2016

Rate Kinetics and Molecular Dynamics of the Structural Transitions in Amyloidogenic Proteins

Timothy Michael Steckmann
tstec001@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

RATE KINETICS AND MOLECULAR DYNAMICS OF THE STRUCTURAL

TRANSITIONS IN AMYLOIDOGENIC PROTEINS

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

Timothy M. Steckmann

2016

To: Dean Michael R. Heithaus
    College of Arts, Sciences, and Education

This dissertation, written by Timothy M. Steckmann, and entitled Rate Kinetics and Molecular Dynamics of the Structural Transitions in Amyloidogenic Proteins, having been approved in respect to style and intellectual content, is referred to you for judgement.

We have read this dissertation and recommend that it be approved.

_____
Jin He

_____
Jaroslava Miksovska

_____
Prem P. Chapagain, Co-major Professor

_____
Bernard S. Gerstman, Co-major Professor

Date of Defense: November 1, 2016

The dissertation of Timothy M. Steckmann is approved.

_____
Dean Michael R. Heithaus
College of Arts, Sciences, and Education

_____
Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2016

DEDICATION

I dedicate this dissertation to my beloved parents.

ACKNOWLEDGMENTS

ABSTRACT OF THE DISSERTATION

RATE KINETICS AND MOLECULAR DYNAMICS OF THE STRUCTURAL

TRANSITIONS IN AMYLOIDOGENIC PROTEINS

by

Timothy M. Steckmann

Florida International University, 2016

Miami, Florida

Professor Bernard S. Gerstman, Co-major Professor

Professor Prem P. Chapagain, Co-major Professor

Amyloid fibril aggregation is associated with several horrific diseases such as

Alzheimer's, Creutzfeld-Jacob, diabetes, Parkinson's and others. The process of

amyloid aggregation involves forming myriad different metastable intermediate

aggregates. Amyloid fibrils are composed of proteins that originate in an innocuous α-

helix or random-coil structure. The α-helices convert their structure to β-strands that

aggregate into β-sheets, and then into protofibrils, and ultimately into fully formed

amyloid fibrils. On the basis of experimental data, I have developed a mathematical

model for the kinetics of the reaction pathways and determined rate parameters for

peptide secondary structural conversion and aggregation during the entire

fibrillogenesis process from random coil to fibrils, including the molecular species that

accelerate the conversions. The specific steps of the model and the rate constants that

are determined by fitting to experimental data provide insight on the molecular species

involved in the fibril formation process. To better understand the molecular basis of the

protein structural transitions and aggregation, I report on molecular dynamics (MD)

computational studies on the formation of amyloid protofibrillar structures in the small model protein ccβ, which undergoes many of the structural transitions of the larger, naturally occurring amyloid forming proteins. Two different structural transition processes involving hydrogen bonds are observed for aggregation into fibrils: the breaking of intrachain hydrogen bonds to allow β-hairpin proteins to straighten, and the subsequent formation of interchain hydrogen bonds during aggregation into amyloid fibrils. For my MD simulations, I found that the temperature dependence of these two different structural transition processes results in the existence of a temperature window that the ccβ protein experiences during the process of forming protofibrillar structures. Both the mathematical modeling of the kinetics and the MD simulations show that molecular structural heterogeneity is a major factor in the process. The MD simulations also show that intrachain and interchain hydrogen bonds breaking and forming is strongly correlated to the process of amyloid formation.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

Aβ   Amyloid beta

AD   Alzheimer's Disease

APP   Amyloid Peptide Precursor

BSE   Bovine Spongiform Encephalopathy (mad cow disease)

CHARMM   Chemistry at Harvard Molecular Mechanics

CTMD   Constant Temperature Molecular Dynamics

EEF1   effective energy function-1

LJ   Lennard-Jones

MD   Molecular Dynamics

MMTSB   Multiscale Modeling Tools for Structural Biology

NAMD   Not Another Molecular Dynamics

NMR   Nuclear Magnetic Resonance

PBC   Periodic Boundary Condition

PDB   Protein Data Bank

PDB ID   Protein Data Bank Identification

pH   power (or potential) of Hydrogen

REMD   Replica Exchange Molecular Dynamics

rk2   Runga Kutta second-order

rk4   Runga Kutta fourth-order

TFE   Trifluoroethanol

1. INTRODUCTION

Amyloid fibrils are protein structures that are associated with many debilitating human diseases such as Alzheimer's(AD), Parkinson's, Creutzfeldt-Jakob and diabetes, and in animal diseases such as BSE.

Amyloid fibrils are composed of proteins that originate in an innocuous α-helix or random coil structure. The proteins convert their structure to β-strands that aggregate into β-sheets, and then into protofibrils, and ultimately into fully formed amyloid fibrils that are composed of about 600 protein chains. Although humans have evolved defenses against aggregation,[1] unfortunately amyloid aggregation still occurs in humans. Pre-fibrillar β-aggregate structures have been found to be highly neurotoxic, and are more neurotoxic than mature amyloid fibrils or amyloid plaques.[2,3,4] It remains uncertain which of the specific aggregates (protofibrils, protofilaments, or small oligomers) is most dangerous.[5,6,7,8] A detailed molecular-level understanding of the formation process of amyloid fibrils is crucial for determining the molecular species that are toxic and for developing methods to slow down or prevent these horrific diseases.

Research on amyloid fibril formation overlaps strongly with prion related diseases.[9] It also has potential nanoscience and material engineering applications since amyloid fibrils are especially strong mechanically.[10,11]

1.1 What Is Already Known About the Fibrillization Process

Amyloid fibrillation is a type of protein aggregation. Small β-aggregates form protofilaments which are composed of β-sheets parallel to each other. These

protofilaments then wrap around each other like strands of rope to make protofibrils.

Protofibrils grow to form fibrils, which are insoluble. All fibrils have rope-like, twisted

structures that reflect that they are assembled from filamentous structures. Fibrils can

grow either by attaching of single monomers or attaching of β-aggregates. Figure 1.1

shows schematics of a single β-sheet, a protofilament, a protofibril and a fibril.

single β-sheet   protofilament   protofibril   fibril
                  two β-sheets
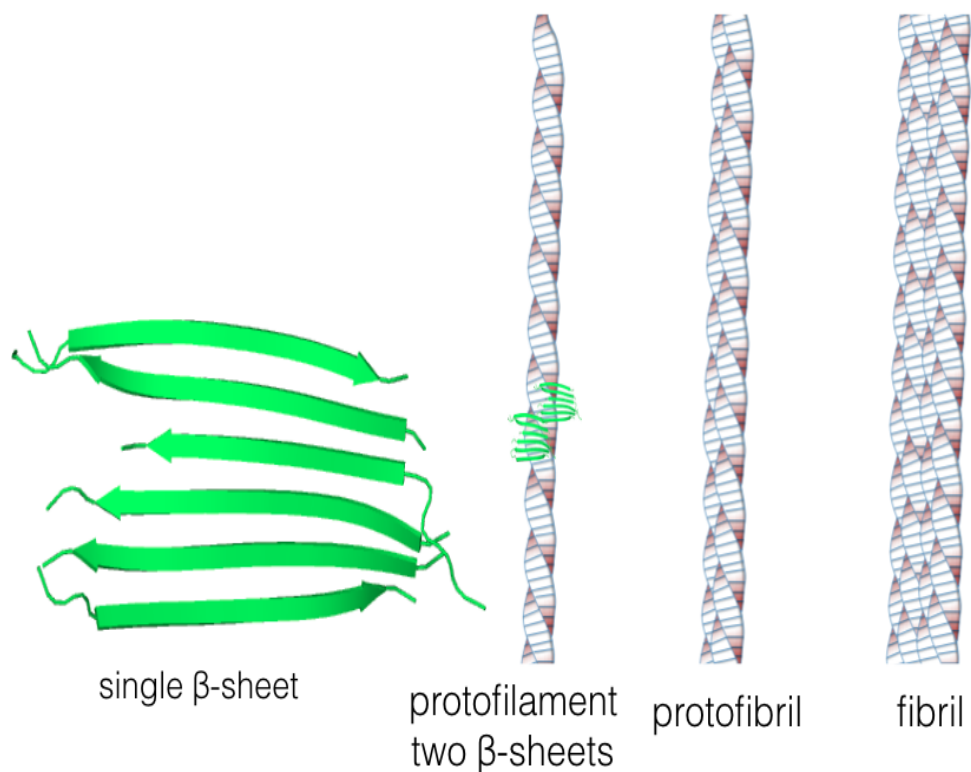
Fig. 1.1 Structure schematics. Schematic of a single β-sheet was done using Pymol.
Pictures of a protofilament, a protofibril, and a fibril were done using PowerPoint
following a diagram by David Talaga.[12]

Experimentalists have created amyloid fibrils *in vitro*. Solid state NMR

measurements have shown the general structure of fibrils. However, detailed

information about the kinetics of the fibrillization process is incomplete.

Unfortunately, *in vivo* experiments are hard to perform and yield only the concentration of fibrils, but not the concentration of the precursors. *In vitro* studies allow measurements of the concentrations of some structural stages but the aggregation kinetics are unclear since the experimental techniques do not allow a distinction between individual intermediate β-structures. Likewise, detailed information about the molecular dynamics of various stages is also incomplete. Some information is available but details on a molecular level are experimentally difficult to obtain because it is hard to crystallize various β precursor structures such as protofibrils.

