \qquad (4)$$

The chosen state and calculated time are appended to the trajectory, at which point the new set of accessible states are determined, two new random numbers are generated, and the algorithm repeats itself. The model can be initiated from any state and terminated in response to arriving at another given state to facilitate the study of any fibril growth process of interest.

Chapter 3 - Validating Model Against Experimental Data

3.1 Amyloid-β Fragment Test System

3.1.1 Introduction

The test system used to investigate the multi-scale computational strategy is a fragment of the Aβ protein ($A\beta_{16-22}$) for which kinetic data of its self-assembly is available from the experimental work performed by the Nilsson group.[60] At neutral pH, $A\beta_{16-22}$ is capable of forming amyloid fibrils with anti-parallel β-sheet structure.[54-57] The fragment contains hydrophobic core residues 16-22 and has the sequence Ac-KLVFFAE-NH_2. This segment of the protein has a Phe-Phe motif that was the subject of investigation by the experimental study due to possibility of its aromatic interactions contributing the chief stabilizing effect on the self-assembling structures. The authors created mutants with pentafluorophenylalanine (F_5-Phe), nonaromatic cyclohexane (Cha), Tyr, or Ala incorporated at one or both of the Phe positions, 19 and 20, in order to examine the importance of the aromaticity of the core, hydrophobic residues. The F_5-Phe, Cha, and Tyr amino acids offer differing levels hydrophobicity and aromatic character than Phe but still maintain the carbon skeleton of the side chain. The F_5-Phe and Cha residues are more hydrophobic than Phe, while Tyr and Ala are less hydrophobic. The Cha and Ala residues are nonaromatic, whereas F_5-Phe and Tyr have similar aromatic interactions to Phe. The non-natural amino acids Cha and F_5-Phe are useful in studying aromatic and hydrophobic properties since their carbon skeleton is not changed from the aromatic, natural amino acids, Phe and Tyr.

3.1.2 Growth Rates Differences Among Amyloid-β Variants

Solutions of disaggregated, wildtype $A\beta_{16-22}$ and various mutants were monitored to characterize their amyloid fibril growth rates. The initial concentration of each solution was 55 μM. Concentrations of the monomeric species in solution for each peptide were measured at intervals to provide an approximate evaluation of the relative self-assembly kinetics. The monomeric concentrations were observed to decrease over time, indicating that aggregates were forming (Figure 3). Eventually, a dynamic equilibrium was reached when the monomer levels stabilized at a final critical concentration (C_r). The association constant for the binding of a single monomer to the aggregate is then computed as

$$K_a = \frac{1}{C_r} \qquad (5)$$

The wildtype sequence exhibited a lag phase of 12 hours before self-assembly began, which then continued for about 14 days before a state of dynamic equilibrium was reached. The sequences with less hydrophobic residues Ala and Tyr substituted at position 19 did not display any growth given the initial concentration of 55 μM. Another trial of the less hydrophobic sequences with a higher initial concentration of 100 μM still did not experience any self-assembly of peptides. The sequences with Cha or F_5-Phe substituted at position 19, position 20, or positions 19 and 20 all experienced a significant increase in the rate of

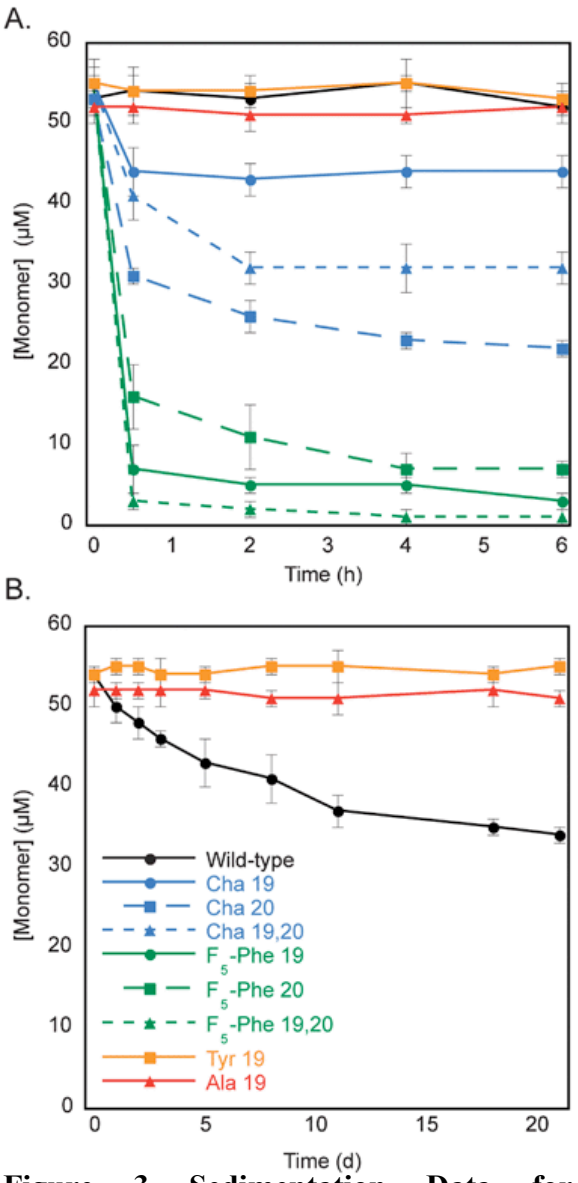


Figure 3 Sedimentation Data for Fibrillization of Wildtype and Mutant $A\beta_{16-22}$ Peptides. The sedimentation data reveals that monomer peptide concentration for Aβ16-22 variants decreases over time, indicating fibrillization. (A) Short-term, 6 hour time range for all variants. (B) Long-term, 21 day time range for wildtype, Phe19 → Ala, and Phe19 → Tyr variants. Figure adapted from the Nilsson group experimental paper.[54]

fibril growth, with dynamic equilibrium being reached within 30 minutes for each case. In addition, the variant critical concentrations differed from that of the wildtype sequence. The experimental sequences and corresponding critical concentrations are listed in Table 1. This data provides an opportunity to substantiate the theoretical model if the model is capable of capturing the differences in self-assembly rates and critical concentrations between the wildtype and variant sequences. The kinetic data of the wildtype peptide and the cyclohexane mutant peptides Cha 19, Cha 20, and Cha 19,20 will be the primary benchmark since MD force field parameters are available for these sequences.

Table 1 $A\beta_{16-22}$ Variants and Critical Concentrations for Self-assembly

Peptide	Sequence	Variant	$C_R/\mu M$
1	Ac-KLVFFAE-NH_2	Native (Phe)	33 ± 3
2	Ac-KLVAFAE-NH_2	Phe19 → Ala	> 100
3	Ac-KLVYFAE-NH_2	Phe19 → Tyr	> 100
4	Ac-KLV(F_5-Phe)FAE-NH_2	Phe19 → F_5-Phe	4 ± 1
5	Ac-KLV(Cha)FAE-NH_2	Phe19 → Cha	44 ± 3
6	Ac-KLVF(F_5-Phe)AE-NH_2	Phe20 → F_5-Phe	7 ± 1
7	Ac-KLVF(Cha)AE-NH_2	Phe20 → Cha	22 ± 1
8	Ac-KLV(F_5-Phe)(F_5-Phe)AE-NH_2	Phe19,20 → F_5-Phe	1 ± 1
9	Ac-KLV(Cha)(Cha)AE-NH_2	Phe19,20 → Cha	32 ± 2

Data from the Nilsson group experimental paper.[60]

3.2 MD Simulation Details

Solvent accessible surface area (SASA) implicit solvent was used to greatly increase the achievable simulation times of the amyloid test system and thus the number of conformational transitions to be observed compared to an explicit solvent system. The total Hamiltonian of an explicit solvent system can be summed into solute-solute, solute-solvent, and solvent-solvent terms:

$$H = H_{solute-solute} + H_{solute-solvent} + H_{solvent-solvent} \qquad (6)$$

To convert to implicit solvent from explicit solvent, the Hamiltonian is integrated over the solvent coordinates to obtain the potential of mean force or effective energy, $W(\mathbf{r})$,

$$W(r) = V_{solute}(r) + V_{solvation}(r) \qquad (7)$$

for a solute having M atoms with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1,\ldots,\mathbf{r}_M)$. The effective energy is composed of a solute-solute term and a mean solvation term. The implicit force field model used to simulate the bilayer structures assumes that the mean solvation term is proportional to the SASA of the solute:

$$V_{solvation}(r) = \sum_{i=1}^{M} \sigma_i A_i(r) \qquad (8)$$

where σ_i and $A_i(\mathbf{r})$ are the atomic solvation parameter and the SASA of atom i, respectively. Integrating over the solvent coordinates is not a rigorous calculation and hence the atomic solvation parameters are required to fine-tune the mean solvation term. The atomic solvation parameters for different atom types were optimized in a trial-and-error approach by the Caflisch group to work in conjunction with the CHARMM force field.[62, 63]

3.3 Simulation of Aβ$_{16-22}$ Oligomers and Fibril Structures

In the initial stages of choosing the amyloid core to represent a portion of a fibril onto which an incoming strand incorporates, it was noticed that unrestrained simulations of β-sheets developed a slight, natural twist along the long axis of the β-sheet within a short amount of time, consistent with experimental observations of twisted fibril structures. To confirm that the simulation was indeed capturing the twisting nature of the peptide and not an incidental occurrence, large fibril-like structures were constructed and simulated to measure the amount and magnitude of any developing twist. The fibril structure consisted of four 100-mer bilayers grouped closely together. After a short 100 picosecond simulation to relax the structure, a twofold fibril developed with four β-sheets twisting together in each fibril (Figure 4). This confirmed that the simulation was in fact capturing the twisting effect of the Aβ$_{16-22}$ fragment sequence.

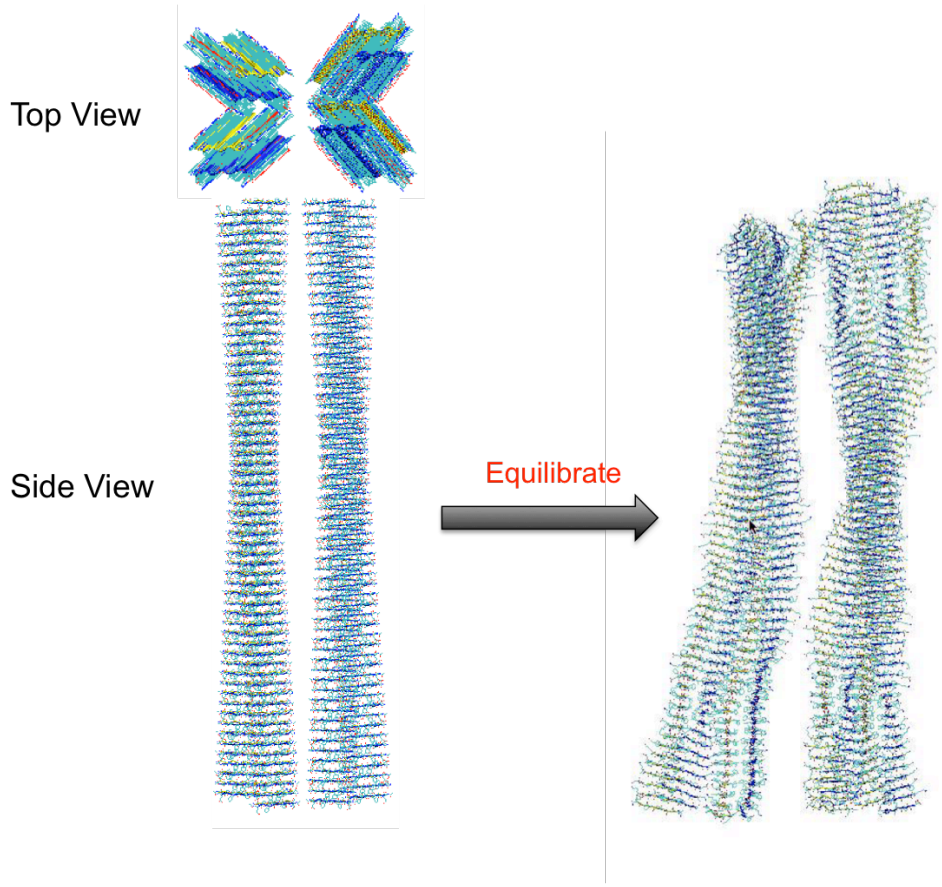


Figure 4 $A\beta_{16-22}$ Fibril Construction and Equilibration. The top and side views show the construction of the fibril. After a short equilibration, the fibril adopts a twisted conformation.

It was desired to retain the twisted nature of the Aβ peptide in the amyloid core to be simulated in order to create a model system that reflects the actual conformational features of the fibril. Additionally, the choice of a reasonably small system was preferred to minimize computation and increase possible sampling. Single layer and bilayer β-sheets fit these requirements, and variously sized sheets were examined for their stability and conformational differences. In test simulations, a single layer exhibited a very sharp twist along the long axis of the β-sheet, whereas a bilayer displayed a more gradual twist. The bilayer was ultimately chosen as the amyloid core as its gradual twist is likely a more representative feature to be found in the interior of an amyloid fibril. In addition, the presence of a neighboring sheet is believed to play a role in determining the templating kinetics.[64]

To construct the bilayer structure to be used in the molecular dynamic simulations, a single strand is first generated with either the wildtype or one of the cyclohexane mutant

sequences using the molecular simulation program CHARMM. The backbone phi and psi angles are set to -140° and 135°, respectively, which correspond to those of a peptide in a representative anti-parallel β-strand conformation. The strand is then duplicated and translated to make two opposing sheets in a bilayer structure. Strands were separated by 5.6 angstroms on the long axis of the bilayer and every other strand was offset by 1.2 angstroms on the perpendicular axis so that the backbone hydrogen bonds would be aligned. The inter-sheet distance was set to 10 angstroms. The geometry was based off of an NMR anti-parallel Aβ fibril structure from the RSCB Protein Data Bank. Unrestrained simulations of several differently sized bilayers indicated that a core of six strands in each sheet was stable over a period of one nanosecond (Figure 5). In order to minimize deviations from a representative amyloid core, therefore, a bilayer with 10-mer sheets and restrained backbone hydrogen bonds was equilibrated long enough for the natural twist of the fibril to develop, and then the outer two peptides on each end of both sheets were deleted to leave a core with few perturbations.