Slight variations in experimental preparation procedures such as purification and synthesis may significantly change the peptide system's fibrillogenic properties. An additional complication is that a single system of proteins of the same species tends to have a heterogeneous combination of aggregates. One common observation is that increases in concentration of proteins tend to speed up aggregation.[13,14,15]

Several processes of protein aggregation often happen concomitantly with amyloid formation. Some of these processes create amorphous aggregates.[16] Although amorphous aggregates are associated with misfolding diseases, some amorphous aggregates never form into amyloids. Some aggregates are small soluble oligomers, which can be swept out by the body.[16]

1.2 Free Energy

The idea for using free energy landscapes to describe different states in the fibrillation process came from protein folding.[11] The free energy is a complicated landscape in amyloid formation since amyloid formation involves both protein folding

and several proteins interacting with each other.[11] However, useful information can still be obtained from free energy landscapes of amyloids and their precursors. An example of a free energy landscape to describe the aggregation process is given in Fig. 1.2. It is taken from Figure 8.2c of *Amyloid Fibrils and Prefibrillar Aggregates*.[11] Amyloid fibrils in Fig. 1.2 clearly correspond to the lowest free energy minima.
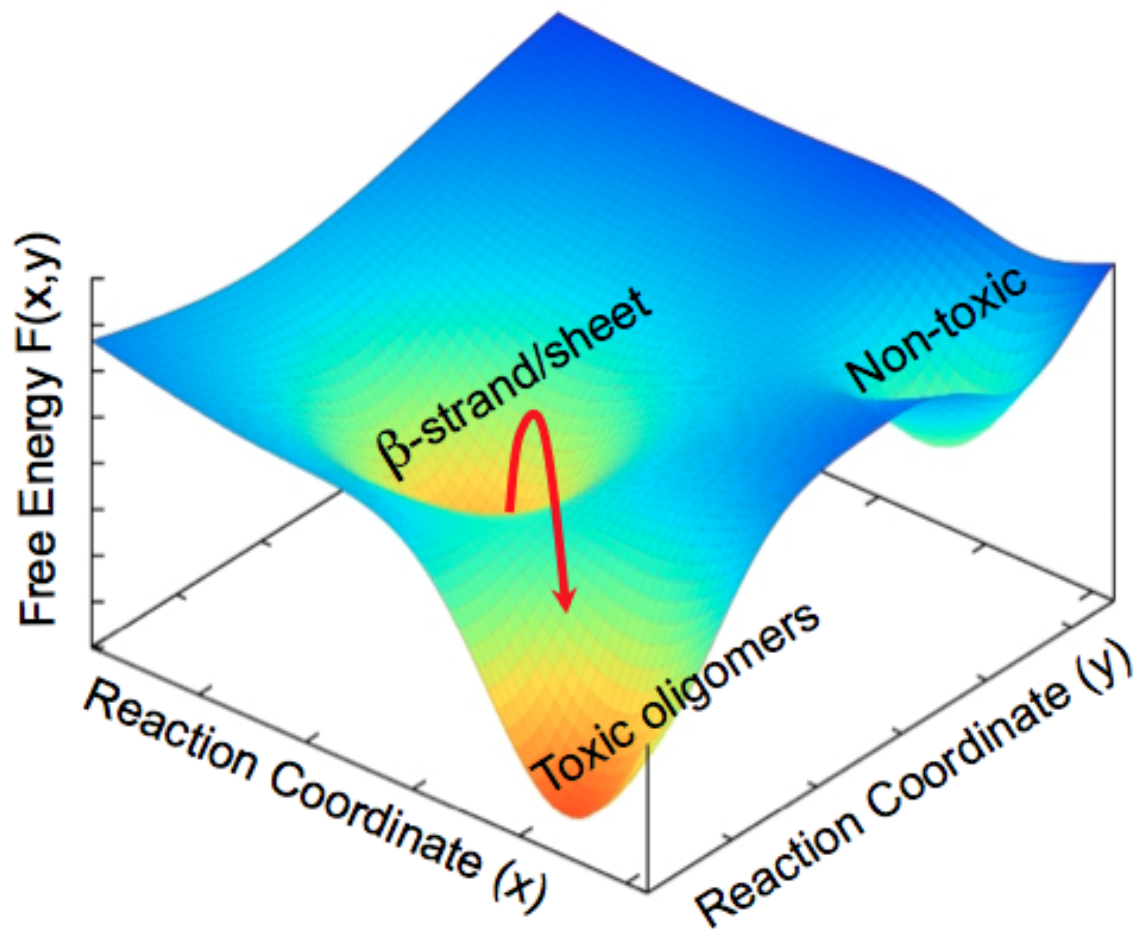
Fig. 1.2. Free energy landscape example.

1.3 Hydrogen Bonds and Aggregation

Any protein can undergo amyloid aggregation,[17] which may be because the formation and stabilization of β-sheets involves interchain hydrogen bonds between protein backbones[17] which are common to all amino acids. Different hydrogen bonding patterns between backbones correspond to the different stages in the folding/aggregation process.

Simulation of two short Aβ(16-22) peptides by Santini et al.[18] showed multiple complex routes to fibrillization involving various networks of hydrogen bonds. Conformational properties of amyloids and their precursors can be partially described by the number of interchain and intrachain hydrogen bonds. There are many intrachain hydrogen bonds when a protein is in an α-helical secondary structure or a hairpin conformation, whereas there are many interchain hydrogen bonds among proteins when they are in a β-sheet. Interchain hydrogen bonds form cooperatively with other interchain hydrogen bonds in the aggregation process.

1.4 Stages in the Fibrillization Process

1.4.1 Nucleation and Secondary Structure

According to classical nucleation theory, a nucleus is an aggregate of several proteins. It is the smallest aggregate size for which the rate of association is higher than the rate of disassociation. Once the nucleus is formed, the aggregation becomes a downhill polymerization process. What typically happens is, by chance, a group of proteins stick together with a number of proteins large enough that the energy loss is

greater than the entropy loss due to another protein joining the aggregate. Although classical nucleation theory talks about a nucleus as being made up of several proteins, a nucleus can be as small as a single protein (as in the case of transthyretin).[19] It could even be made up of a few residues of a protein.

The typical aggregation process involves some α-helices or proteins that have some regions of α-helix in them that eventually convert to β-strands as the aggregates undergo reorganization. In an article written by Lomakin et al.[2] the proteins come together to form micelles, which are loosely-bound disordered aggregates. These micelles then break apart into nuclei, as described in my article[20] and described later in Chapter 3. The initial assembly of proteins is like a micelle. The micelle or micelle-like aggregate breaks up into single proteins or small aggregates mostly in the β-strand state.[20] The process from single proteins to β-sheet is described as a "nucleated conformational conversion" by Cheon et al.[21]

1.4.2 Steric Zipper

The steric zipper conformation helps to stabilize amyloid fibrils. A steric zipper is a pair of parallel β-sheets. β sheets come together due to the hydrophobic force and their side chains interlock. Their van der Waals forces hold them together. The steric zipper has no water molecules between the two β-sheets, causing it to be a dry interface.[22,23,24,11] Two different steric zippers can join through a wet interface.[24] A wet interface has water and many hydrogen bonds between the side chains. The steric zipper is a common configuration of the protofilament.[25] Some protofilaments contain only one steric zipper whereas some contain several.

1.4.3 Hairpin Straightening

Hairpins are stabilized by intrachain hydrogen bonds and hydrophobicity. For general amyloid proteins, the intermediate aggregates are partly composed of β-strands in the hairpin state and/or ones that are extended. The oligomers tend to have a higher percentage of β-strands in the hairpin state than the fibrils. The book *Molecular Modeling at the Atomic Scale* by Zhou[26] explains that the hairpin tends to serve as the growing end of the sheet. However, hairpins at the edges of oligomers tend to impede other proteins joining the aggregate. These hairpins needs to straighten out before bonding to other proteins. Zhou explains that hairpins can straighten out becaue of electrostatic forces.[26] Many have suggested that key nucleation steps in the amyloid fibril formation process involve β-hairpin conformations.

1.5 Temperature

Increases in temperature have the effect of speeding up the aggregation process.[27] Hairpins can also straighten out when exposed to high temperature and not interacting with other proteins. High temperature causes the protein not interacting with other proteins to be entropy-driven and remain highly flexible in random-coil configurations, rather than trying to lower its energy by intrachain hydrogen bonds. If the temperature is high enough, the aggregation process may not involve any α-helices and instead just β-strands and random coils.

It has been suggested that most proteins have segments capable of forming amyloid fibrils if flexible enough, but most proteins have evolved to effectively conceal these segments.[1] Although only certain proteins undergo aggregation in vivo, almost

any globular protein can be made to form amyloid aggregates by denaturing it with heat or chemicals.[1]

1.6 Heterogeneity

It is hard to overstate the heterogeneity of amyloid precursors. Simulation of two short Aβ(16-22) peptides by Santini et al.[18] showed significant heterogeneity in forming a dimer. The heterogeneity of much larger amyloid forming systems is likely to have significantly more heterogeneity than the system of Aβ(16-22) peptides by Santini et al.[18] Amyloid precursors can potentially assume many conformations that are structurally distinct.[11, 20, 28,29] No structure element common to all oligomers has been found. Some aggregates have parallel β-sheets and some have antiparallel β-sheets.[30] Some aggregates have both parallel and antiparallel β-sheets. Oligomers tend to have different numbers of proteins.

Another cause of heterogeneity is that for an aggregating system, many of the proteins may be undergoing secondary structure conversions.[11] Although α-helices are not a main constituent of the ordered fibril structure, there can be α-helices in the intermediate aggregates. There is also a significant amount of heterogeneity in the fibrils. For instance, fibrils can be composed of both anti-parallel and parallel β-sheets. They also can have different β-sheet registries. [31]

1.7 Work Performed on Amyloid-Like Fibrillization

1.7.1 Mathematical Modeling of Real Aβ Proteins to Gain Insight into Various Steps

An extensively studied, naturally occurring peptide system is the amyloid beta precursor Aβ peptide, which is responsible for amyloid deposits in AD. Naturally occurring Aβ peptides have a length in the range of 39–43 amino acids. Experiments have elucidated valuable structural information of variants such as Aβ(10-35), Aβ (1-40) and Aβ (1-42) peptides. Fibrillar structures are formed when Aβ acquires a β secondary structure. Though the molecular mechanisms of fibrillation are still uncertain, it is known that Aβ fibril formation is a multi-step nucleated polymerization process that involves soluble oligomeric intermediates called protofibrils.[30]

Many different theoretical models have been created for the rates of the processes of aggregation.[32,33,34,35] These models, however, are unable to explicitly deal with the fact that amyloid precursors are heterogeneous.

Kinetic models that include reaction pathways and rate parameters for the various stages of the process can be helpful toward understanding the dynamics on a molecular level. Rates for peptide structural transformations that nucleate β-structure, as well as rates for aggregation/polymerization into fibrils are especially interesting. There are experimental data on rate parameters for some of the stages of the fibril formation process. The stages that have experimental data can have protein specifics be accurately measured. These include the investigations on the elongation and nucleation

rate parameters for fibrillogenesis at low pH, rates of nuclei production and association, and lag time and rate of aggregation. Experiments have shed light on aspects of the process of Aβ aggregation from random coil to fibril. Kirkitadze et al.[28] were able to measure the temporal changes of concentrations of different types of secondary structures: random coil, α-helix, β-strand/sheet, and β-turn for both Aβ(1-40) and Aβ(1-42). These measurements were done over a period of 20 days for Aβ(1-40) and 10 days for Aβ(1-42), allowing enough time for fibrillogenesis to occur. Fezoui and Teplow[36] measured the concentration of prefibrillar and fibrillar β-structure as a function of time in different solvent conditions. Walsh et al. [37] measured the concentration of various secondary structures as a function of time during the aggregation process.

On the basis of the experimental data, I have developed a mathematical model to determine viable reaction pathways and rate parameters for peptide secondary structural conversion and aggregation during the entire fibrillogenesis process from random coil to fibrils, including the species that accelerates the conversions. The numerical solutions yield graphs of concentrations of different molecular species versus time. With the proper choice of transition rate parameters, my model is able to nicely fit the time dependence of the concentrations measured experimentally by Kirkitadze et al.,[28] Walsh et al.,[37] and Fezoui and Teplow.[36] The model and the rate constants include different molecular structural stages in the nucleation and polymerization process and provide insight on the molecular species involved in the fibril formation process.

Determination of rate parameters for the various peptide structural conversion steps during fibril formation may be helpful in developing efficacious therapies for AD.

1.7.2 Ccβ Molecular Dynamics Computer Simulations to Uncover Molecular Level Details

Molecular dynamics (MD) computer simulations provide insight into the molecular details of various steps, such as the rearrangement of hydrogen bonds. I simulated a system with 12 ccβ proteins. [38,39,40,41,42,43] ccβ is a de novo protein that is naturally in coiled-coil trimers at low temperatures but transitions to amyloid fibrils at high temperatures. I used Replica Exchange Molecular Dynamics (REMD) to model the aggregation process of ccβ. REMD makes it so that the system being simulated climbs over energy barriers in much shorter times than in constant temperature MD. The temperatures I used were in a range of 380 - 600K.

I observed ccβ developing protofilaments with extended β-strands, corroborating what was found by Strodel et al.[40] From these MD computational studies, I was able to investigate changes in the hydrogen bond arrangements and calculate thermodynamic properties that characterize the temperature dependence of the structural transitions.

2. METHODS

The association of fibrillar Aβ with AD has stimulated interest in the kinetics and dynamics of Aβ fibril formation. Fibrillar Aβ is a fibril-like structure composed of many Aβ proteins. The Aβ proteins have 39 to 43 amino acids. They are produced by cleavage from a larger precursor protein, APP.

I used two different theoretical and computational approaches for studying fibrillization. In order to investigate the kinetics of the different oligomerization stages, I developed a system of coupled, non-linear differential equations describing the time evolution of the population of the different oligomer species. My model incorporated the heterogeneity of the oligomers. To study the molecular dynamics, I used a computational approach. It is very difficult to gain from experiments a detailed description of small oligomer formation, because of the structural heterogeneity and transient nature of the small oligomers. In order to investigate the molecular details of the fibrillation process, I used MD computer simulations.

2.1 Mathematical Modeling of Oligomer Population Kinetics

The important issues that must be incorporated into a mathematical model of the kinetics of the fibrillization process are the many different stages of oligomerization, as well as the heterogeneity within some of the stages. Compartmental modeling is an important concept for the development of a good model.

2.1.1 Introduction to Compartmental Modeling

Many people are familiar with compartmental analysis from radioactive decay. The R compartment is the number of radioactive atoms that have not yet decayed. The D compartment is the number of atoms that have decayed. The system is homogenous, with each radioactive atom having an identical rate constant k of decay. Figure 2.1 shows the radioactive decay scheme.

$$R \xrightarrow{k} D$$
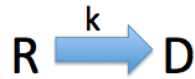
Fig. 2.1 Radioactive decay schematic.

The rate constant k has the units of 1/time (sec^{-1}) and $1/k \equiv \tau$ gives the characteristic time for decay of R. If the system is homogenous, the D compartment follows a simple exponential with time, eq. (2.4). Assuming at time t=0, $[R]=[R_0]$ and $[D]=0$

$$\frac{d[R]}{dt} = -k[R] \qquad (2.1)$$

$$\rightarrow \frac{d[R']}{[R']} = -kdt' \rightarrow \int_{[R_0]}^{[R]} \frac{d[R']}{[R']} = -k \int_0^t dt' \rightarrow ln\frac{[R]}{[R_0]} = -kt$$

$$\rightarrow [R] = [R_0]e^{-kt} \qquad (2.2)$$

$$\frac{d[D]}{dt} = k[R] = k[R_0]e^{-kt} \qquad (2.3)$$

$$\rightarrow \int_{[D']=0}^{[D]} d[D'] = k[R_0]\int_{t'=0}^t e^{-kt'}dt' \rightarrow [D] = -[R_0](e^{-kt} - 1)$$

$$\rightarrow [D] = [R_0](1 - e^{-kt}) \qquad (2.4)$$

The starting equation (2.1) has a simple form because each radioactive atom had the same rate constant of decay, k. This led to a simple exponential function, equation (2.4) with respect to time for the growth of compartment D.

Now, let us consider a transition process in general, as shown in Figure 2.2.
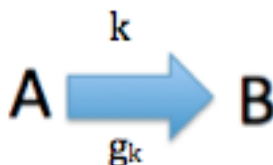
Fig. 2.2 Generalized schematic for transitions between two species. Species A may have heterogeneity g_k in the rate constant for transitions to B.

Using a simple equation like (2.1), $\frac{d[A]}{dt} = -k[A]$ would give a simple exponential for the population of compartment B, similar to equation (2.4) for radioactive decay. However, if each entity in A has a different rate constant, equation (2.1) and the resulting equation (2.4) for the time-dependence of the population are no longer valid. The heterogeneity in k can be expressed as

$$\frac{d[A]}{dt} = -\int_k g_k k[A]dk \tag{2.5}$$

where g_k is the fraction of A molecules with a specific value for the rate constant k, and the integral is over all possible values of k. If we substitute equation (2.5) for (2.1) above, the derivation leading to (2.4) becomes much more complicated than if we had used (2.1). Analytical solutions are not possible except for very special probability distributions of g_k. Especially important for the present work and for most systems, the distribution function g_k of the molecular components is usually not known, and therefore it is not possible to include g_k in the analysis. Fortunately, there is a way to

create a mathematical model that allows the investigation of the kinetics of the different species that are known to transition amongst each other.