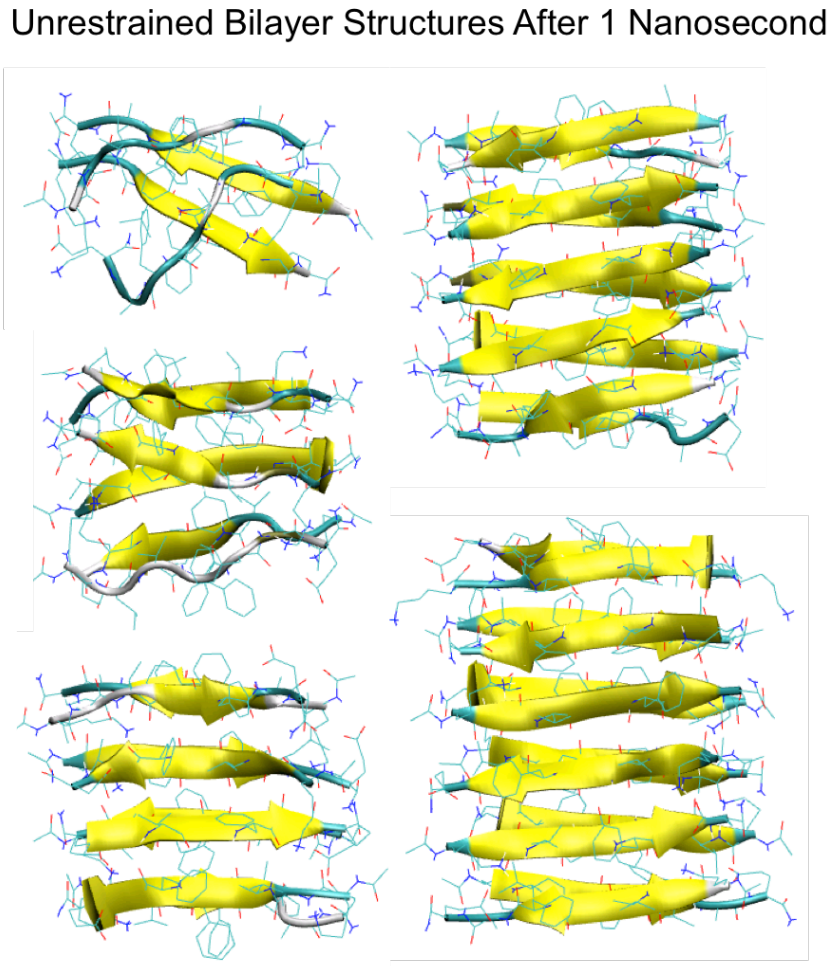


Figure 5 Stability of Differently Sized Bilayers. Final conformations of variously sized bilayers after a 1 nanosecond MD simulation exhibit a varying degree of stability. A bilayer with six peptides in each β-sheet was determined to be the smallest, stable structure to be used as the amyloid core.

Restraints were applied to the backbone hydrogen bonds of the bilayer structure that took effect when the distance between the heavy atoms in the hydrogen bond were farther than 3.5 angstroms apart. The system was then equilibrated for 100 picoseconds to relax the structure and allow the natural fibril twist to develop (Figure 6). This relaxed bilayer structure was the template from which all in-register and mis-registered configurations were derived. To create a mis-registered configuration, the strands on each end of the fibril were first duplicated. The duplicate strands were then translated and rotated so that the backbone hydrogen bond acceptors and donors of the residue imitating first contact with the core superimposed those of the residue

on the original, in-register strand whose place it was taking. The end in-register strands were then deleted, leaving only the mis-registered strand.

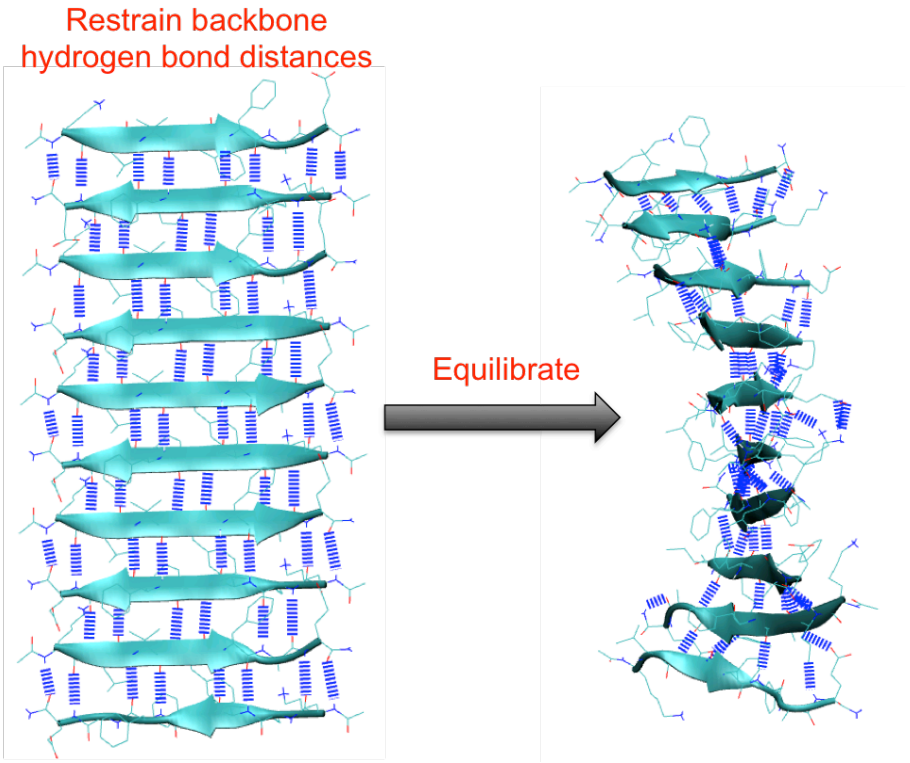


Figure 6 Twisting Effect of β-sheet After Equilibration. Restraints were applied to the backbone hydrogen bonds of the bilayer. After a short equilibration, the structure develops a dramatic twist. The test system on which this was performed is a bilayer, however, a single β-sheet is shown here for clarity in visualizing the twist.

The final step in creating each configuration was modifying the fully bound end strand to appear as though it has made only a single contact with the core. To create this effect, heavy atom restraints were applied to the amyloid core and to the residues on the end strands that were imitating the first contact. The system was then simulated at 1000 Kelvin for 100 picoseconds, resulting in the unrestrained portions of the end strands to dissociate from the core and assume a disordered conformation. For each configuration, fifty such preparative conformations were generated. Heavy atom restraints were again applied to the amyloid core as in the heating procedure to prevent disruptive interior events in detracting from the observation of transition dynamics of the end strands. Two sets of 50 nanosecond simulations were carried out at 300 Kelvin using this setup, one with an additional heavy atom restraint on the residues of the end strands imitating the first contact with the core and one with no additional restraints added. Thus,

the restrained set enforces the end strands to remain in their assigned configuration, which facilitates focused sampling of hydrogen bonding dynamics of a given registry. The unrestrained end strands in the second set of simulations were able to explore other configurations or dissociate from the amyloid core completely. Overall, each configuration was simulated for a total of 500 nanoseconds, equivalent to1 μs of sampling with two independent peptides (Figure 7). An illustrative initial conformation for each configuration and an explanation of the notation used to identify each can be found in Appendix A.

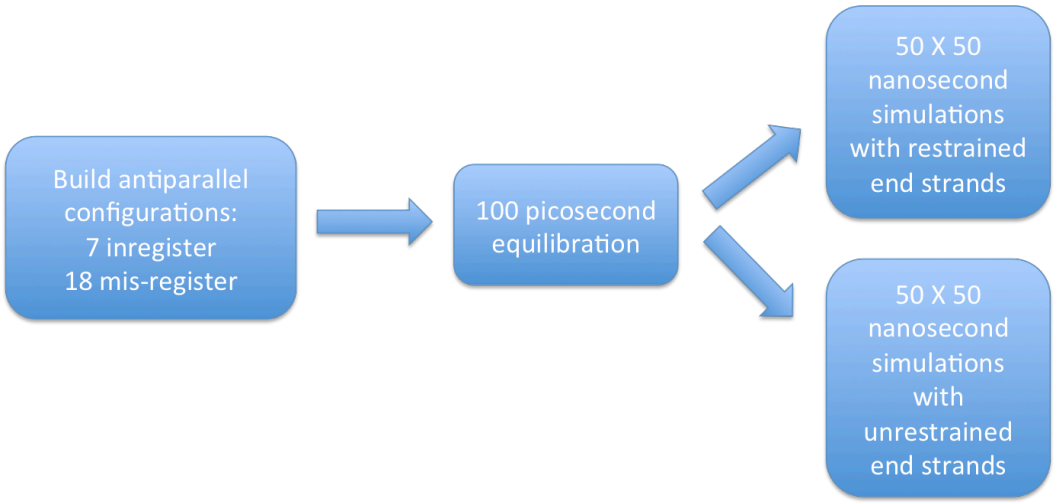


Figure 7 Overview of Simulation Procedure

3.4 Results

3.4.1 Overview of Transition Data Matrix

The complete 50 nanosecond trajectories for each configuration were analyzed for any conformational transitions as described in Chapter 2. The times that elapsed for each type of transition to occur follows a single exponential distribution as shown in Figure 8. This indicates that most transitions are indeed dominated by single barriers and are suitable for MSM description and the Gillespie algorithm.

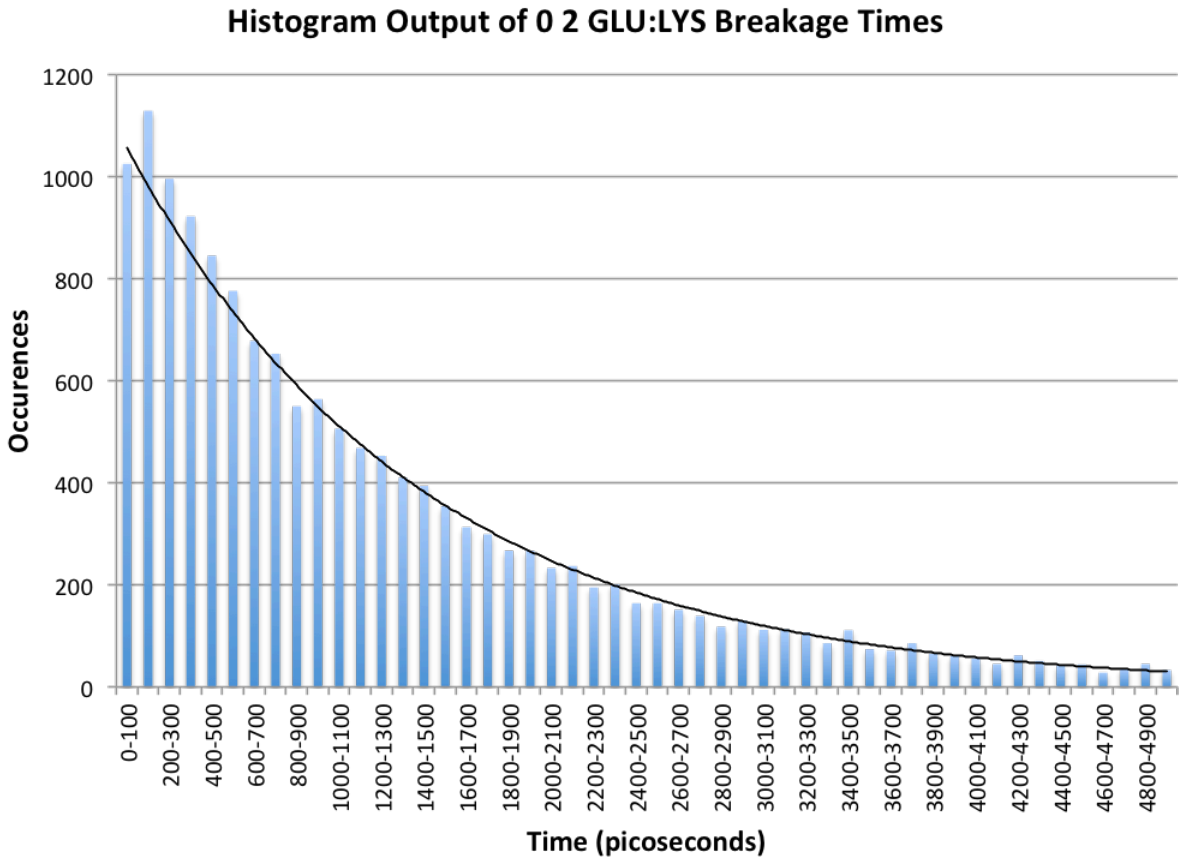


Figure 8 Histogram Output of a Sample Conformational Transition. The horizontal axis is a range of times that elapsed before the transition occurred. The vertical axis is the number of times that transition was observed to occur. The sample transition shown here is the breaking of hydrogen bonds that results in the passage from a free chain length of 0 to 2 and the residues involved were a glutamate on the end strand and a lysine on the amyloid core. The transition times follow a single exponential distribution.

The times for each type of transition were averaged and are listed in Appendix B. Overall, the mutant sequences experienced significantly slower hydrogen bond breakage transitions than those of the wildtype sequence (Table 2). There is not as strong a trend for bond formation transitions, as there is a slight increase in rates for transitions involved in in-register configurations and a mixed difference in rates for those in mis-registered configurations. In general, there does not seem to be an apparent dependence of the increase or decrease of transition times on the residue types involved. Some transitions that do not involve the mutant cyclohexane residue may exhibit a similar rate increase or decrease for all mutant sequences while other transitions that also involve a mutant cyclohexane residue may not have any

discernable pattern. The observed increase of experimental fibril growth rates for the mutant structures is therefore not likely explained solely by the specific set of interactions of paired side-chains for a given sequence, but the collection of small effects propagated to all interactions by the introduction of a point mutation.