Mathematical modeling of heterogeneous kinetics can be performed without knowledge of g_k if we substitute for the heterogeneity g_k in the rate constant of one step by inserting several fictitious steps using fictitious sub-compartments. The population of each fictitious sub-compartment is treated as being homogeneous and therefore each transition from one compartment to the next can be described by a single rate constant, as shown in Figure 2.3.
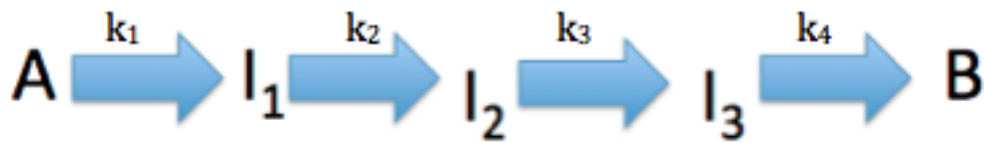
Fig. 2.3 Fictitious, homogeneous sub-compartment species inserted to mimic heterogeneity in species *[A]*.

The entities in *[A]* are not really transforming into those intermediate compartments *[I$_1$]*, *[I$_2$]*, *[I$_3$]* before arriving in *B*. However, the *ansatz* in the mathematical modeling that the entities in *[A]* go through those fictitious intermediate compartments leads to sigmoidal time kinetics for *[B]*. This type of sigmoidal curve for *[B]* mimics the dynamics of the real system that has a significant amount of rate constant heterogeneity in species *[A]*.
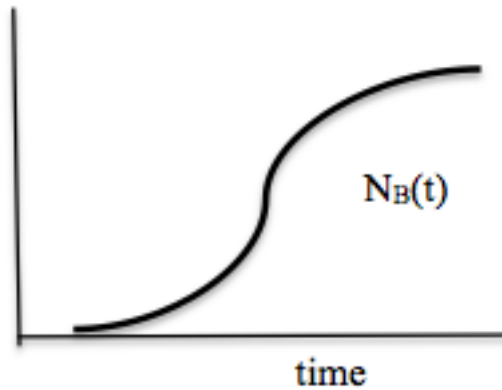
Fig. 2.4 Schematic of sigmoidal growth of the population of species *[B]*.

To illustrate why these fictitious intermediate compartments produce a sigmoidal shape, I will use a simple case, with only one sub-compartment. Let us suppose that at t=0, 100% of the entities are in compartment *A*, as shown in Fig. 2.5.
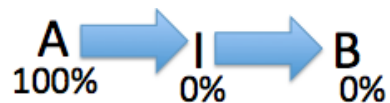
Fig. 2.5 Initial, t=0 populations for a system with one fictitious intermediate species.

Entities in *A* need to go through the intermediate compartment *I* before the process of conversion from *[I]* → *[B]* commences. Because there is initially no population in *[I]*, the process of conversion to *[B]* starts off slowly, with a lag time before the population of *[B]* starts to increase.

Now, look at the situation at some intermediate time, after many transformations from *[A]* to *[I]* and transformations from *[I]* to *[B]* have occurred, in Figure 2.6.

A 10%   I 40%   B 50%
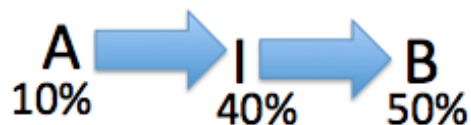
Fig. 2.6 Populations for a system with one fictitious intermediate species (Fig. 2.4) after enough time has elapsed for significant transitions to occur.

Now that there is significant population in *[I]*, the process of transformation to

[B] is much quicker.

Next, look at the situation at long times, when most of the entities are in

compartment *[B]*:

A 1%   I 10%   B 89%
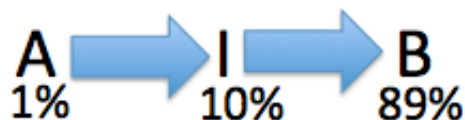
Fig. 2.7 Populations for a system with one fictitious intermediate species (Fig. 2.4) at long times when most of the particles have completed the transitions to *B*.

At long times, there are not many entities in *[I]* left to transform to *[B]*. The

population of *[B]* slows in growth rate as the system approaches its asymptotic

population distribution. The greater number of intermediate compartments between *[A]*

and *[B]*, the more sigmoidal and less exponential *[B]* will be.

So far, I have described only the behavior of *[B]* from Fig. 2.2, and how Fig. 2.3

can be used to get numerical results for *[B(t)]* when g_k in Fig. 2.2 is not known. Since

experimental measurements may be available for other species, e.g. *[A]*, it is worth

looking at their time development. In Figures 2.2, 2.3, and 2.5, the reactions occur only

from left to right. This means that eventually, *[A]* will disappear and *[I]* will disappear,

but with different time profiles. The species *[A]* experiences a simple first order

reaction that depends only on k_1, and therefore *[A]* decays exponentially, Fig. 2.8. In Fig. 2.5 with a single intermediate species, fictitous or real, the initial population of *[I]* is zero. It then rises (Fig. 2.9) by gaining population from *[A]* with a rate constant k_1, and then falls by converting to *[B]* with a rate constant k_2. The relative values of k_1 and k_2 determine the details of the shape of the time dependence of *[I]*, which is unlikely to be symmetric.
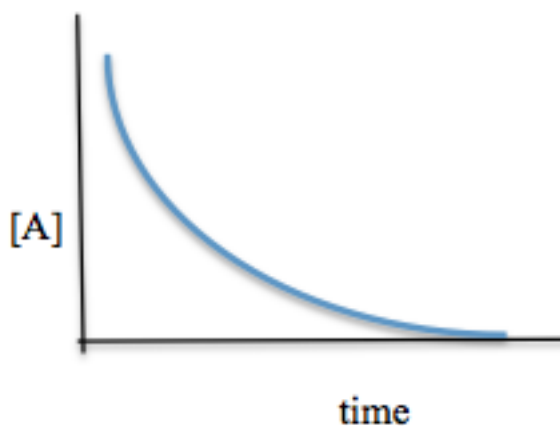
Fig. 2.8 Exponential decay of *[A]* in Fig. 2.5.
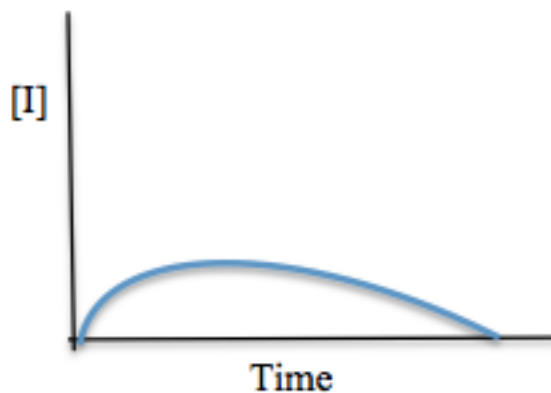
Fig. 2.9 Rise and fall of *[I]* in Fig. 2.5.

The model that I described above and showed schematically in Fig. 2.3 through Fig. 2.9 is overly simple. The direction of transitions is only from left to right. Therefore, at long enough times, the only species that remains is the last one. Real systems are usually more complicated. Reactions can go in both directions, and there is also the possibility for alternate paths that skip steps, such as going directly from *A* to *B*. Also, there may be back reactions that provide feedback, such as from *[B]→[A]*. All of these possibilities must be taken into account to produce a realistic mathematical model of the kinetics that is more complicated than the simple example given above.

2.1.2 Compartmental Model Used In My Work

The use of differential equations for studying the change in protein compartments with time is described in Chapter 4 of *Protein Folding Kinetics* by Nolting.[44] Nolting shows coupled differential equations that are simple enough to have analytic solutions. The differential equations in *Protein Folding Kinetics* did not explictily take into consideration the heterogeneity in the rate constants.

I used differential equations to study the changes in population in the following secondary structure configurations (species in the examples given above): random coil/β-turn, α-helix, β-strand/sheet. Each compartment was the percentage of amino acids from all the proteins in the system with a characteristic secondary structure. Many of the compartments also had structural characteristics in addition to the secondary structure.

On the basis of experimental data, I developed a mathematical model to determine viable reaction pathways and rate parameters for peptide secondary structural

conversion and aggregation during the entire fibrillogenesis process from random coil

to mature fibrils, including the species that accelerates the conversions. The numerical

solutions yield graphs of concentrations of different molecular species versus time. The

model and the rate constants include different molecular structural stages in the

nucleation and polymerization process and provide insight on the molecular species

involved in the fibril formation process. Determination of rate parameters for the

various peptide structural conversion steps during fibril formation may be helpful in

developing efficacious therapies for AD.

2.1.3 Numerically Solving the Equations

Finding an analytical solution to the complicated set of differential

equations that I developed (see Chapter 3) would be extremely difficult, if not

impossible. The number of equations that are coupled with each other is large,

plus many of them are non-linear. Instead, I solved them numerically.

There are different numerical algorithms for solving differential

equations. An overview of some of these techniques is given in the book

Computational physics by Newman.[45] The numerical algorithms involve

iterations over time steps.

A simple algorithm is Euler's Rule. The formula for Euler's rule with

one dependent variable is:

$$x(t + h) = x(t) + h \cdot f(x, t) \qquad (2.6)$$

with $dx/dt = f(x, t)$ and $\Delta t = h$ the size of the time step. The formula for

f(x,t) is assumed known, so that its value can be determined at each step.

Iteratively using this equation yields a solution with an error on the order of h^2.

The second order dependence is good because h is a small quantity. There is

only one equation to use for Euler's method and it's a simple equation that

comes from a Taylor expansion. However, by making things more complicated,

we can reduce the algorithmic error to a higher power of h.

The second-order Runge Kutta method uses the slope at the midpoint

between points at t and $t+h$, $t + \frac{1}{2}h$. The second-order Runge Kutta method

requires a knowledge of $x(t + \frac{1}{2}h)$, which can be approximated using Euler's

method: $x\left(t + \frac{1}{2}h\right) = x(t) + \frac{1}{2}hf(x, t)$. The equations used for second-order

Runga Kutta (rk2) for a single dependent variable are:

$$k_1 = hf(x, t) \tag{2.7}$$

$$k_2 = hf\left(x + \tfrac{1}{2}k_1, t + \tfrac{1}{2}h\right) \tag{2.8}$$

$$x(t + h) = x(t) + k_2 \tag{2.9}$$

The formula for *f(x,t)* is assumed known, so that its value can be determined at

each step. You use the first equation to get k_1. You then use the second equation

to calculate k_2. Based on knowledge of k_2, you get the value of *x(t+h)*. The error

for rk2 is of the order of h^3.

Can we do better than rk2? How about fourth order Runga Kutta (rk4)? By adding linear combinations of Taylor expansions around various points, we can cancel terms in h^3, h^4 and so on. The equations for rk4 for a single dependent variable are:

$$k_1 = hf(x, t) \tag{2.10}$$

$$k_2 = hf(x + \tfrac{1}{2}k_1, t + \tfrac{1}{2}h) \tag{2.11}$$

$$k_3 = hf\left(x + \tfrac{1}{2}k_2, t + \tfrac{1}{2}h\right) \tag{2.12}$$

$$k_4 = hf(x + k_3, t + h) \tag{2.13}$$

$$x(t + h) = x(t) + \tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{2.14}$$

For each iteration, you start with the first equation and do the rest of the equations the order in which they are written above. There are two more equations in rk4 than in rk2, which requires additional computational time, but the benefit is that it has an algorithmic error on the order of h^5, which means larger values of $h \equiv \Delta t$ can be used. With larger values of h, the overall number of time steps can be reduced. Many people feel rk4 offers the best balance between number of calculations needed for each time-step and algorithm accuracy. It is by far the most commonly used method. For a system of coupled differential equations with more than one dependent variable, like my system, the equations are iteratively integrated for each dependent variable in parallel with the other dependent variables.

Although in general rk4 is the best method for solving differential equations, it depends on the particular situation. Remember that algorithmic errors are not the only errors. There are also round-off errors. Round-off errors come from the fact that a computer does not use an infinite number of digits after the decimal point. Round-off errors may seem trivial, but they can accumulate to produce large errors.[46] Round-off errors are especially prevalent in the above-mentioned techniques because they involve getting slopes numerically, which involves subtracting two numbers that are close to each other. Subtracting two numbers that are close to each other can produce serious round-off errors.[46] The idea is to have a good balance between algorithmic accuracy and the number of steps in the calculations. This is accomplished through the proper selection of the algorithm and the size of the time step. There is even the idea of varying the step size while doing the rk4 integration.[46]

The differential equations representing the reaction pathway for my model were solved numerically using the computer software package Mathematica 7.0. Mathematica analyzes the differential equations and then determines which algorithm to use. Mathematica sometimes even switches after already starting to solve the equations if a better choice is found.

2.2 Molecular Dynamics of Amyloid Formation

2.2.1 ccβ Peptide

In order to perform detailed MD computational investigations, I focused my attention on a small, engineered peptide, ccβ (PDB ID code 1S9Z).[38,39,40,41,42,43] ccβ, which has amino acid sequence SIRELEARIRELELRIG, responds to temperature changes by undergoing many of the structural transitions that occur in the amyloid formation processes of the larger, naturally occurring amyloid forming proteins and it has been demonstrated experimentally that coiled-coil trimers of ccβ can transform their structure and form amyloid-like protofibrils.[42] Experimental observations[42] find that at a temperature of 277K, ccβ exists as a stable coiled-coil α-helix trimer, whereas at 310K the helical structure converts to β-strands. MD investigations on a single trimer by Strodel et al.[40] found that a β-sheet composed of three ccβ chains is stable up to a temperature of 440K.

There are two important advantages to using ccβ in investigations of amyloid formation: its small size, and the fact that the control parameter for inducing structural transitions is temperature. Strodel et al.[40] performed MD simulations of a trimer of ccβ. In order to speed up the process, they used replica exchange MD (REMD) simulations with the EEF1 force field. They found many pathways for the transforming of the coiled-coil ccβ trimer into a trimer β-sheet. They also found that only parts of the ccβ peptides first convert to β strands, which then further catalyzes the α to β secondary structure conversion.[40]

2.2.2 Molecular Dynamics Computer Simulations

Molecular dynamics is a computational technique that simulates the dynamics of a system of atoms as a function of time. On the most fundamental level, the evolution of the system is described by quantum mechanics. However, using quantum mechanics would be computationally very expensive. Instead, the dynamics of the atoms are handled classically. Although the energy functions that are used are classical, the parameters (such as the spring constants representing bonds) come from experiments or from results of *ab initio* quantum mechanics calculations involving electronic orbitals. The use of classical dynamics with quantum mechanically determined parameters makes it so that the calculations are quick enough to model many-atom systems such as proteins, but accurate enough to produce simulations that reasonably agree with experiment. Figure 2.10 is a schematic of the basics of the MD algorithm.

Basic MD Algorithm

Set initial conditions $r_i(t_0)$ and $v_i(t_0)$

Get new forces $F_i(r_i)$

Solve the equations of motion numerically over a short step Δt
$$r_i(t) \rightarrow r_i(t + \Delta t)$$
$$v_i(t) \rightarrow v_i(t + \Delta t)$$

$t = t + \Delta t$

Is $t > t_{max}$?

Calculate results and finish
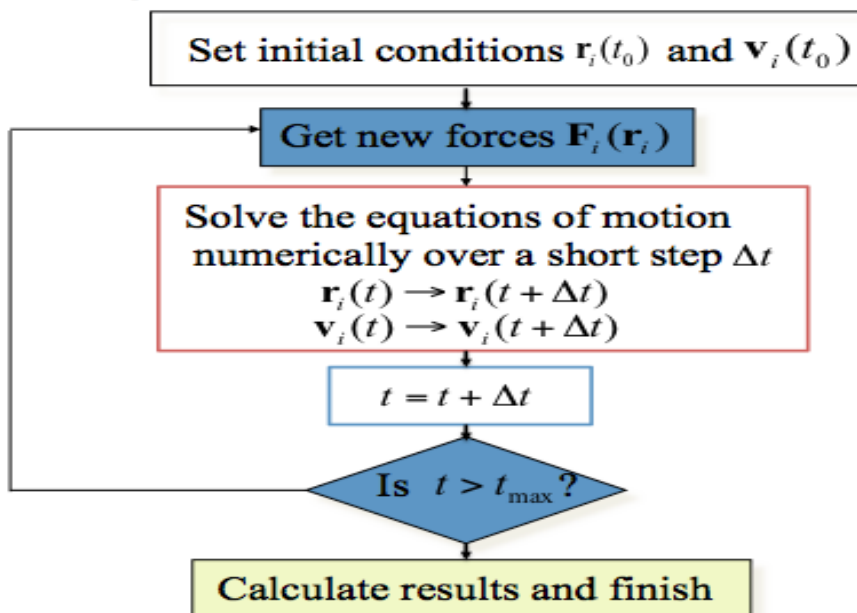
Figure 2.10 Schematic of the MD algorithm.

There are different packages that perform MD, such as NAMD, Amber, and

CHARMM. I used CHARMM (Chemistry at Harvard Molecular Mechanics) for my

MD simulations. CHARMM is a versatile MD simulations package.

2.2.3 Details of Molecular Dynamics

The MD calculations compute the forces and energies based on the

positions of all the atoms in the system. The interactions among atoms are

classified into bonded and non-bonded.

2.2.3.1 Bonded Interactions

Bond Stretching

A chemical bond is modeled as a spring with a classical form of the

potential energy (coming from Hooke's Law), but with the value of the bond

spring constant k_b derived from experiments or *ab initio* quantum mechanics

calculations:

$$U(r) = \frac{k_b}{2}(r - r_0)^2 \qquad\qquad (2.15)$$

with r_0 the equilibrium bond length.

Angle Bending

The angle bending energy is the energy needed to bend a bond from its equilibrium value, θ_0:

$$U(\theta) = \frac{k_\theta}{2}(\theta - \theta_0)^2 \qquad (2.16)$$

in which k_θ is the force constant whose value is derived from experiment or *ab initio* quantum calculations. The form of the energy is that of a simple harmonic oscillator, which is valid for small θ.