Bond Breakage Transitions							
	Avg Time (ps)	Avg Time (ps)	% Diff from WT	Avg Time (ps)	% Diff from WT	Avg Time (ps)	% Diff from WT
	WT	CHA19		CHA1920		CHA20	
In-register	1532	1656	28%	1814	55%	1842	48%
Mis-register	1705	1668	9%	2038	33%	1861	31%

Bond Formation Transitions							
	Avg Time (ps)	Avg Time (ps)	% Diff from WT	Avg Time (ps)	% Diff from WT	Avg Time (ps)	% Diff from WT
	WT	CHA19		CHA1920		CHA20	
In-register	1046	1307	13%	1279	25%	1297	16%
Mis-register	984	912	-6%	797	-6%	1059	6%

Table 2 Average Times of Backbone Hydrogen Bond Formation and Breakage.
Gradient fill colors are used to show the magnitude of the percent difference among mutant sequence transition rates in comparison to that of the wildtype sequence. Solid red refers to ≥ +100% difference, while solid green refers to ≤ -100% difference.

Although there is not a discernable trend of the transition times in relation to the residue types involved, there is however a clear dependence on the free chain length. Hydrogen bond formation transition times increase as the free chain length increases, which is consistent with the idea that a longer disordered chain will be less dynamic due to viscous drag and slower to form hydrogen bonds on the amyloid core. It is not immediately clear if the breaking of hydrogen bonds is dependent on FCL.

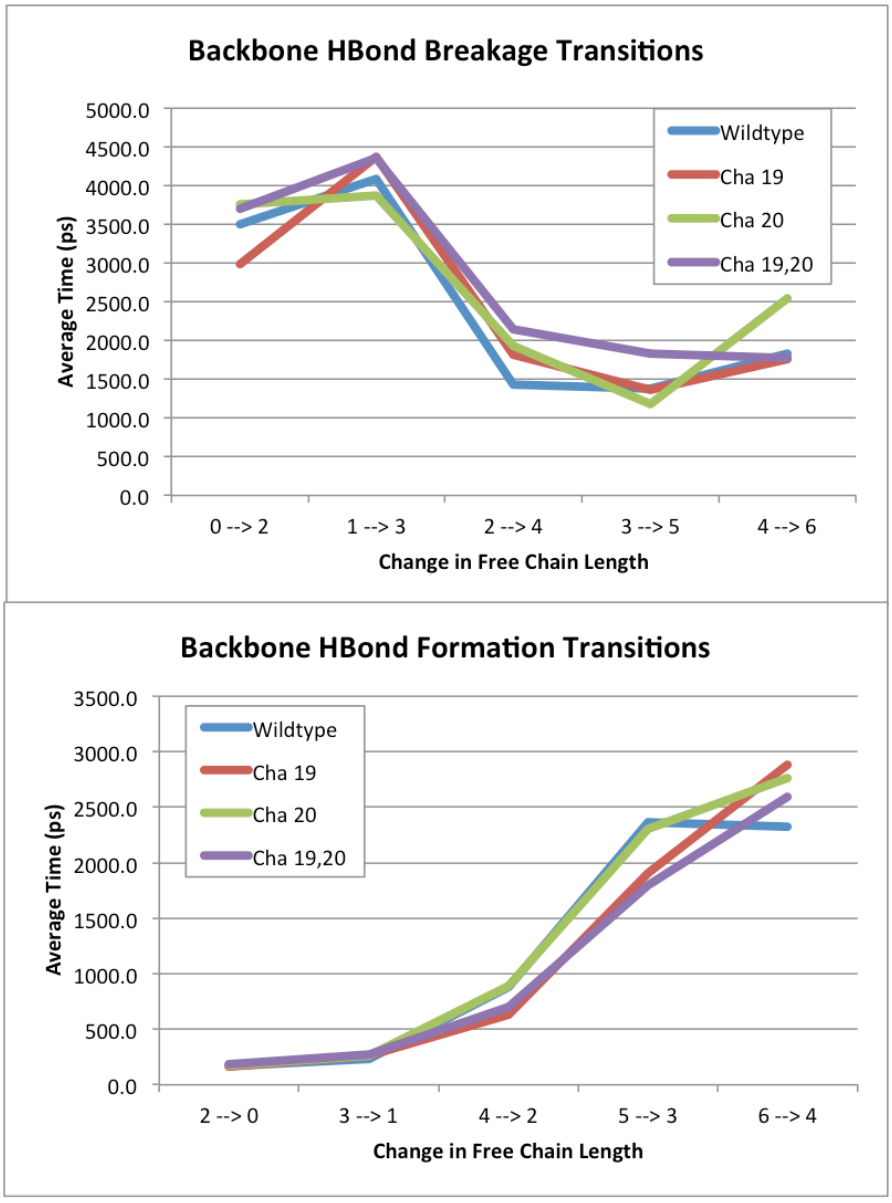


Figure 9 Average Times of Backbone Hydrogen Bond Formation and Breakage Transitions as a Function of Free Chain Length. The average time for a transition to occur is plotted here in terms of the initial and final FCL. For bond breakage transitions, the final FCL is greater than the initial FCL, while the opposite is true for bond formation transitions. The average time to form hydrogen bonds increases with increasing FCL, whereas the average time to break hydrogen bonds decreases with increasing FCL.

3.4.2 Folding Probability

The folding probability p_{fold} of a protein structure recorded in an MD trajectory is the probability for the protein to fold before unfolding, starting from an intermediate conformation

between the native and unfolded state. This order parameter represents the kinetic distance of a structure from its native, folded state. A p_{fold} value of 1 indicates all structures folded over the course of the trajectory, while a p_{fold} value of 0 indicates all structures became unfolded. To calculate the p_{fold} probability, folding and unfolding criteria must first be determined. To achieve this, a histogram plot was created of the probability of the two different amyloid core structures to have a number of formed backbone hydrogen bonds (Figure 10). The core structures are termed *even* and *odd* to refer to the number of hydrogen bond pairs available on the core peptide.

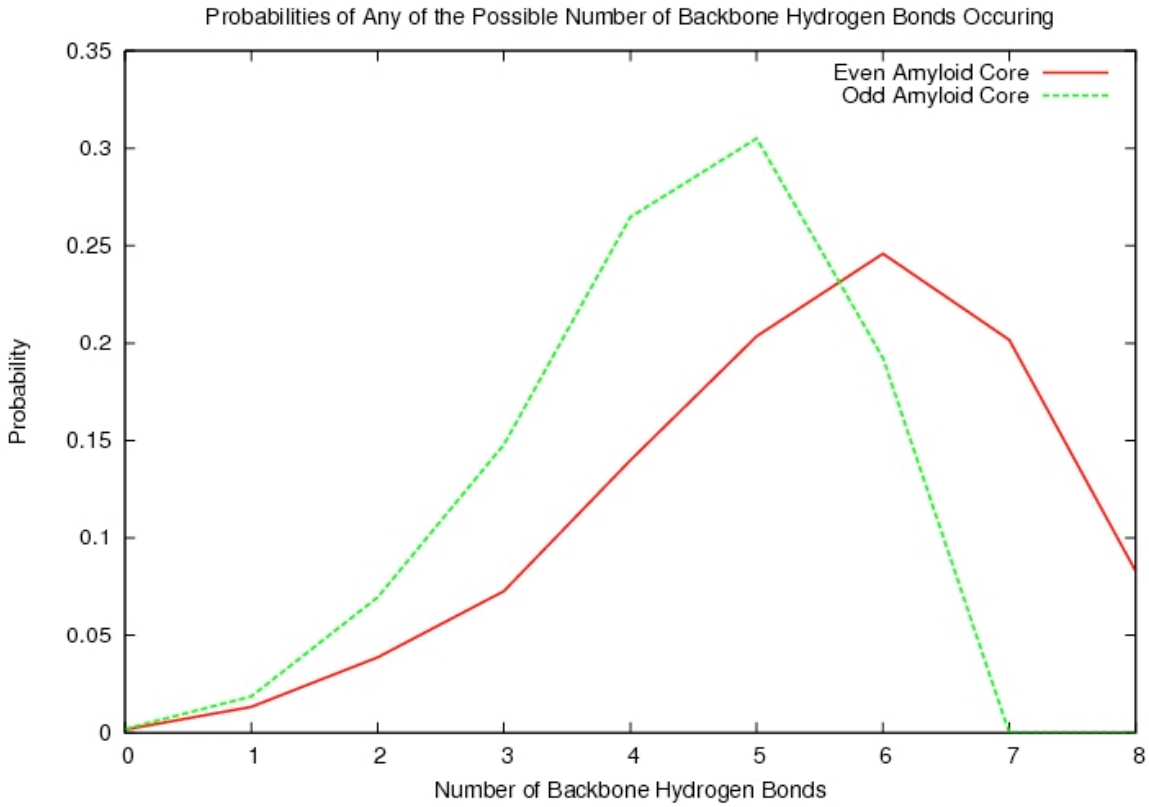


Figure 10 Probability of Even and Odd Amyloid Cores Forming a Number of Backbone Hydrogen Bonds with the End Peptide. All trajectories were scanned for the number of hydrogen bonds formed between the end strand and the amyloid core. The occurrences for each possible number of hydrogen bonds for even and odd cores were recorded and divided by the total time. The resulting number is the probability of that number of hydrogen bonds to have occurred. The probabilities are plotted here for each number of hydrogen bonds.

The data point slightly to the left of the peak bin will designate the most appropriate choice for the folding criteria. From the histogram data, a folding criterion of 4 and 5 backbone

hydrogen bonds for the odd and even core configurations, respectively, will likely describe the folded state of the peptide for the majority of conformations, and as such these parameters were used to define trajectories in which the end peptide became folded for the p_{fold} analysis. End peptides were considered to become unfolded if their number of backbone hydrogen bonds dropped below one. To minimize noise in the data, the running average of the backbone hydrogen bonds for a particular configuration was computed over a time frame of 25 picoseconds. In addition, any trajectories in which the end peptide already existed in the folded or unfolded state at the first frame were not included in the analysis, as p_{fold} is the probability of the end peptide initially in an intermediately folded state moving to the folded conformation. The p_{fold} values for in-register even and odd core configurations are plotted in Figure 11 and listed in Table 3. The notation on the horizontal axis used to define configurations is explained in Appendix A.

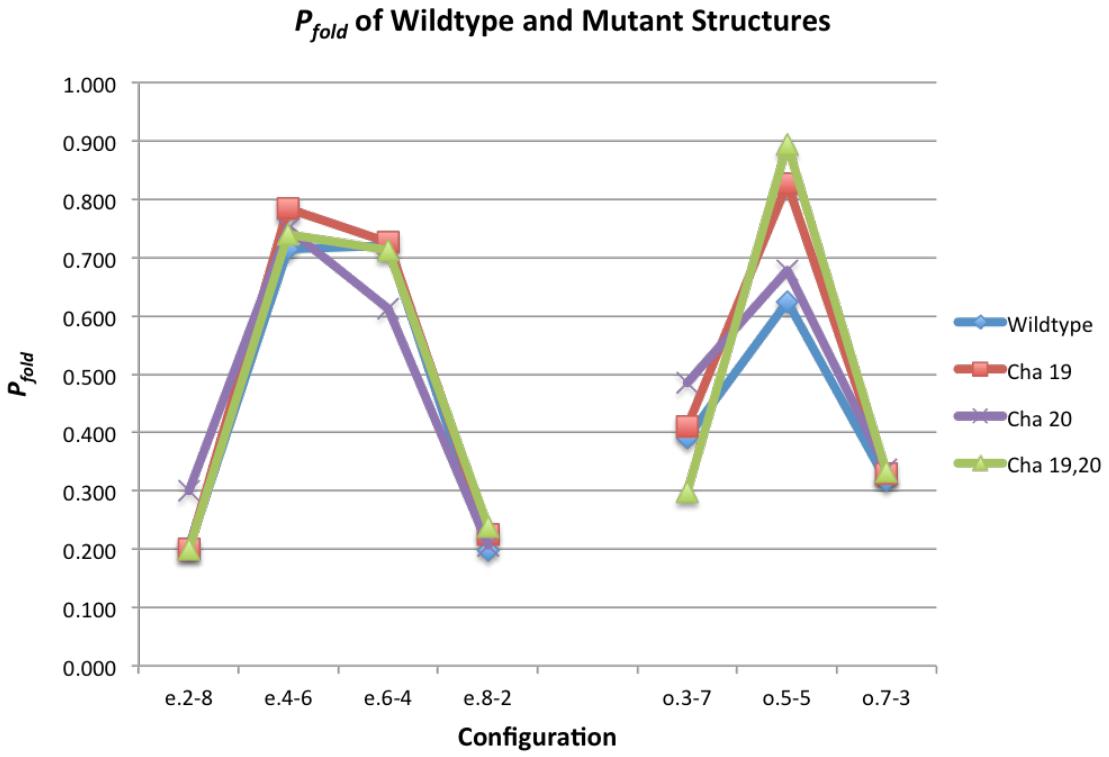


Figure 11 P_{fold} of Wildtype and Mutant Structures After 50 Nanoseconds. The plotted p_{fold} values correspond to in-register even and odd core configurations for the wildtype and mutant peptides. Lines between data points are plotted solely as a visual guide and do not signify any data evolution.

Table 3 Outcome and P_{fold} of End Peptides After 50 Nanoseconds

		Configuration						
	Sequence	e.2-8	e.4-6	e.6-4	e.8-2	o.3-7	o.5-5	o.7-3
# Peptides Do Not Bind	WT	70	22	21	69	45	18	56
	CHA19	68	14	20	69	43	9	53
	CHA1920	72	21	21	64	54	6	54
	CHA20	58	18	29	69	38	18	53
# Peptides Bind	WT	18	55	54	17	29	30	26
	CHA19	17	51	53	20	30	43	26
	CHA1920	18	60	52	20	23	51	27
	CHA20	25	54	46	18	36	38	27
# Already Unfolded	WT	10	0	5	11	3	5	4
	CHA19	13	4	2	10	9	7	8
	CHA1920	9	1	5	15	15	5	8
	CHA20	8	3	4	10	6	3	3
# Already Bound	WT	2	23	20	3	23	47	14
	CHA19	2	31	25	1	18	41	13
	CHA1920	7	25	21	3	18	41	15
	CHA20	4	34	22	3	21	35	18
P_{fold}	WT	0.205	0.714	0.720	0.198	0.392	0.625	0.317
	CHA19	0.200	0.785	0.726	0.225	0.411	0.827	0.329
	CHA1920	0.200	0.741	0.712	0.238	0.299	0.895	0.333
	CHA20	0.301	0.750	0.613	0.207	0.486	0.679	0.338

The kinetic theory has predicted that the transition state of templating occurs early, around when the peptide forms its first hydrogen bond pair of a native registry. The MD simulations here are consistent with this prediction, showing that all initial configurations with one registered hydrogen bond pair have a p_{fold} value in the region of 0.5, ranging from 0.2 to 0.9. Considering that the p_{fold} values are exponentially related to the location of the structure on the energy landscape, the observed values exhibiting the same magnitude as 0.5 suggest an initial structure close to both binding or unbinding, which is consistent with the predicted outcome of a peptide making a single contact with the amyloid core. The simulation also captures interesting heterogeneity in the p_{fold} values of different configurations. The in-register peptides whose initial contact is more lateral with respect to the amyloid core exhibit a lower probability of folding since they have a longer initial free chain length, which causes the end peptide to be more dynamic and more likely to unfold. If the initial contact is more central between the end and core strands, the end strand is more stable and the p_{fold} value is logically larger.

3.4.3 Estimating Fibril Growth Through the Markov State Model

3.4.3.1 Algorithm Details

The averaged transition data was used to provide the kinetic rates between states in a MSM. The Gillespie algorithm was then used to generate a trajectory of a peptide interacting with an amyloid core given that at least a single contact is already formed. The same states as specified in the MD simulations are involved in the MSM given by free chain length and the interacting residues between the end and core peptides at the site adjacent to the free chain, which will be termed the residue pair. At each iteration of the algorithm, the current state of the peptide is calculated and then the possible states to which the peptide can transition are determined. Next, two random numbers are generated which determine which state will occur and how much time elapses before the peptide assumes that state, as described in Chapter 2. The algorithm continues until a desired state is reached.

3.4.3.2 Using Experimental and MSM Data to Calculate Theoretical Rates

To calculate the fibril growth rate as depicted by the Markov state model, equation 1 is adapted to

$$k_{grow} \simeq \frac{P_+}{(2L - 1)(t_{wait} + t_{diff}(c))} = \frac{P_+}{(2L - 1)(t_{wait} + a/c)} \qquad (9)$$

The diffusion time is expected to be inversely proportional to the concentration of the monomeric peptide in solution, c. The fitting parameter a is also introduced here and represents the probability of a random collision leading to the formation of the first hydrogen bond pair. At equilibrium, the rate of fibril growth is equal to the rate of fibril dissolution. The rate of fibril dissolution is equal to the quantity $1/t_{off}$, where t_{off} is the average residence time of a peptide fully bound to the amyloid core in its optimal configuration before it dissociates. Since the equilibrium concentration of the monomeric peptide, C_R, is known from the experimental study, the experimental constant a can be solved for:

$$a = C_R\left(\frac{t_{off}P_+}{(2L - 1)} - t_{wait}\right) \qquad (10)$$

Once a is known, the fibril growth rate and diffusion time can be calculated from equation 9.

From the fully bound state in each configuration, 3000 MSM simulations were performed to calculate the average residence time, t_{off}. The times are plotted in Figure 12 and Appendix A provides an explanation of the notation used on the horizontal axis to describe the configurations with fully bound end peptides. There is a clear distinction in residence times among the different sequences, but there is not a general trend for a given sequence to have a longer or shorter residence time across all configurations.

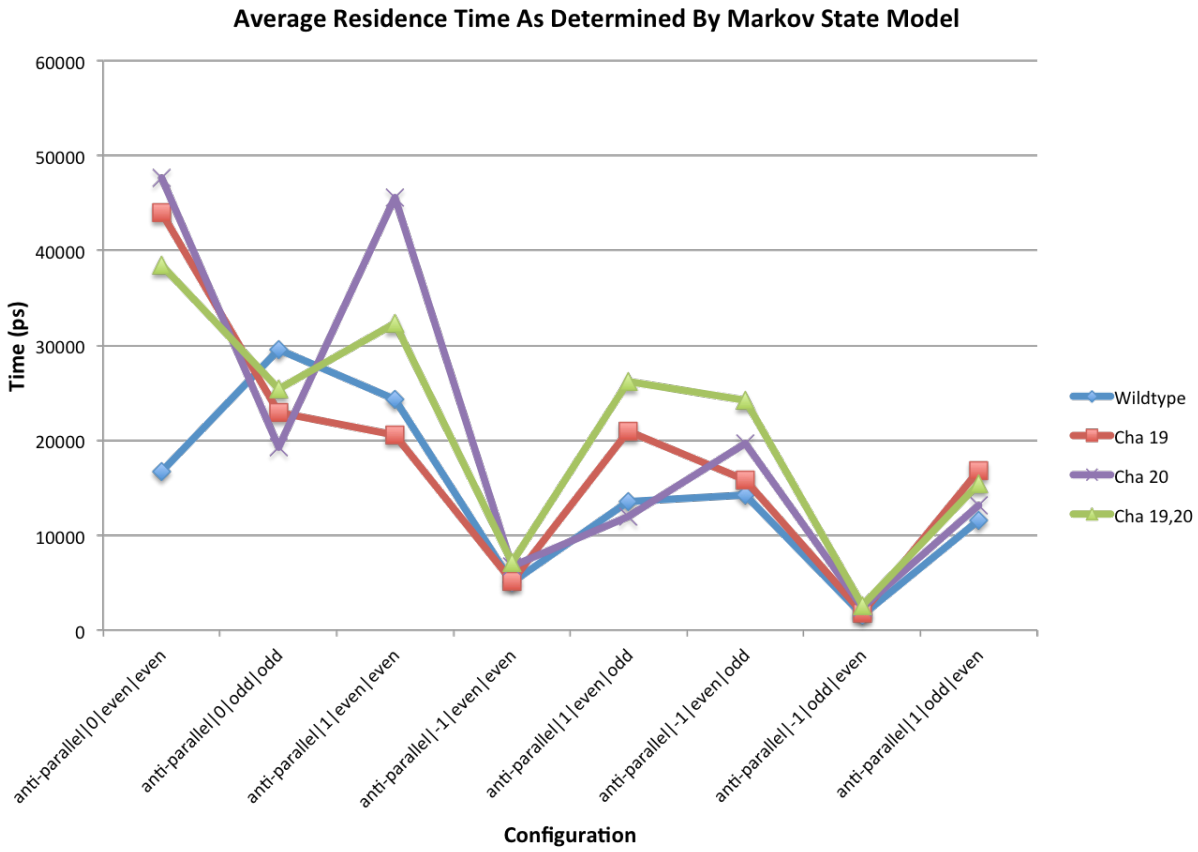


Figure 12 Average Residence Times as Determined by Markov State Model for Configurations Previously Simulated with Molecular Dynamics. The configurations that were simulated with MD were simulated with the MSM. The average residence time for the end peptide as calculated from 3000 MSM simulations are plotted here for each configuration.

In addition, 10,000 simulations were performed starting from a random configuration making a single contact with the amyloid core. The success rate of peptides becoming fully bound in the optimal, in-register configuration was recorded and this quantity is equivalent to $P_+/(2L - 1)$. The average time of mis-registered peptides to dissociate was calculated as the value

for t_{wait}. With all the unknowns of equation 10 determined, the diffusion time for each sequence was then calculated. The calculated values for each sequence are listed in Table 4.

Table 4 Calculated Values for Determining the Diffusion Time at Fibrillization Equilibrium

Sequence	$P_+/(2L-1)$	t_{off} (ps)	t_{wait} (ps)	t_{diff} (ps)
Wildtype	0.0181	22213	994	-592
Cha 19	0.0181	34984	1268	-633
Cha 20	0.0193	35517	1578	-894
Cha 19,20	0.0190	32857	1809	-1183

Noticeably, the calculated diffusion times are negative for each sequence, which suggests that one of the values derived from the Markov state model is incorrect. An assumption of the MSM is that once a peptide breaks its last contact with the core, it dissociates and contributes to the monomeric population in solution. However, another plausible scenario is that the peptide actually persists in a nonspecifically bound state for an amount of time comparable to the templating time scales before it completely dissociates or reforms hydrogen bonds. If true, the current MSM model will underestimate t_{off} by ignoring the critical nonspecifically bound states. To check if such events have significant occurrence, 100 nanosecond MD simulations of fully bound, in-register configurations were performed with no restraints placed on the end peptide, allowing it the ability to dissociate. The resulting trajectories were analyzed to find the register state of the unrestrained peptides.

The register state of the end peptides are defined in terms of the β-strand orientation of the peptide, either parallel or anti-parallel; the shift of the peptide, which refers to the degree of displacement in relation to the core with 0 corresponding to an in-register strand and +/- *N* corresponding to a mis-registered strand shifted +/- *N* hydrogen backbone pairs to one side; and whether there is an even or odd number of hydrogen bond pairs available for binding on the amyloid core and the end strand (see Appendix A for further explanation of register state notation).

If the strand was involved in making backbone hydrogen bond contacts that did not unanimously agree with one register state, it was said to exist in a nonspecific state. In addition,

if no backbone hydrogen bond contacts were formed, but the strand was still proximal to the bilayer core, the register state was also classified as nonspecific. Finally, if the strand was not proximal to the bilayer core, it was said to exist in a dissociated state. The register states of the end peptides were computed as a function of time, a sample plot of which can be seen in Figure 13. Table 5 lists the number of end peptides that changed their register state or remained in their original configuration at the end of a 100 nanosecond simulation. Fifty trajectories were generated for all four possible in-register configurations to offer 100 unrestrained end peptides capable of sampling different register states. All configurations exhibited some peptides that changed their register state by the end of the 100 nanosecond period. Clearly, changes from one register state to another through a transitional, nonspecifically bound state for the end peptides are a common event and will need to be accommodated in the Markov state model.

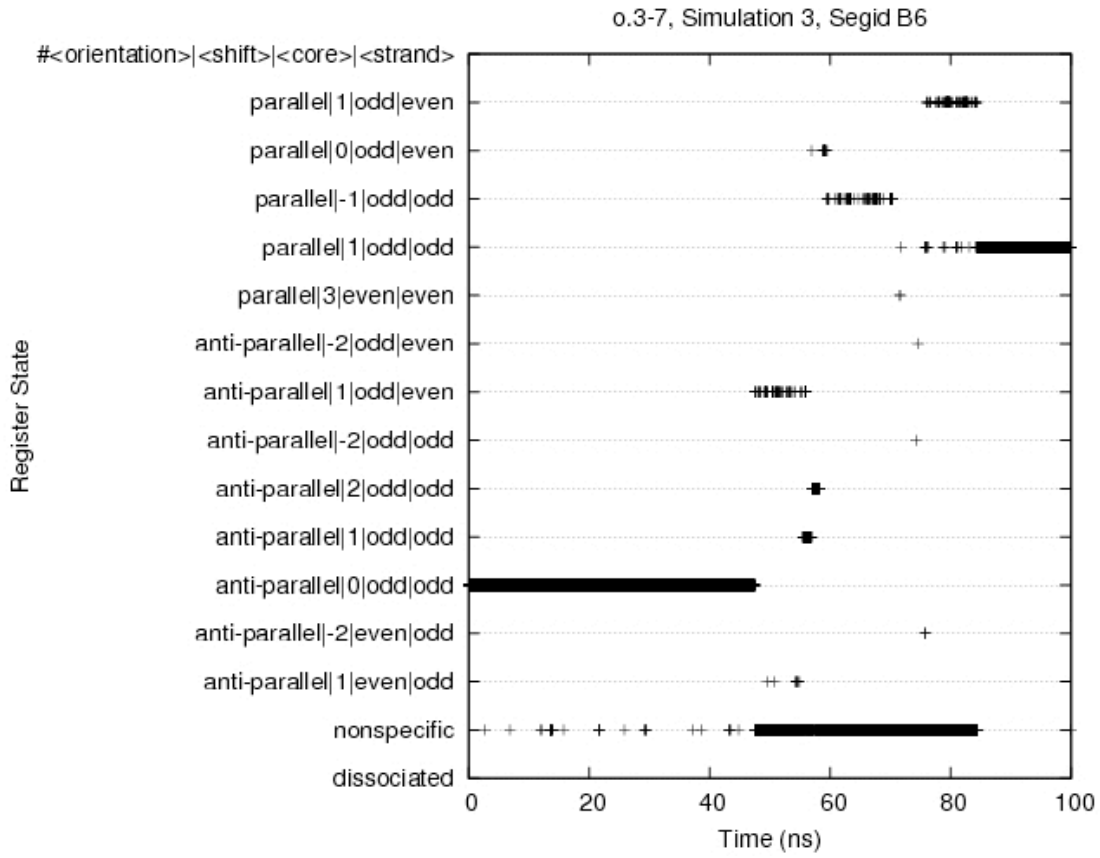


Figure 13 Sample Plot for the Register State of an Unrestrained End Peptide as a Function of Time. The register state of the end peptides were tracked along the course of the MD simulations. At each frame in the trajectory (marked on the horizontal axis), the peptide was assigned to have one of the register states listed on the vertical axis. The sample plot shown here

indicates that this particular peptide stayed in its initial anti-parallel|0|odd|odd register state for the first half of the simulation before assuming a nonspecifically bound register state for a significant amount of time and then finally settling on the parallel|1|odd|odd register state. For $A\beta_{16\text{-}22}$, there are a total of 52 possible register states, however, the vertical axis only lists the register states the peptide experienced in the sample trajectory for clarity.