Torsions

If four atoms i,j,k,and l form a bonded quartet (i bonded to j, j bonded to k, and k bonded to l), a dihedral angle can be defined. The dihedral angle for a sequence of atoms i-j-k-l is the angle between the normal to the plane defined by the atoms i-j-k and the normal to the plane defined by the atoms j-k-l.[47] The terms dihedral angle and torsion angle are often used interchangeably. The energy for the torsion angle φ is:

$$U(\varphi) = \sum_{n=1}^{N} \frac{k_\varphi}{2}[1 + \cos(n\varphi - \delta)] \qquad (2.17)$$

In the above equation, k_φ is a force constant (derived from *ab initio* quantum calculations or experiments). δ is the phase factor that puts the minimum potential energy at the correct torsion angles.

The idea is that n is the periodicity of the dihedral angle, or the number of cycles for a full 360^0 rotation.[48] For instance, N=2 means there is one potential energy function which has one minima and another energy function which has two minima, each 180^0 apart. N is usually only one, although it can go up to six.[49]

Improper torsions are used, when needed, to enforce the plane structure of a molecule. If instead of atoms i,j,k,l bonded in a sequence, you have a star-like conformation with k bonded to atoms i,j,l you use an improper torsion angle. Atom k is called a branching point. Improper torsion energy can be used to obtain a particular geometry. The energy of an improper torsion angle is:

$$U(\omega) = \tfrac{1}{2}k_\omega(\omega - \omega_0)^2 \qquad (2.18)$$

with ω being the angle between the normal to the plane i,j,k and the normal to the plane j,k,l.

2.2.3.2 Non-Bonded Interactions

Most of the non-bonded interactions are the electrostatic and van der Waals interactions. Non-bonded interactions are neglected between covalently-bonded atom pairs and atom pairs that are separated by only two covalent bonds[49] because the covalent bonds incorporate the non-bonded interactions.

Electrostatic Interactions

These interactions are the attractive and repulsive forces between charged atoms and molecules. The electrostatic energy function comes from Coulomb's law:

$$U(r) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\varepsilon r_{ij}} \qquad (2.19)$$

In the equation above, N_A is the number of point charges in molecule A and N_B is the number of point charges in molecule B.

There are also electrostatic interactions within a molecule:

$$U(r) = \sum \frac{q_i q_j}{4\pi\varepsilon r_{ij}} \qquad (2.20)$$

where the sum is over all pairs of atoms in the molecule besides nearest and next-nearest pairs. The permanent partial atomic charges that are assigned at the beginning of an MD simulation are determined by making an initial guess and then performing calculations in an attempt to get the results of quantum mechanics calculations or experiments. The assigned partial charges are modified based on the discrepancy between the result of the initial guess and results from experiments or quantum mechanics calculations. The process is repeated until there is good agreement between assigned charges and experiments or quantum mechanics calculations.

Lennard-Jones Interactions

The van der Waals interaction, which is the total energy of interaction between two non-bonded atoms, is approximated by the Lennard-Jones (LJ) potential[50]:

$$U(r) = E_0[(\frac{r_0}{r})^{12} - 2(\frac{r_0}{r})^6] \qquad (2.21)$$

E_0 is the depth at the minimum in energy and r_0 is the separation of the atoms at the minimum. The first term gives the repulsion that comes from the Pauli Exclusion Principle when the electron clouds overlap. The last term gives the attraction that comes from induced dipole moments. The LJ energy makes it so that two atoms' nuclei stay a reasonable distance (but not too far) from each other. The reason eq. (2.21) is approximate is that it assumes atoms are spherical, since the energy function is independent of direction.

2.2.3.3 Non-Bonded Interaction Cut-Off Distances

Calculating every non-bonded interaction would not be feasible computationally for protein systems, since the number of non-bonded interactions scales as the square of the number of atoms and proteins have thousands or more atoms. A common way to reduce the number of calculations for the non-bonded interactions is to use a cut-off distance.

A cutoff distance is chosen by the MD user, beyond which no interaction is calculated. Care must be used when deciding where to have the non-bonded interactions cut-off, since electrostatic energies have infinite range. This cutoff creates a discontinuity if the energy abruptly drops to zero, which can lead to unreliable results or program termination. A switching function is used to avoid any discontinuities in energy.

Three parameters are associated with the cutoff distance: ctonnb, ctofnb, and ctnb. The switching function affects the non-bonded energies beginning at ctonnb, the smallest of the three parameters. The switching function takes the interaction smoothly to zero at ctofnb. The distance for generating the non-bonded pair list is given by the value of ctnb. The non-bonded pair list is updated periodically, and saves computational time by not having to check all of the atoms at every step to see if they are within ctofnb. The value of ctnb is slightly larger than ctofnb. For my system, cuton was 16Å, ctofnb was 18Å, and cutnb was 20Å.

2.2.4 Numerical Time Integration for the Evolution of My System

Molecular dynamics makes use of $\Sigma \boldsymbol{F} = m\boldsymbol{a}$ and $\boldsymbol{F} = -\boldsymbol{\nabla} U$ to calculate the acceleration of each atom based on the bonded and non-bonded forces of the surrounding atoms for each MD time-step. The acceleration is used to update the position and velocity of each atom for the next MD time-step. A time

interval of ~1fs is required to capture the highest frequency of motions of biomolecular systems, which is the vibrational motion of the chemical bond between a heavy atom and hydrogen. However, if the SHAKE command is used to constrain the bond length between a hydrogen and heavy atom a time-step of ~2fs can be employed. SHAKE can be appllied to all the bonds and angles if selected. It uses an iterative scheme to exactly fix bond lengths.[51] In the case of constraining the bond length between hydrogen and heavy atoms, SHAKE is an iterative procedure to adjust the hydrogen atoms after each step that the heavy atoms are integrated.[48]

Some techniques that numerically update the position and velocity of each atom given the acceleration are described below. These techniques are referred to as finite-difference methods. These numerical integration techniques have much in common with the numerical techniques used in the mathematical modeling of the kinetics that is described above.

Taylor Expansion

The simple Taylor expansion technique to advance the position and velocity with time of atom n uses these two equations[48]:

$$r_{n+1} = r_n + v_n \Delta t + \frac{1}{2}\frac{F_n}{m}(\Delta t)^2 \qquad (2.22)$$

$$v_{n+1} = \frac{r_{n+1} - r_n}{2\Delta t} \qquad (2.23)$$

with Δt as the time-step and F_n represents the sum of the forces on atom n. This formulation is quite simple, but results in large errors.

Verlet Integrator

The Verlet Integrator is an improvement over the simple Taylor expansion. It is derived from using a backward Taylor Expansion and a forward Taylor expansion.[48] The Verlet integrator uses this equation for advancing the position with time

$$r_{n+1} = 2r_n - r_{n-1} + \frac{F_n}{m}(\Delta t)^2 \qquad (2.24)$$

The velocity at each point can be estimated by:

$$v_n = \frac{r_{n+1} - r_{n-1}}{2\Delta t} \qquad (2.25)$$

The Verlet integrator gives high-precision values of position (errors on the order of Δt^4) and is independent of velocity propagation.[48] However, the velocity propagation from the Verlet integrator can have large errors on the order[48] of Δt^2 and the velocity can only be computed after the position is computed. It also requires a value of the position for the n-1 time step. Sometimes, a simple Taylor expansion is used to get the position at a fictitious -1 time step to get the algorithm started from the initial structure. (The starting structure is at time step zero.)

Leap-Frog Integrator

The Leap-Frog integrator (also called Verlet Leap-Frog integrator) is a modification to the basic Verlet integrator.[48] The Leap-Frog integrator uses these two equations for advancing the system with time:

$$r_{n+1} = r_n + v_{n+1/2}\Delta t \qquad\qquad (2.26)$$

$$v_{n+1/2} = v_{n-1/2} + \frac{F_n}{m}\Delta t \qquad (2.27)$$

As can be seen from equations (2.26) and (2.27), the Leap-Frog Integrator is so called because during a time step Δt the velocty leaps ½ time step ahead of the position as it is calculated. When the position is calculated, it leaps ½ time step ahead of the velocity. The current velocity can be calculated using:

$$v_n = (v_{n+1/2} + v_{n-1/2})/2 \qquad (2.28)$$

or other methods that introduce smaller errors than equation (2.28). The position in the Leap-Frog algorithm has an error for position of Δt^4, the same as Verlet. However, the error in the velocity calculated using Leap-Frog at half time steps is Δt^3, which improves over the Verlet algorithm of Δt^2.

Simply decreasing Δt to try to increase the precision would require more time steps to be calculated. This not only increases the computer time needed, but also increases the total round-off error.[46] A rule of thumb for selecting Δt is:

$$\Delta t = {}^{\tau}/_{20} \qquad\qquad (2.29)$$

in which τ is the period of the fastest motion.[48]

2.2.5 Replica Exchange Molecular Dynamics

A complex system like a protein with many moveable and interacting parts is likely to have a complicated energy landscape with many hills and valleys. Exploring such a landscape to find preferred, low enegy configurations with MD simulations can be difficult. The initial configuration is often chosen for reasons of convenience, but because of high barriers, the system may be unable to leave the initial location in configuration space if the temperature of the simulation is low. Conversely, if the temperature of the simulation is set very high, the system can quickly change its configuration from the initial one, but may never settle into any configuration because high temperatures cannot distinguish potential energy minima. REMD[52,53,54] is used to allow the system to move around configuration space but also settle into a region that has a potential energy minimum. Because of this, REMD is known as a technique of enhanced sampling.