Table 5 Outcome of Unrestrained End Peptides After 100 Nanosecond Simulation

	Number of End Peptides	
Configuration	Remained In Original Register State	Changed Register State
e.2-8	44	56
o.3-7	26	74
e.3-3	91	9
o.4-4	30	70

To check the accuracy of the *toff* values used in the diffusion time calculation, the Markov state model residence times for the wildtype sequence were compared with the residence times as observed in the MD simulations (Figure 14). Starting from a fully bound state, the residence time is measured as the amount of time that elapses before the end peptide breaks its last backbone hydrogen bond pair or assumes the previously defined "non-specific" register state as observed in the Markov state model or the MD simulations, respectively. The residence times for the Markov state model were averaged from 3000 unbinding events for each configuration and thus can be considered a precise number for the model. The labels above the wildtype MD simulation data points indicate the number of times the end peptide was observed to assume the non-specific register state given that it started in the configuration specified on the horizontal axis. Several configurations experienced few end peptide unbinding events, and so the corresponding residence times must be considered tentative values. The MD simulation residence times must also be taken with reservations because trajectories in which the end peptide did not change its register state were not accounted for in the residence times. Therefore, the MD residence times are likely less than the actual time that end peptides stay bound to the core, which may be a contributing factor to the negative diffusion times that were calculated. Despite the unreliability of the MD data, the plotted comparison shows that the Markov state

model is able to capture the general trend of the MD simulation residence times. This observation further supports the notion that the MSM as implemented is correct, although incomplete since the critical nonspecifically bound states have been ignored.

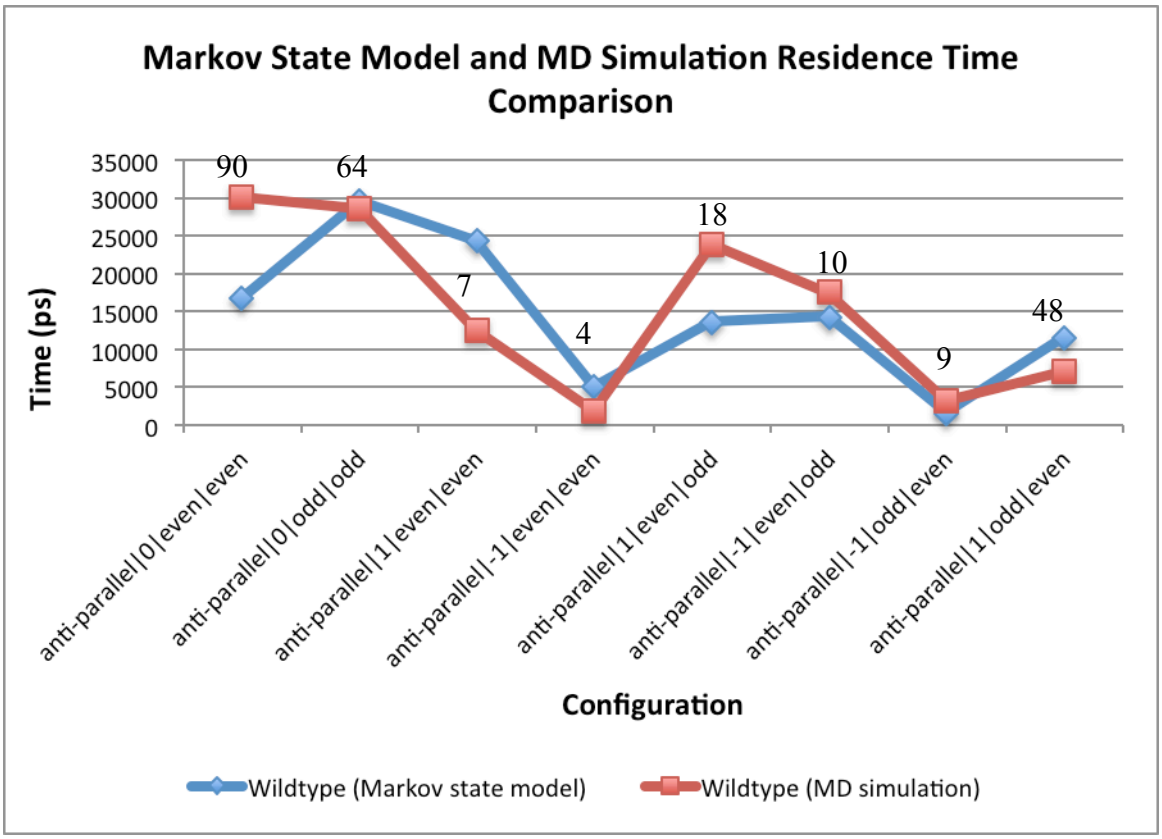


Figure 14 Average Residence Time Comparison Between the Markov State Model and MD Simulations for the Wildtype Sequence. The plot shown here compares the residence times of the wildtype end peptide as observed in the MD simulations and MSM. The labels above the MD simulation data points indicate the number of times the end peptide was observed to assume the non-specific register state given that it started in the configuration specified on the horizontal axis.

3.5 Discussion and Future Work

Conformational transition data from MD simulations of $A\beta_{16-22}$ wildtype and mutant systems in terms of free chain length and the interacting residues were not found to be identical among the different sequences. When the transition data was used in a Markov state model to simulate a fully bound peptide dissociating from an amyloid core structure, the resulting residence times vary depending on from which sequence the transition data were collected and also the residence times from the model reflect those determined from the MD simulations. This

indicates that the order parameters that define the conformational transitions and their implementation in the model are capable of capturing kinetic differences that are sequence dependent. However, when p_{fold}, t_{off}, and t_{wait} numbers were determined from the model and used in conjunction with experimental data to determine the fibril growth rate, the calculation incorrectly yields negative values for the peptide diffusion time. Investigation revealed that the model might be estimating the residence time to be too short. Also, the model assumes that once a peptide breaks its last backbone hydrogen bond pair, the peptide dissociates free of the amyloid core and becomes part of the monomeric population in solution, whereas a more probable scenario is one where the peptide becomes nonspecifically associated with the amyloid core once it breaks its last contact. These nonspecifically bound peptides then have the opportunity of dissociating or reincorporating into the amyloid core, not necessarily in its initial configuration. Future work on the model, therefore, should focus on adding the possibility of peptides assuming a nonspecific state. In combination with the experimental data, the revised kinetics of the model may then be able to provide an accurate estimate of amyloid fibril growth.

Appendix A - Configuration Figures and Notation

A.1 Notation for Configurations with a Single Initial Contact

Due to the many in-register and mis-registered conformations capable of being formed by the $A\beta_{16\text{-}22}$ fragment model system, a three character notation system was devised to identify each configuration that was simulated. The first character describes the amyloid core onto which the end peptide joins. Since the model peptide has 7 residues, the amyloid core will either have an odd or an even number of backbone hydrogen bond pairs available for bonding with the end peptide, depending on the orientation of the most exterior peptide in the core. Hence, the first character is either "e" or "o" for an even or an odd number of available backbone hydrogen bond pairs, respectively. As such, the terms *even core* or *odd core* will be used as an abbreviated phrase to describe the number of available backbone hydrogen bond pairs on the amyloid core (Figure 15).

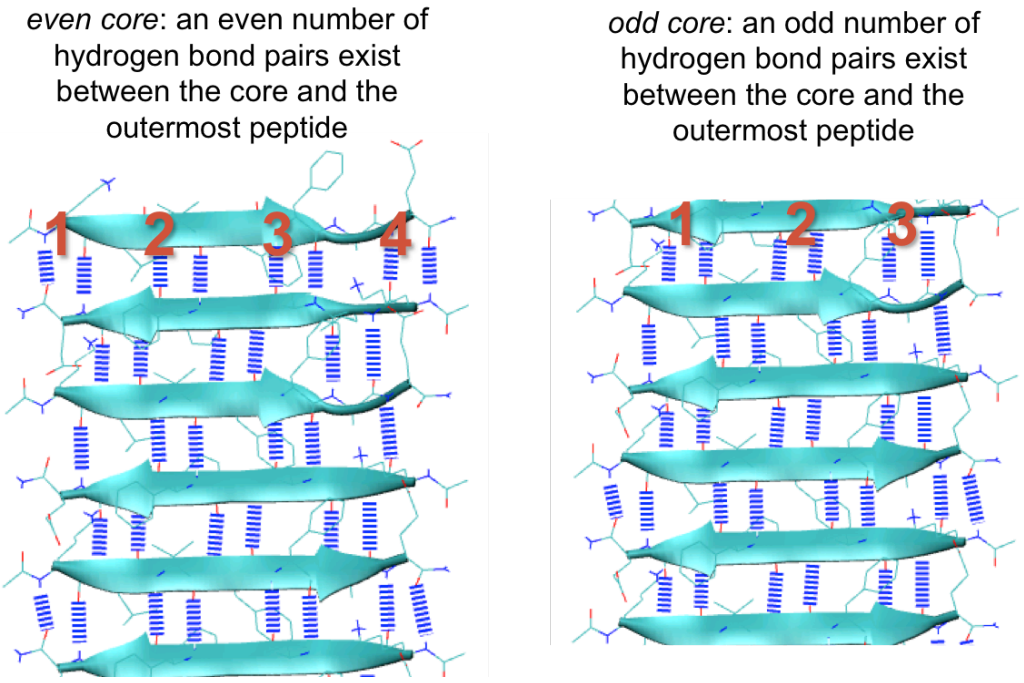


Figure 15 Bonding Differences of the Amyloid Core. The two types of amyloid cores are shown here (odd or even) and the number of backbone hydrogen bond pairs is indicated by the bold red numbers.

The next two characters identify which residue pair was restrained during the heating simulation to give the end peptide a conformation that mimics an incoming strand having just formed a single pair of backbone hydrogen bonds. The second character refers to the strand residue ID and the third character refers to the core residue ID in pdb format. The first residue ID of the model system refers to an acetyl group capping residue that imitates a peptide bond connecting the fragment to the omitted residues. Residue IDs 2-8 correspond to the fragment residues 16-22. See Figure A.2 for an example of an in-register configuration name.

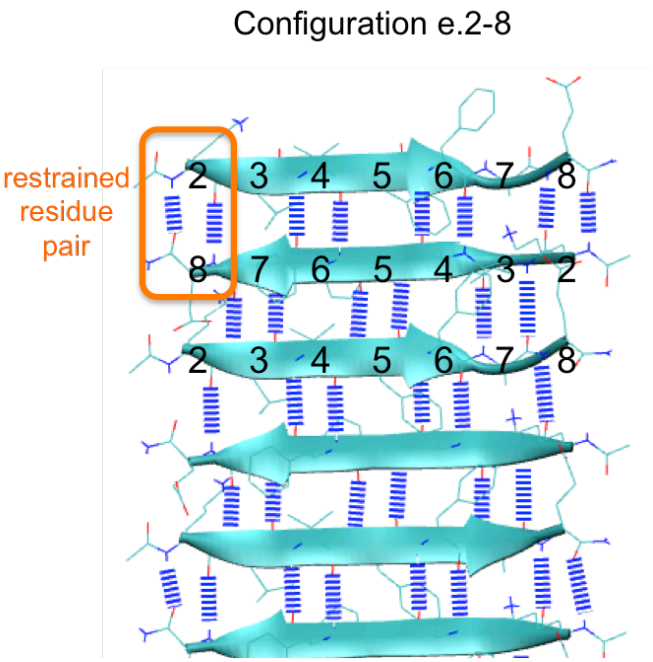


Figure 16 Schematic of Configuration e.2-8. The notation for the configuration shown here is e.2-8. The letter "e" indicates that the end strand is contacting an even core, an amyloid core with an even number of backbone hydrogen bond pairs. The numeric pair "2-8" indicates that residue number 2 on the end strand and residue number 8 on the core are restrained to maintain their hydrogen bonds.

A.2 Register State Notation

The notation used to describe the register states an incoming peptide of any sequence can assume involves four parameters. The first simply denotes the β-strand orientation: parallel or anti-parallel. The second parameter is termed the peptide's "shift" and relates to what degree the strand is mis-aligned on the amyloid core. A shift of zero corresponds to an in-register strand,

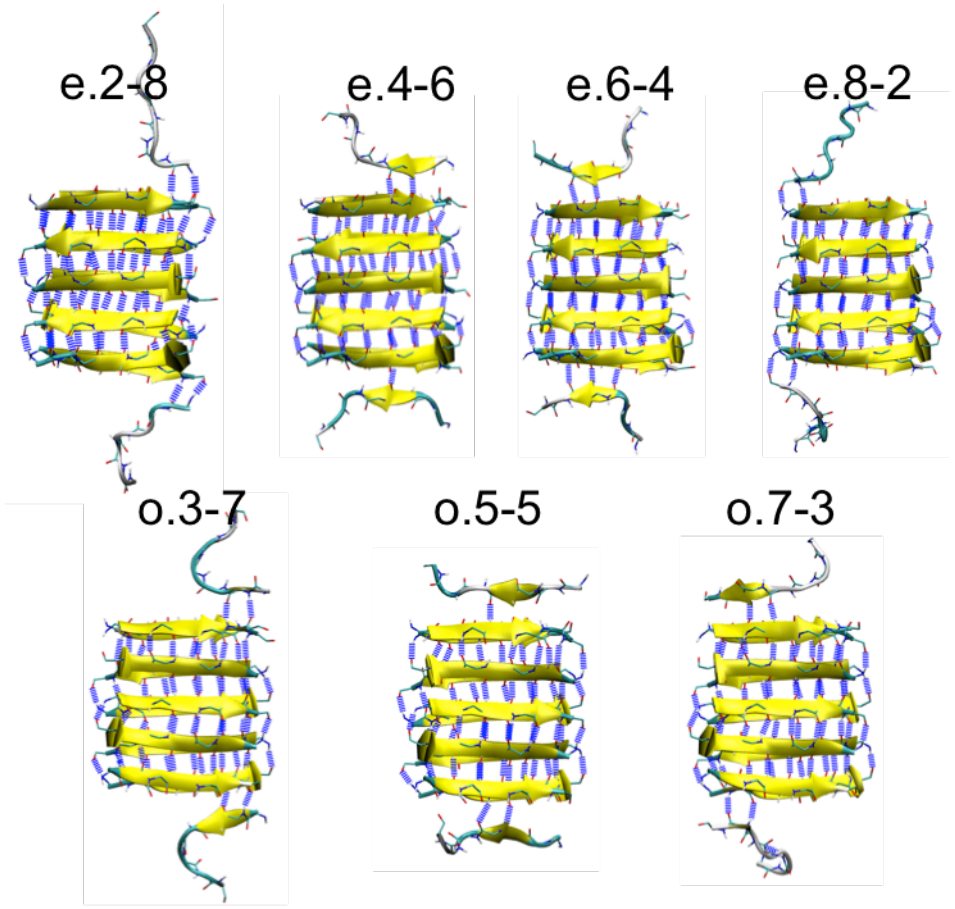
while positive or negative integers indicate a strand mis-registered in the direction of its C-terminus or N-terminus, respectively. The third parameter specifies whether the residue numbers of the end strand making backbone hydrogen bonds with the core are even or odd. Likewise, the fourth parameter is even or odd based on the residue numbers of the core strand that form backbone hydrogen bonds with the end strand.

This notation system can describe any register state for a particular peptide. The difference between the notation outlined here and the one detailed in the previous section is that the initial single contact between the end strand and amyloid core is not specified in this system. In addition, the notation described earlier is specific to the $A\beta_{16-22}$ peptide; a different system would need to be created to describe other sequences. As an example to delineate the two different systems, the notation "anti-parallel|0|even|even" describes the in-register end peptide shown in Figure A.2. That same peptide is also capable of being classified as configuration e.2-8, e.4-6, e.6-4, or e.8-2, depending on the initial contact with which it was set up.

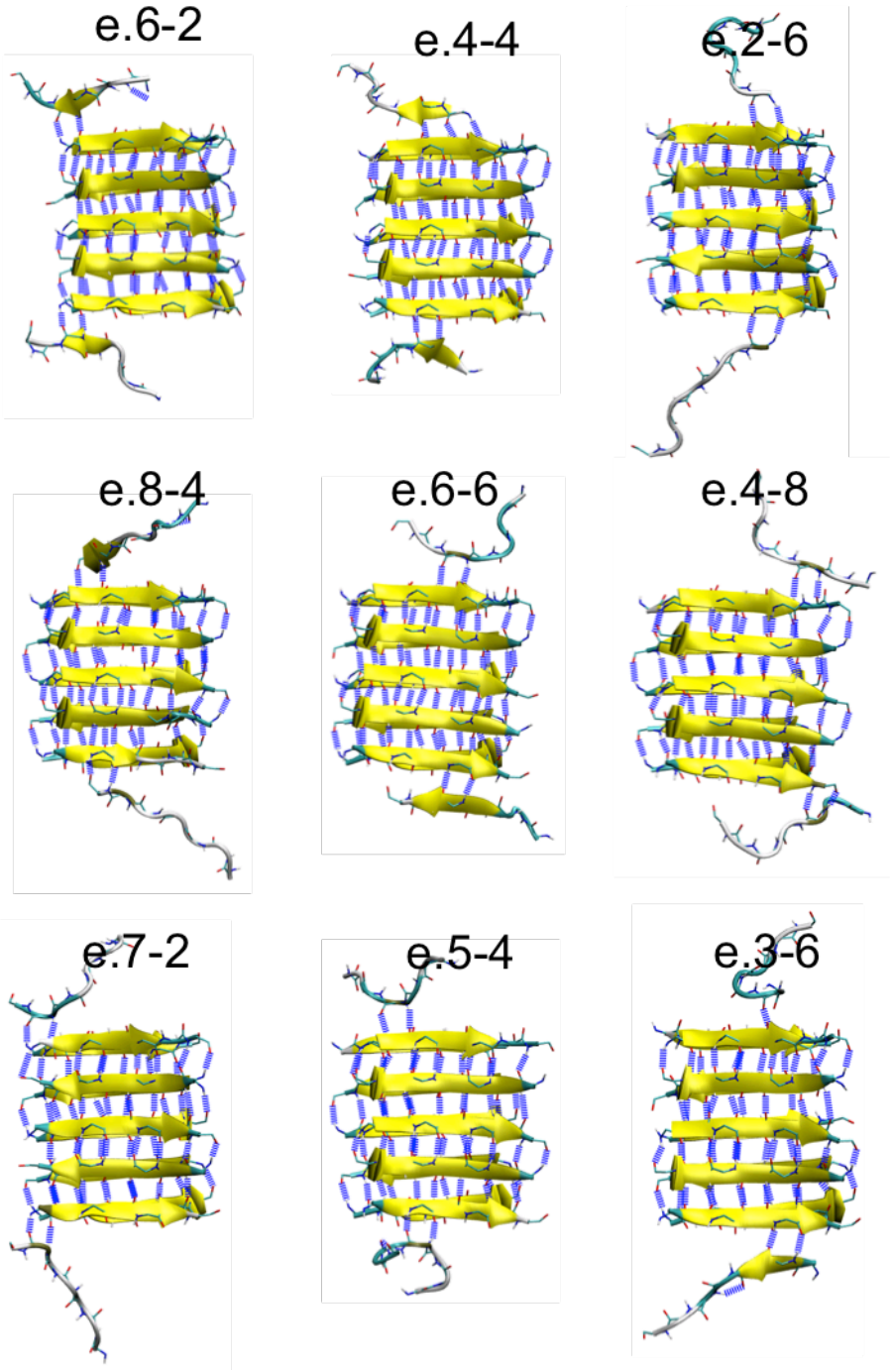
A.3 Configuration Figures

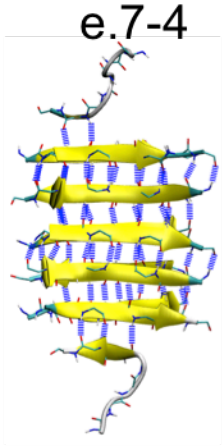
The following figures depict a representative structure of all the configurations that were simulated as described in section 3.3 at the end of the 100 picosecond heating simulation. The dashed blue bonds represent backbone hydrogen bonds. The heated simulation produces a structure that mimics an incoming peptide making a single contact with the amyloid core.

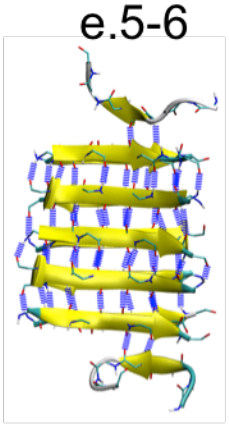
A.3.1 In-register, Anti-parallel Configurations

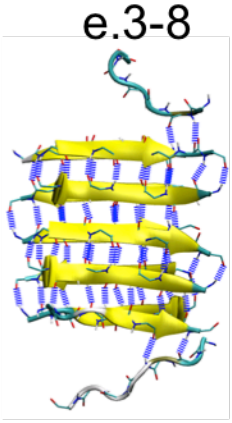


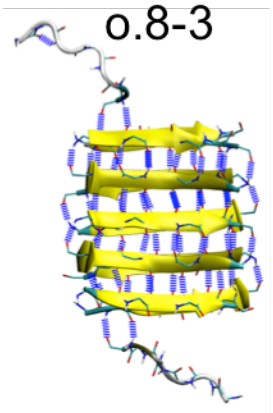
A.3.2 Mis-registered, Anti-parallel Configurations

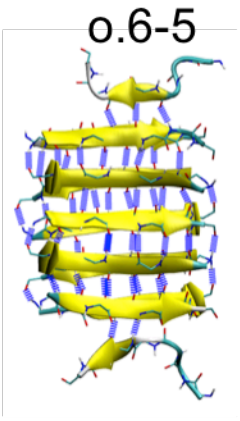


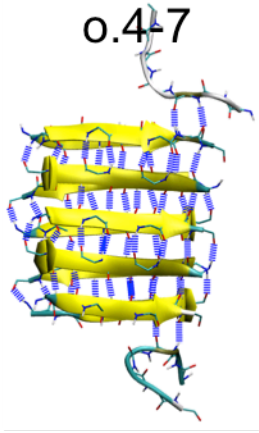

e.7-4

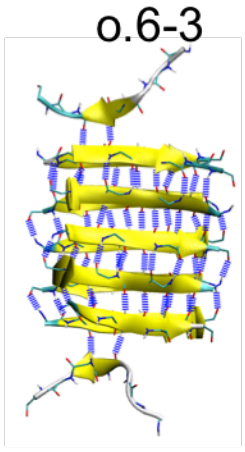

e.5-6

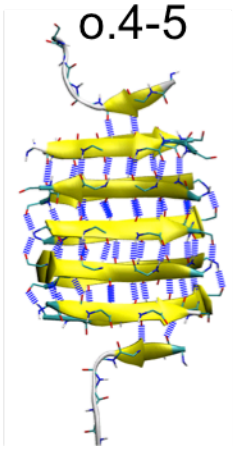

e.3-8

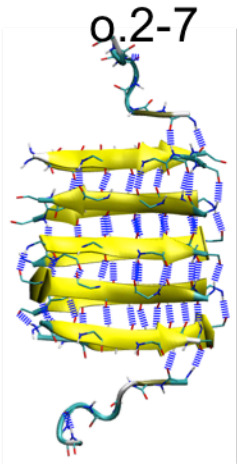

o.8-3


o.6-5


o.4-7


o.6-3


o.4-5


o.2-7

Appendix B - Conformational Transition Data

Averaged transitions
FCLi=Initial free chain length, refers to how many "free" dissociated residues there are on end of incoming peptide
FCLf=Final free chain length refers to how many "free" dissociated residues there are on end of incoming peptide
Ts=Residue type on end strand, Tc=Residue type on amyloid core
For bond breaks, Ts:Tc refers to the residue pair of the breaking bond
For bond formations, Ts:Tc refers to the residue pair of the bond that is forming
Simulations were 50ns long
Solid red refers >= 100% increase in time from wt (slower)
Solid green refers >= 100% decrease in time from wt (faster)

BOND BREAKING

				AvgTime(ps)							NumberofTransitions			
FCLi	FCLf	Register	Ts:Tc	WT	CHA19		CHA1920		CHA20		WT	CHA19	CHA1920	CHA20
0	2	inregister	GLU:LYS	1443.1	1606.6	11%	2241.4	55%	2098.2	45%	15500	13808	10009	10490
0	2	inregister	LYS:GLU	1973.4	2231.0	13%	2849.6	44%	2969.8	50%	8928	8111	6205	5890
0	2	misregister	GLU:LEU	341.6	334.6	-2%	440.3	29%	441.4	29%	18953	20271	18376	16673
0	2	misregister	GLU:VAL	1356.0	1496.5	10%	1814.8	34%	1818.6	34%	6341	5070	5570	5712
0	2	misregister	LYS:ALA	3511.2	3239.1	-8%	3663.3	4%	3210.2	-9%	3148	3432	3281	3352
0	2	misregister	LYS:CHA				11210.2	-9%	12033.7	-3%			657	507
0	2	misregister	LYS:PHE	12354.1	8988.6	-27%					423	739		
1	3	inregister	ALA:LEU	4193.1	4468.7	7%	3702.9	-12%	3485.5	-17%	3177	3114	3642	3793
1	3	inregister	LEU:ALA	6274.5	6500.7	4%	5961.3	-5%	6472.4	3%	1746	1710	1852	1575
1	3	misregister	ALA:LYS	3093.6	3414.8	10%	2725.7	-12%	2169.8	-30%	3672	3620	4585	5461
1	3	misregister	ALA:VAL	4039.7	4204.7	4%	6120.3	52%	4100.3	1%	2468	2524	1458	2496
1	3	misregister	LEU:CHA				6037.8	21%	4663.3	-7%			1637	2293
1	3	misregister	LEU:GLU	1901.0	1585.1	-17%	1625.1	-15%	2299.8	21%	5048	5751	5684	4608
1	3	misregister	LEU:PHE	5002.0	6052.3	21%					2025	1579		
2	4	inregister	CHA:VAL				3313.1	410%	1581.2	143%			142	139
2	4	inregister	PHE:VAL	650.3	2095.3	222%					177	133		
2	4	inregister	VAL:CHA				297.8	18%	234.8	-7%			295	975
2	4	inregister	VAL:PHE	253.4	529.0	109%					1315	652		
2	4	misregister	CHA:CHA				985.8	16%	580.1	-32%			333	349
2	4	misregister	CHA:LEU				6093.3	75%	3892.8	12%			1694	2199
2	4	misregister	CHA:LYS				5104.3	4%	6587.8	34%			2093	1737
2	4	misregister	CHA:PHE						1693.7	99%				120
2	4	misregister	PHE:CHA		1568.5	85%						88		
2	4	misregister	PHE:LEU	3472.0	5439.3	57%					3061	1934		
2	4	misregister	PHE:LYS	4929.1	4591.7	-7%					1919	1852		
2	4	misregister	PHE:PHE	849.4	824.0	-3%					426	537		
2	4	misregister	VAL:ALA	304.4	447.7	47%	732.3	141%	496.7	63%	8564	8127	7158	7785
2	4	misregister	VAL:CHA		665.0	113%	665.7	113%				185	272	
2	4	misregister	VAL:GLU	1393.2	1346.7	-3%	1642.1	18%	2207.3	58%	3743	3694	4285	2920
2	4	misregister	VAL:PHE	312.1					543.3	74%	253			300
2	4	misregister	VAL:VAL	700.7	645.7	-8%	1668.8	138%	1112.8	59%	103	135	84	60
3	5	inregister	CHA:CHA		1289.3	-22%	1373.1	-17%				222	289	
3	5	inregister	PHE:PHE	1653.8					1387.9	-16%	161			195
3	5	misregister	CHA:CHA				2615.6	55%					170	
3	5	misregister	CHA:PHE		1742.6	3%						173		
3	5	misregister	CHA:VAL		1056.0	33%	1482.5	87%				280	301	
3	5	misregister	PHE:CHA						1371.2	-19%				175
3	5	misregister	PHE:PHE	1687.3							276			
3	5	misregister	PHE:VAL	793.2					778.7	-2%	240			293
4	6	inregister	CHA:VAL				3433.1	-4%	4844.2	35%			91	80
4	6	inregister	PHE:VAL	3575.7	2893.8	-19%					148	161		
4	6	inregister	VAL:CHA				946.2	21%	1439.9	85%			46	44
4	6	inregister	VAL:PHE	779.9	658.9	-16%					67	41		
4	6	misregister	CHA:CHA				2422.4	3%	2519.4	7%			536	283
4	6	misregister	CHA:PHE						2796.8	19%				189
4	6	misregister	PHE:CHA		2380.5	1%						173		
4	6	misregister	PHE:PHE	2356.5	2616.9	11%					627	515		
4	6	misregister	VAL:CHA		1134.4	6%	802.3	-25%				104	84	
4	6	misregister	VAL:PHE	1072.3					1349.3	26%	192			162
4	6	misregister	VAL:VAL	1401.7	870.3	-38%	1255.3	-10%	2352.9	68%	266	460	353	201