Replica Exchange Molecular Dynamics is performed by first creating many replicas of the initial configuration of the system and assigning each replica a different

temperature. Each replica is independently but simultaneously simulated by MD. After

a fixed number of MD time steps, a Metropolis-type test is applied to the replicas to

decide whether to swap replicas at adjacent temperatures. The temperature exchange

probability is given by $P_{ij}=min\{1,exp(-\Delta)\}$ with $\Delta=(E_j-E_i)*[1/(kT_i)-1/(kT_j)]$ and E_j the

potential energy of replica j. After this test for swapping is performed and the replicas

that fit the criteria are temperature swapped, another fixed number of MD steps is

performed on all replicas, followed by another test for swapping adjacent temperatures,

and so on. During stages when a replica is at high temperatures, it is able to escape

from configurational kinetic traps (energy minima) and sample large regions of

structural configuration space, but is unlikely to settle into any specific, low energy

configurations. However, a replica that finds a low energy configuration is likely to

swap to a lower temperature and thus increase its probability to remain in a low energy

configuration.

2.2.6 Implicit Solvation

The natural environment for a protein system is water. Explicitly including

water atoms in addition to protein atoms is computationally expensive. Ninety

percent or more of the computation time may be spent on water-water

interactions, which are not directly relevant to the protein dynamics. In a protein

system, often the water-water dynamics are not of major importance. To speed

up the calculations, implicit solvents can been used in MD simulations. Chapter

7 in *Computational Biochemistry and Biophysics*[48] gives general descriptions of

different ways to represent the water solvent implicitly, rather than using explicit water molecules.

One such implicit solvation model is EEF1.[55,40,56] It uses an excluded volume implicit solvation model and the CHARMM 19 polar hydrogen energy.[49] It also implicitly models the dielectric screeing of water molecules by a distance-dependent dielectric function, $\varepsilon = r$,[57] as well as neutralizing ionic sidechains.[57,49] It also uses the solvent exclusion effect, which assumes that the solvation free energy of each group is equal to a reference solvation free energy of that group minus the solvation it loses due to solvent exclusion by other atoms around it.[57]

Strodel et al. performed REMD and basin hopping global optimization to study the KFFE monomer and dimer energy landscapes, using four implicit solvent models.[40] They found that of the four models they tested (EEF1 and three generalized Born models) EEF1 provided the best description for KFFE when comparing to experimental results.

In order to model solvent friction and Brownian motion of the solvent, the Langevin dynamics method is often used with an implicit solvation model. The Langevin equation is:

$$m_i \frac{d^2 r_i}{dt^2} = -\nabla_i U(r) - m_i \beta v_i(t) + R_i(r) \qquad (2.30)$$

β is the reduced friction coefficient. $R_i(t)$ is the random force representing stochastic collisions between solvent molecules and solute.

2.2.7 Confining Potentials and Periodic Boundary Conditions

Periodic Boundary Conditions (PBC) can be used in MD simulations to decrease the volume of the system that is included in the simulations and thus allowing the simulations to run faster.[49] A boundary is created in the shape of a cube. When an atom exits on the left side of the boundary, it enters on the right side of the boundary. The same thing occurs for the top and bottom edges of the PBC cube, as well as the front and back edges of the cube.

Another way to confine the system is by using a repulsive spherical potential to push the atoms back once they have encountered the confining potential.

2.3 My Particular Molecular Dynamics Simulation

To speed up the structural conversion process so that I could investigate the β formation and aggregation dynamics, I performed an REMD simulation using the CHARMM EEF1.1 force field with the MMTSB toolset,[58] using CHARMM version 35b2. EEF1.1 is the same as EEF1 mentioned above, except that EEF1.1 contains some adjustments in the parameters describing partial charges. The adjustments are empirical adjustments and are made because some of the parameters in the EEF1 force field model simulate the ionizable side chains too strongly.

I used an MD step of 2fs and I had CHARMM write every 500 steps (making each frame 1ps apart). My REMD simulation checked every 500 MD steps (every 1ps) for whether to swap temperatures between replicas of adjacent temperatures. I used exponential spacing of the temperatures with a range of 380K-600K.

In order to make up for the Brownian motion and frictional effects of the solvent that are not simulated in the EEF1 implicit solvent, I used Langevin Dynamics with a reduced friction coefficient of 1.

The proteins were confined by a spherical quartic potential that had the value zero for $r<DROFF$ and a value of $Force*\Delta^2*(\Delta^2-P1)$ for $r>DROFF$ with $\Delta=r-DROFF$. I used $Force=0.2$ and $P1=2.25$. A $DROFF$ of 600Å was used to minimize the effect of the confining potential on the aggregation process.

The number of interchain and intrachain hydrogen bonds were calculated using Visual Molecular Dynamics. Many of the parameters were calculated using Python programs analyzing large PDB files containing the coordinates of each atom during each frame.

3. MOLECULAR RATE CONSTANTS FOR AMYLOID β FIBRILLOGENESIS

An extensively studied, naturally occurring peptide system is the amyloid beta precursor Aβ peptide, which is responsible for amyloid deposits in AD.[59] On the basis of experimental data, I have developed a mathematical model for the reaction pathways and determined rate parameters for peptide secondary structural conversion and aggregation during the entire fibrillogenesis process from random coil to mature fibrils, including the molecular species that accelerate the conversions. The model and the rate parameters include different molecular structural stages in the nucleation and polymerization processes and the numerical solutions yield graphs of concentrations of different molecular species versus time that are in close agreement with experimental results. My model also allows for the calculation of the time-dependent increase in aggregate size.

In Section 3.1, I give a general description of Aβ Proteins. Section 3.2 describes the various structures in my model. In Section 3.3, I describe model details and the results of fitting my model to experimental data.

3.1 Description of Aβ Proteins

A protein that is produced normally in the human body throughout life is Aβ. It is not known what function the Aβ protein in humans has, if any. It is cleaved from a larger precursor protein, AβPP. Aβ amyloids in the neuropil are associated with AD.

The primary structure of Aβ is:

DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA.

Naturally occurring Aβ peptides have a length in the range of 39–43 amino acids. Experiments have elucidated valuable structural information of variants such as Aβ(10-35), Aβ(1-40) and Aβ(1-42) peptides.[60,61,62] The predominant forms of Aβ produced in vivo are Aβ(1-40) and Aβ(1-42).

Different segments of Aβ have different properties. For instance, Aβ(1-40) has 48% hydrophobic residues and 40% hydrophilic residues whereas Aβ(16-22) has 71% hydrophobic residues and 29% hydrophilic residues. Different segments within Aβ(1-40) have different secondary structure propensities.[63]

3.2 Description of Various Structures in My Model

After dissolution, Aβ proteins tend to exist as monomers, dimers, and very small oligomers of mostly unstructured random coil and β-turn peptides, along with a small amount of structured β-strands and β-sheets. They then form micelles, which are loosely bound disordered aggregates. These micelles then break apart into nuclei. The nuclei are small and reactive to proteins and brain cells. They are β-aggregates and grow by forming with other nuclei and having monomers attach to them. When they have these monomers attach to them, they tend to catalyze the added proteins to the β-strand state. These nuclei continue to grow until they are large enough to be called protofilaments.

Protofilaments are composed of β-sheets parallel to each other. Protofilaments wrap around each other like strands of rope to make protofibrils. Protofibrils are soluble and can grow either by attaching of single monomers or attaching of all kinds of β-

aggregates. When protofibrils are large enough they become fibrils, also called mature fibrils.

The description above is a simple one. However, in real Aβ proteins, there are many different oligomers and pathways toward fibrils.[9] Assuming a simple process toward aggregation allowed me to be able to define a functional model for Aβ fibrillation.

3.3 Results

The role of fibrillar Aβ in AD has stimulated interest in the kinetics of Aβ fibril formation. Determination of rate parameters for the various peptide structural conversion steps during fibril formation may be helpful in developing efficacious therapies for AD. Various research groups have been successful in performing very difficult experiments.

Kinetic models that include reaction pathways and rate parameters for the various stages of the process can be helpful toward understanding the dynamics on a molecular level. Rates for peptide structural transformations that nucleate β-structure, as well as rates for aggregation/polymerization into mature fibrils are especially interesting. There are experimental data on rate parameters for some of the stages of the fibril formation process. These include the investigations on the elongation and nucleation rate parameters for fibrillogenesis at low pH,[2] rates of nuclei production and association,[64] and lag time and rate of aggregation.[65] Kinetic rate equations were used by Pallitto and Murphy[7] to model folding and aggregation. Kinetic models can further benefit from experiments that have shed light on aspects of the process of Aβ

aggregation from random coil to mature fibril. Kirkitadze et al.[28] were able to measure the temporal changes of concentrations of different types of secondary structures: random coil, α-helix, β-strand/sheet, and β-turn for both Aβ(1-40) and Aβ(1-42). These measurements were done over a period of 20 days for Aβ(1-40) and 10 days for Aβ(1-42), allowing enough time for fibrillogenesis to occur. Fezoui and Teplow[36] measured the concentration of prefibrillar and fibrillar β-structure as a function of time in different solvent conditions.