BOND BREAKING

				AvgTime(ps)							NumberofTransitions			
FCLi	FCLf	Register	Ts:Tc	WT	CHA19		CHA1920		CHA20		WT	CHA19	CHA1920	CHA20
bound	unbound	inregister	ALA:LEU	849.1	1540.8	81%	961.4	13%	892.6	5%	248	201	293	327
bound	unbound	inregister	CHA:CHA		552.1	-2%	659.0	17%				134	167	
bound	unbound	inregister	CHA:VAL				604.3	233%	836.5	361%			74	91
bound	unbound	inregister	GLU:LYS	397.4	307.8	-23%	308.3	-22%	382.4	-4%	358	596	300	344
bound	unbound	inregister	LEU:ALA	353.2	585.5	66%	681.2	93%	400.4	13%	98	79	90	113
bound	unbound	inregister	LYS:GLU	965.2	549.0	-43%	1241.2	29%	1264.7	31%	137	81	116	118
bound	unbound	inregister	PHE:PHE	565.3					717.5	27%	163			167
bound	unbound	inregister	PHE:VAL	181.6	294.9	62%					115	71		
bound	unbound	inregister	VAL:CHA				454.6	11%	462.6	13%			36	53
bound	unbound	inregister	VAL:PHE	409.7	385.8	-6%					95	44		
bound	unbound	misregister	ALA:LYS	711.0	750.0	5%	578.5	-19%	575.9	-19%	133	139	98	143
bound	unbound	misregister	ALA:VAL	1224.2	1525.5	25%	3996.3	226%	2234.7	83%	272	124	133	176
bound	unbound	misregister	CHA:CHA				422.4	-1%	781.0	83%			605	224
bound	unbound	misregister	CHA:LEU				2670.1	67%	2119.7	32%			234	325
bound	unbound	misregister	CHA:LYS				915.0	-53%	1206.5	-38%			116	99
bound	unbound	misregister	CHA:PHE		498.1	17%			456.6	7%		118		279
bound	unbound	misregister	CHA:VAL		217.5	-58%	341.4	-34%				83	88	
bound	unbound	misregister	GLU:LEU	214.7	204.7	-5%	235.9	10%	269.2	25%	174	158	175	177
bound	unbound	misregister	GLU:VAL	283.8	392.2	38%	394.1	39%	259.4	-9%	367	191	203	365
bound	unbound	misregister	LEU:CHA				731.4	17%	854.9	37%			193	346
bound	unbound	misregister	LEU:GLU	644.5	1036.0	61%	1071.1	66%	2525.2	292%	131	173	144	81
bound	unbound	misregister	LEU:PHE	624.9	708.6	13%					227	143		
bound	unbound	misregister	LYS:ALA	349.9	205.4	-41%	343.5	-2%	275.1	-21%	196	69	187	209
bound	unbound	misregister	LYS:CHA				877.3	61%	902.6	65%			206	219
bound	unbound	misregister	LYS:PHE	545.4	701.9	29%					254	214		
bound	unbound	misregister	PHE:CHA		382.3	-10%			722.1	69%		304		119
bound	unbound	misregister	PHE:LEU	1600.3	1919.8	20%					312	228		
bound	unbound	misregister	PHE:LYS	1937.3	1914.5	-1%					143	103		
bound	unbound	misregister	PHE:PHE	426.6	352.3	-17%					604	160		
bound	unbound	misregister	PHE:VAL	515.3					256.0	-50%	125			110
bound	unbound	misregister	VAL:ALA	178.8	169.4	-5%	271.7	52%	176.8	-1%	50	52	52	36
bound	unbound	misregister	VAL:CHA		424.6	5%	489.6	22%				65	66	
bound	unbound	misregister	VAL:GLU	508.6	259.5	-49%	518.2	2%	425.2	-16%	87	71	82	97
bound	unbound	misregister	VAL:PHE	402.7					488.3	21%	132			117
bound	unbound	misregister	VAL:VAL	327.4	419.6	28%	405.0	24%	894.9	173%	23	41	30	35

BOND FORMING

				AvgTime(ps)							NumberofTransitions			
FCLi	FCLf	Register	Ts:Tc	WT	CHA19		CHA1920		CHA20		WT	CHA19	CHA1920	CHA20
2	0	inregister	GLU:LYS	75.2	77.4	3%	89.8	19%	88.2	17%	15757	14071	10258	10747
2	0	inregister	LYS:GLU	159.1	134.2	-16%	149.3	-6%	174.3	10%	9181	8379	6521	6166
2	0	misregister	GLU:LEU	213.1	198.9	-7%	175.0	-18%	192.9	-9%	19003	20310	18433	16730
2	0	misregister	GLU:VAL	223.6	260.0	16%	258.6	16%	195.6	-13%	6480	5202	5719	5820
2	0	misregister	LYS:ALA	137.0	130.8	-5%	209.6	53%	213.5	56%	3215	3530	3393	3420
2	0	misregister	LYS:CHA				229.4	14%	162.7	-19%			771	620
2	0	misregister	LYS:PHE	200.7	154.6	-23%					533	843		
3	1	inregister	ALA:LEU	105.8	138.7	31%	140.7	33%	115.3	9%	3291	3278	3792	3885
3	1	inregister	LEU:ALA	374.6	480.6	28%	516.5	38%	514.0	37%	1837	1835	1970	1687
3	1	misregister	ALA:LYS	124.2	143.7	16%	120.3	-3%	120.7	-3%	3710	3712	4665	5510
3	1	misregister	ALA:VAL	200.7	168.4	-16%	233.7	16%	265.4	32%	2575	2634	1622	2622
3	1	misregister	LEU:CHA				303.7	-6%	300.0	-7%			1829	2445
3	1	misregister	LEU:GLU	293.3	330.1	13%	319.9	9%	278.4	-5%	5176	5908	5843	4711
3	1	misregister	LEU:PHE	322.1	309.2	-4%					2191	1758		
4	2	inregister	CHA:VAL				409.6	-6%	352.7	-19%			334	318
4	2	inregister	PHE:VAL	435.8	314.3	-28%					378	327		
4	2	inregister	VAL:CHA				731.3	76%	440.5	6%			505	1130
4	2	inregister	VAL:PHE	415.6	442.6	6%					1496	831		
4	2	misregister	CHA:CHA				610.7	7%	565.1	-1%			462	396
4	2	misregister	CHA:LEU				309.8	2%	322.7	6%			1888	2371
4	2	misregister	CHA:LYS				271.0	-16%	310.8	-3%			2223	1886
4	2	misregister	CHA:PHE						889.7	55%				175
4	2	misregister	PHE:CHA		529.2	-8%						156		
4	2	misregister	PHE:LEU	305.2	298.5	-2%					3209	2092		
4	2	misregister	PHE:LYS	320.9	293.3	-9%					2041	1953		
4	2	misregister	PHE:PHE	572.7	487.7	-15%					526	590		
4	2	misregister	VAL:ALA	675.4	630.7	-7%	644.3	-5%	618.9	-8%	8547	8110	7151	7773
4	2	misregister	VAL:CHA		1899.7	-26%	1730.2	-33%				237	305	
4	2	misregister	VAL:GLU	732.4	537.1	-27%	513.2	-30%	852.5	16%	3787	3750	4345	3002
4	2	misregister	VAL:PHE	2573.3					1877.2	-27%	268			312
4	2	misregister	VAL:VAL	1936.8	858.4	-56%	1089.4	-44%	2628.5	36%	152	194	156	127
5	3	inregister	CHA:CHA		2408.3	-18%	2259.9	-23%				355	410	
5	3	inregister	PHE:PHE	2938.4					3455.1	18%	264			275
5	3	misregister	CHA:CHA				1804.8	-14%					299	
5	3	misregister	CHA:PHE		2026.3	-4%						301		
5	3	misregister	CHA:VAL		1282.9	-37%	1336.0	-35%				383	411	
5	3	misregister	PHE:CHA						2000.3	-5%				292
5	3	misregister	PHE:PHE	2101.3							358			
5	3	misregister	PHE:VAL	2040.0					1474.4	-28%	301			348
6	4	inregister	CHA:VAL				3157.1	41%	2642.2	18%			184	175
6	4	inregister	PHE:VAL	2236.1	2704.1	21%					232	247		
6	4	inregister	VAL:CHA				4052.8	52%	3886.8	45%			111	97
6	4	inregister	VAL:PHE	2673.0	5063.7	89%					134	102		
6	4	misregister	CHA:CHA				1656.7	-20%	2455.0	19%			620	326
6	4	misregister	CHA:PHE						2848.6	38%				211
6	4	misregister	PHE:CHA		3801.9	84%						229		
6	4	misregister	PHE:PHE	2071.4	788.1	-62%					674	533		
6	4	misregister	VAL:CHA		2744.9	63%	2069.3	23%				162	151	
6	4	misregister	VAL:PHE	1679.8					1303.0	-22%	238			225
6	4	misregister	VAL:VAL	2954.3	2196.4	-26%	2056.8	-30%	3415.5	16%	292	481	395	256

References

[1] Carrell, R. W., and Lomas, D. A. (1997) Conformational disease, *Lancet 350*, 134-138.

[2] Dobson, C. M. (1999) Protein misfolding, evolution and disease, *Trends Biochem Sci 24*, 329-332.

[3] Soto, C. (2001) Protein misfolding and disease; protein refolding and therapy, *Febs Lett 498*, 204-207.

[4] Forno, L. S. (1996) Neuropathology of Parkinson's disease, *J Neuropath Exp Neur 55*, 259-272.

[5] Gutekunst, C. A., Li, S. H., Yi, H., Mulroy, J. S., Kuemmerle, S., Jones, R., Rye, D., Ferrante, R. J., Hersch, S. M., and Li, X. J. (1999) Nuclear and neuropil aggregates in Huntington's disease: Relationship to neuropathology, *J Neurosci 19*, 2522-2534.

[6] Terry, R. D., Masliah, E., Salmon, D. P., Butters, N., Deteresa, R., Hill, R., Hansen, L. A., and Katzman, R. (1991) Physical Basis of Cognitive Alterations in Alzheimers-Disease - Synapse Loss Is the Major Correlate of Cognitive Impairment, *Ann Neurol 30*, 572-580.

[7] Hardy, J., and Gwinn-Hardy, K. (1998) Genetic classification of primary neurodegenerative disease, *Science 282*, 1075-1079.

[8] Goate, A., Chartierharlin, M. C., Mullan, M., Brown, J., Crawford, F., Fidani, L., Giuffra, L., Haynes, A., Irving, N., James, L., Mant, R., Newton, P., Rooke, K., Roques, P., Talbot, C., Pericakvance, M., Roses, A., Williamson, R., Rossor, M., Owen, M., and Hardy, J. (1991) Segregation of a Missense Mutation in the Amyloid Precursor Protein Gene with Familial Alzheimers-Disease, *Nature 349*, 704-706.

[9] Hsiao, K., Baker, H. F., Crow, T. J., Poulter, M., Owen, F., Terwilliger, J. D., Westaway, D., Ott, J., and Prusiner, S. B. (1989) Linkage of a Prion Protein Missense Variant to Gerstmann-Straussler Syndrome, *Nature 338*, 342-345.

[10] Macdonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S. A., James, M., Groot, N., Macfarlane, H., Jenkins, B., Anderson, M. A., Wexler, N. S., Gusella, J. F., Bates, G. P., Baxendale, S., Hummerich, H., Kirby, S., North, M., Youngman, S., Mott, R., Zehetner, G., Sedlacek, Z., Poustka, A., Frischauf, A. M., Lehrach, H., Buckler, A. J., Church, D., Doucettestamm, L., Odonovan, M. C., Ribaramirez, L., Shah, M., Stanton, V. P., Strobel, S. A., Draths, K. M., Wales, J. L., Dervan, P., Housman, D. E., Altherr, M., Shiang, R., Thompson, L., Fielder, T., Wasmuth, J. J., Tagle, D., Valdes, J., Elmer, L., Allard, M., Castilla, L., Swaroop, M., Blanchard, K., Collins, F. S., Snell, R., Holloway, T., Gillespie, K., Datson, N., Shaw, D., and Harper, P. S. (1993) A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntingtons-Disease Chromosomes, *Cell 72*, 971-983.

[11] Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., Stenroos, E. S., Chandrasekharappa, S., Athanassiadou, A., Papapetropoulos, T., Johnson, W. G., Lazzarini, A. M., Duvoisin, R. C., DiIorio, G., Golbe, L. I., and Nussbaum, R. L. (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease, *Science 276*, 2045-2047.

[12] Rosen, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., Oregan, J. P., Deng, H. X., Rahmani, Z., Krizus, A., Mckennayasek, D., Cayabyab, A., Gaston, S. M., Berger, R., Tanzi, R. E., Halperin, J. J., Herzfeldt, B., Vandenbergh, R., Hung, W. Y., Bird, T., Deng, G., Mulder, D. W., Smyth, C., Laing, N. G., Soriano, E., Pericakvance, M. A., Haines, J., Rouleau, G. A., Gusella, J.

S., Horvitz, H. R., and Brown, R. H. (1993) Mutations in Cu/Zn Superoxide-Dismutase Gene Are Associated with Familial Amyotrophic-Lateral-Sclerosis, *Nature 362*, 59-62.

[13] Davies, S. W., Turmaine, M., Cozens, B. A., DiFiglia, M., Sharp, A. H., Ross, C. A., Scherzinger, E., Wanker, E. E., Mangiarini, L., and Bates, G. P. (1997) Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation, *Cell 90*, 537-548.

[14] Games, D., Adams, D., Alessandrini, R., Barbour, R., Berthelette, P., Blackwell, C., Carr, T., Clemens, J., Donaldson, T., Gillespie, F., Guido, T., Hagopian, S., Johnsonwood, K., Khan, K., Lee, M., Leibowitz, P., Lieberburg, I., Little, S., Masliah, E., Mcconlogue, L., Montoyazavala, M., Mucke, L., Paganini, L., Penniman, E., Power, M., Schenk, D., Seubert, P., Snyder, B., Soriano, F., Tan, H., Vitale, J., Wadsworth, S., Wolozin, B., and Zhao, J. (1995) Alzheimer-Type Neuropathology in Transgenic Mice Overexpressing V717f Beta-Amyloid Precursor Protein, *Nature 373*, 523-527.

[15] Gurney, M. E., Pu, H. F., Chiu, A. Y., Dalcanto, M. C., Polchow, C. Y., Alexander, D. D., Caliendo, J., Hentati, A., Kwon, Y. W., Deng, H. X., Chen, W. J., Zhai, P., Sufit, R. L., and Siddique, T. (1994) Motor-Neuron Degeneration in Mice That Express a Human Cu,Zn Superoxide-Dismutase Mutation, *Science 264*, 1772-1775.

[16] Hsiao, K. K., Scott, M., Foster, D., Groth, D. F., Dearmond, S. J., and Prusiner, S. B. (1990) Spontaneous Neurodegeneration in Transgenic Mice with Mutant Prion Protein, *Science 250*, 1587-1590.

[17] Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trottier, Y., Lehrach, H., Davies, S. W., and Bates, G. P. (1996) Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice, *Cell 87*, 493-506.

[18] Masliah, E., Rockenstein, E., Veinbergs, I., Mallory, M., Hashimoto, M., Takeda, A., Sagara, Y., Sisk, A., and Mucke, L. (2000) Dopaminergic loss and inclusion body formation in alpha-synuclein mice: Implications for neurodegenerative disorders, *Science 287*, 1265-1269.

[19] Klement, I. A., Skinner, P. J., Kaytor, M. D., Yi, H., Hersch, S. M., Clark, H. B., Zoghbi, H. Y., and Orr, H. T. (1998) Ataxin-1 nuclear localization and aggregation: Role in polyglutamine-induced disease in SCA1 transgenic mice, *Cell 95*, 41-53.

[20] Moechars, D., Dewachter, I., Lorent, K., Reverse, D., Baekelandt, V., Naidu, A., Tesseur, I., Spittaels, K., Van Den Haute, C., Checler, F., Godaux, E., Cordell, B., and Van Leuven, F. (1999) Early phenotypic changes in transgenic mice that overexpress different mutants of amyloid precursor protein in brain, *J Biol Chem 274*, 6483-6492.

[21] Cohen, A. S., and Calkins, E. (1959) Electron Microscopic Observations on a Fibrous Component in Amyloid of Diverse Origins, *Nature 183*, 1202-1203.

[22] Serpell, L. C., Blake, C. C. F., and Fraser, P. E. (2000) Molecular structure of a fibrillar Alzheimer's A beta fragment, *Biochemistry-Us 39*, 13269-13275.

[23] Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., and Blake, C. C. F. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction, *J Mol Biol 273*, 729-739.

[24] Serpell, L. C., Berriman, J., Jakes, R., Goedert, M., and Crowther, R. A. (2000) Fiber diffraction of synthetic alpha-synuclein filaments shows amyloid-like cross-beta conformation, *P Natl Acad Sci USA 97*, 4897-4902.

[25] Teplow, D. B. (1998) Structural and kinetic features of amyloid beta-protein fibrillogenesis, *Amyloid 5*, 121-142.
[26] Castano, E. M., Ghiso, J., Prelli, F., Gorevic, P. D., Migheli, A., and Frangione, B. (1986) Invitro Formation of Amyloid Fibrils from 2 Synthetic Peptides of Different Lengths Homologous to Alzheimers-Disease Beta-Protein, *Biochem Bioph Res Co 141*, 782-789.
[27] Hilbich, C., Kisterswoike, B., Reed, J., Masters, C. L., and Beyreuther, K. (1992) Substitutions of Hydrophobic Amino-Acids Reduce the Amyloidogenicity of Alzheimers-Disease Beta-A4 Peptides, *J Mol Biol 228*, 460-473.
[28] Jarrett, J. T., Berger, E. P., and Lansbury, P. T. (1993) The C-Terminus of the Beta-Protein Is Critical in Amyloidogenesis, *Ann Ny Acad Sci 695*, 144-148.
[29] Soto, C., Castano, E. M., Frangione, B., and Inestrosa, N. C. (1995) The Alpha-Helical to Beta-Strand Transition in the Amino-Terminal Fragment of the Amyloid Beta-Peptide Modulates Amyloid Formation, *J Biol Chem 270*, 3063-3067.
[30] Wood, S. J., Wetzel, R., Martin, J. D., and Hurle, M. R. (1995) Prolines and Amyloidogenicity in Fragments of the Alzheimers Peptide Beta/A4, *Biochemistry-Us 34*, 724-730.
[31] Zoghbi, H. Y., and Orr, H. T. (2000) Glutamine repeats and neurodegeneration, *Annu Rev Neurosci 23*, 217-247.
[32] Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G. P., Davies, S. W., Lehrach, H., and Wanker, E. E. (1997) Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo, *Cell 90*, 549-558.
[33] Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. (1994) Glutamine Repeats as Polar Zippers - Their Possible Role in Inherited Neurodegenerative Diseases, *P Natl Acad Sci USA 91*, 5355-5358.
[34] Mrak, R. E., Griffin, W. S. T., and Graham, D. I. (1997) Aging-associated changes in human brain, *J Neuropath Exp Neur 56*, 1269-1275.
[35] Scherzinger, E., Sittler, A., Schweiger, K., Heiser, V., Lurz, R., Hasenbank, R., Bates, G. P., Lehrach, H., and Wanker, E. E. (1999) Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: Implications for Huntington's disease pathology, *P Natl Acad Sci USA 96*, 4604-4609.
[36] Wood, S. J., Wypych, J., Steavenson, S., Louis, J. C., Citron, M., and Biere, A. L. (1999) alpha-synuclein fibrillogenesis is nucleation-dependent - Implications for the pathogenesis of Parkinson's disease, *J Biol Chem 274*, 19509-19512.
[37] Kuo, Y. M., Emmerling, M. R., VigoPelfrey, C., Kasunic, T. C., Kirkpatrick, J. B., Murdoch, G. H., Ball, M. J., and Roher, A. E. (1996) Water-soluble A beta (N-40, N-42) oligomers in normal and Alzheimer disease brains, *J Biol Chem 271*, 4077-4081.
[38] Lambert, M. P., Barlow, A. K., Chromy, B. A., Edwards, C., Freed, R., Liosatos, M., Morgan, T. E., Rozovsky, I., Trommer, B., Viola, K. L., Wals, P., Zhang, C., Finch, C. E., Krafft, G. A., and Klein, W. L. (1998) Diffusible, nonfibrillar ligands derived from A beta(1-42) are potent central nervous system neurotoxins, *P Natl Acad Sci USA 95*, 6448-6453.
[39] Levine, H. (1995) Soluble Multimeric Alzheimer Beta(1-40) Pre-Amyloid Complexes in Dilute-Solution, *Neurobiol Aging 16*, 755-764.

[40] Walsh, D. M., Klyubin, I., Fadeeva, J. V., Cullen, W. K., Anwyl, R., Wolfe, M. S., Rowan, M. J., and Selkoe, D. J. (2002) Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo, *Nature 416*, 535-539.
[41] Walsh, D. M., Hartley, D. M., Kusumoto, Y., Fezoui, Y., Condron, M. M., Lomakin, A., Benedek, G. B., Selkoe, D. J., and Teplow, D. B. (1999) Amyloid beta-protein fibrillogenesis - Structure and biological activity of protofibrillar intermediates, *J Biol Chem 274*, 25945-25952.
[42] Harper, J. D., Wong, S. S., Lieber, C. M., and Lansbury, P. T. (1999) Assembly of A beta amyloid protofibrils: An in vitro model for a possible early event in Alzheimer's disease, *Biochemistry-Us 38*, 8972-8980.
[43] Martin, J. B. (1999) Molecular basis of the neurodegenerative disorders, *New Engl J Med 340*, 1970-1980.
[44] Tuckerman, M. E., and Martyna, G. J. (2000) Understanding modern molecular dynamics: Techniques and applications, *J Phys Chem B 104*, 159-178.
[45] Straatsma, T. P., and Mccammon, J. A. (1992) Computational Alchemy, *Annu Rev Phys Chem 43*, 407-435.
[46] Simonson, T., Archontis, G., and Karplus, M. (2002) Free energy simulations come of age: Protein-ligand recognition, *Accounts Chem Res 35*, 430-437.
[47] Berendsen, H. J. C., and Hayward, S. (2000) Collective protein dynamics in relation to function, *Curr Opin Struc Biol 10*, 165-169.
[48] Roux, B., and Simonson, T. (1999) Implicit solvent models, *Biophys Chem 78*, 1-20.
[49] Eisenberg, D., and Mclachlan, A. D. (1986) Solvation Energy in Protein Folding and Binding, *Nature 319*, 199-203.
[50] Nguyen, H. D., and Hall, C. K. (2004) Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides, *P Natl Acad Sci USA 101*, 16180-16185.
[51] Gsponer, J., Haberthur, U., and Caflisch, A. (2003) 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, *P Natl Acad Sci USA 100*, 5154-5159.
[52] Klimov, D. K., and Thirumalai, D. (2003) Dissecting the assembly of A beta(16-22) amyloid peptides into antiparallel beta sheets, *Structure 11*, 295-307.
[53] Ma, B. Y., and Nussinov, R. (2002) Stabilities and conformations of Alzheimer's beta-amyloid peptide oligomers (A beta(16-22 ') A beta(16-35 ') and A beta(10-35)): Sequence effects, *P Natl Acad Sci USA 99*, 14126-14131.
[54] Liang, Y., Pingali, S. V., Jogalekar, A. S., Snyder, J. P., Thiyagarajan, P., and Lynn, D. G. (2008) Cross-strand pairing and amyloid assembly, *Biochemistry-Us 47*, 10018-10026.
[55] Petkova, A. T., Buntkowsky, G., Dyda, F., Leapman, R. D., Yau, W. M., and Tycko, R. (2004) Solid state NMR reveals a pH-dependent antiparallel beta-sheet registry in fibrils formed by a beta-amyloid peptide, *J Mol Biol 335*, 247-260.
[56] Mehta, A. K., Lu, K., Childers, W. S., Liang, Y., Dublin, S. N., Dong, J. J., Snyder, J. P., Pingali, S. V., Thiyagarajan, P., and Lynn, D. G. (2008) Facial symmetry in protein self-assembly, *J Am Chem Soc 130*, 9829-9835.
[57] Balbach, J. J., Ishii, Y., Antzutkin, O. N., Leapman, R. D., Rizzo, N. W., Dyda, F., Reed, J., and Tycko, R. (2000) Amyloid fibril formation by A beta(16-22), a seven-residue fragment of the Alzheimer's beta-amyloid peptide, and structural characterization by solid state NMR, *Biochemistry-Us 39*, 13748-13759.

[58] Nguyen, P. H., Li, M. S., Stock, G., Straub, J. E., and Thirumalai, D. (2007) Monomer adds to preformed structured oligomers of A beta-peptides by a two-stage dock-lock mechanism, *P Natl Acad Sci USA 104*, 111-116.
[59] Schmit, J. D. (2013) Kinetic theory of amyloid fibril templating, *J Chem Phys 138*.
[60] Senguen, F. T., Lee, N. R., Gu, X. F., Ryan, D. M., Doran, T. M., Anderson, E. A., and Nilsson, B. L. (2011) Probing aromatic, hydrophobic, and steric effects on the self-assembly of an amyloid-beta fragment peptide, *Mol Biosyst 7*, 486-496.
[61] Gillespie, D. T. (1977) Exact Stochastic Simulation of Coupled Chemical-Reactions, *J Phys Chem-Us 81*, 2340-2361.
[62] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations, *J Comput Chem 4*, 187-217.
[63] Ferrara, P., Apostolakis, J., and Caflisch, A. (2002) Evaluation of a fast implicit solvent model for molecular dynamics simulations, *Proteins 46*, 24-33.
[64] Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., and Tycko, R. (2002) A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR, *P Natl Acad Sci USA 99*, 16742-16747.