Determination of rate parameters for the various peptide structural conversion steps during fibril formation may be helpful in developing efficacious therapies for AD. The numerical solutions of my model yield graphs of concentrations of different molecular species versus time. With the proper choice of transition rate parameters, my model is able to nicely fit the time dependence of the concentrations measured experimentally by Kirkitadze et al.,[28] Walsh et al.,[37] and Fezoui and Teplow.[36] My model and the rate constants include different molecular structural stages in the nucleation and polymerization process and provide insight on the molecular species involved in the fibril formation process.

3.3.1 Model of Reaction Pathways and Experimental Data

On the basis of the experimental and computational results, Bitan et al.[66] hypothesized that hydrophobic residues on different helical peptides might aggregate to create a paranucleus, which can then transform to amyloid protofibrils. In addition, it has been suggested that the process of structural conversion may be accelerated by the presence of prefibrillar β-aggregates of varying sizes.[33,67] Experimental data[28,2,37]

43

suggest that the population of the α-helix species increases rapidly (and non-linearly) as the β-structure begins to appear near the end of a lag phase. The lag phase (also called lag time) is the time it takes for a significant number of nuclei to form. During the lag phase, not much structural conversion takes place, which shows that the appearance of β-species plays a role in the structural conversion. Although a detailed molecular description of the nucleation mechanism is still not fully understood, Auer et al.[68] suggest that the surface of a growing β-sheet acts as a substrate for the attachment of α-helical peptides during the aggregation process. Therefore, I assumed that the random coil→α-helix transition is affected by the presence of β-strands. Some experiments[69,70] suggest that fibrillogenesis proceeds from random coils directly to β-structure. The prefibrillar structures can serve as toxic elements that facilitate the structural conversion from random coil to α-helix, and also in the conversion from α-helix to β-structure.

Using differential equations for studying the change in protein compartments with time is described in Chapter 4 of *Protein Folding Kinetics* by Nolting.[44] Nolting shows coupled differential equations that are simple enough to have analytic solutions. The differential equations in *Protein Folding Kinetics* do not take into consideration the heterogeneity in the rate constants. The set of differential equations for my model is more complicated.

On the basis of experimental observations described in an article by Kirkitadze et al.,[28] I made the following assumptions in formulating my reaction pathway model: (i)The conformational transitions which occur during fibrillogenesis proceed as random coil → α-helix → β-strand/sheet → β-aggregate/protofibril → mature fibril, (ii)the processes of random coil → α-

44

helix, and α-helix → β-strand conversion are accelerated by the presence of

toxic β-strand aggregates, (iii) random coil/β-turn structures incorporated in the

mature fibril state retain their unstructured properties.

With these basic assumptions, I developed a kinetic, state-transition

model (Figure 3.1) in which the amino acid configurations are grouped into

various stages during the fibril formation process. I used differential equations

to study the changes in population in these secondary structure configurations:

random coil/β-turn, α-helix, β-strand/sheet. Each compartment represents the

percentage of amino acids from all the proteins in the system with a

characteristic secondary structure. Many of the compartments also had another

characterstic in addition to the secondary structure, as described below.
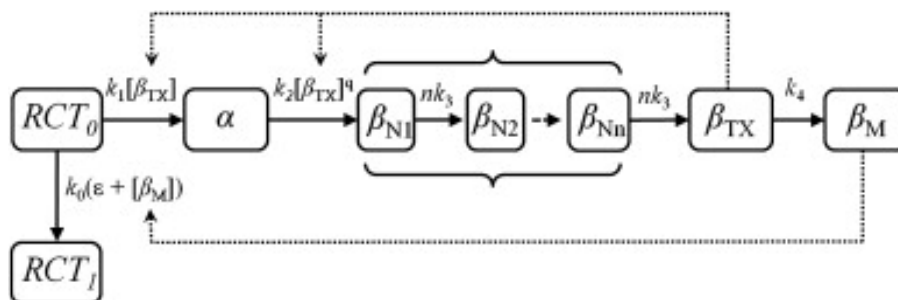
Figure 3.1 A state transition model of the structural conversion process of the
Aβ formation process. To simulate the heterogeneity in β_N structure, the β_N
phase is treated purely mathematically as a series of n fictitious sub-
compartments. For all sub-compartments, the rate of transition to the next sub-
compartment is given the same value, nk_3. The reason for this is given in detail
below.

The notation for the species and kinetic parameters included in the

schematic diagram of Figure 3.1 are explained in Table 3.1.

45

RCT_0	Initial combined concentration of random coil/β-turn
RCT_1	Fraction of inital concentration of random coil/β-turn that retain the random coil structure and are either directly incorporated into mature fibrils or remain isolated as monomers/dimers
α	α-helix concentration
β_N	Concentration of non-toxic β-strands/sheets/aggregates such as micelles
β_{Tx}	Concentration of β-strand aggregates such as protofibrils
β_M	Concentration of larger, non-toxic β-aggregates such as mature fibrils
k_0, k_1, k_2, k_3, k_4	Rate parameters for transitions from one structural state to another
ε	Fraction of initial monomers/dimers that are stable and do not participate in fibrillogenesis

Table 3.1 Parameters Used in This Model.

The differential equations used to quantify my model presented in Fig 3.1 are equations (3.1)-(3.8).

$$\frac{d[RCT_0]}{dt} = -k_0[\beta_M][RCT_0] - k_1[\beta_{TX}][RCT_0] \qquad (3.1)$$

$$\frac{d[\alpha]}{dt} = k_1[\beta_{TX}][RCT_0] - k_2[\beta_{TX}]^q[\alpha] \qquad (3.2)$$

$$\frac{d[\beta_{N1}]}{dt} = k_2[\beta_{TX}]^q[\alpha] - nk_3[\beta_{N1}] \qquad (3.3)$$

$$\frac{d[\beta_{N2}]}{dt} = nk_3[\beta_{N1}] - nk_3[\beta_{N2}] \qquad (3.4)$$

$$\vdots$$

$$\frac{d[\beta_{Nn}]}{dt} = nk_3[\beta_{N\,n-1}] - nk_3[\beta_{Nn}] \qquad (3.5)$$

$$\frac{d[\beta_{TX}]}{dt} = nk_3[\beta_{Nn}] - k_4[\beta_{TX}] \qquad (3.6)$$

$$\frac{d[\beta_M]}{dt} = k_4[\beta_{TX}] \qquad (3.7)$$

$$\frac{d[RCT_1]}{dt} = k_0[\beta_M][RCT_0] \qquad (3.8)$$

Summing the right side of all the equations (3.1)-(3.8) gives zero because my model had no proteins entering or leaving, and the differential equations are consistent with the mass action principle.

Figure 3.1 and equations (3.1)-(3.8) represent a model with the following details. An initial combined concentration of random coil/β-turn configurations is represented by RCT_0. A fraction of the RCT_0 molecules remain in the random coil configuration and are denoted as RCT_1. The RCT_1 molecules are composed of two types. One component of RCT_1 contains molecules that undergo rapid and irreversible commitment to stable monomers/dimers, as found experimentally in Pallitto and Murphy,[7] and represented by the path with rate constant $k_0\varepsilon$.

The other contribution to the population of RCT_1 is from RCT_0 amino acids that are either part of a mature fibril or sequestered in a mature fibril. The process of RCT_0 amino acids becoming either part of a mature fibril or sequestered in a mature fibril requires mature fibrils and its rate depends upon $[\beta_M]$. The experimental results in Fig. 1b of Kirkitadze et al.[28] show that the combined random coil/β-turn content levels off at long time scales, which implies that as the process of structure conversion proceeds and prefibrillar and

fibrillar structures start to form, some initial random coil/β-turn components retain their structures and may never convert to α-helix. Another study[37] found significant amounts of random coil/β-turn incorporated into both protofibril and fibril structures. Since mature fibrils are stable, the random coil/β-turn components in these structures can no longer transform to other structures.[63,71,72]

The fraction of RCT_0 that does not convert to RCT_1 follows the main pathway for fibrillogenesis in Fig. 3.1 and transforms into α-helix secondary structure configuration with a rate k_1, and this process is accelerated by the toxic β-aggregate species that are grouped in β_{TX}. The β_{TX} compartment includes toxic species of varying sizes, including small aggregates.

There are two important issues that must be explained by the model: determining a molecular mechanism that generates the lag phase observed in the experiments, and determining a reaction pathway mechanism to correctly reproduce a rapid drop-off at long times in the concentration of α-helix. I first address the issue of a molecular explanation of the lag phase.

3.3.2 Modeling of the Lag Phase

For the specific experiments that are analyzed in my work, there is little change in the concentration of any of the different species during the first four days of observation. Circular dichroism measurements revealed that the Aβ(1-40) oligomerization was preceded by an increase in α-helical content from 3% to >30%. The α-helix content begins to rise after four days of incubation

whereas the fibrils are observed only after six days. Similar results from MD simulations were obtained by Bumketner et al.,[73] in which a modest α-helical content was observed.

The sub-compartmentalization of the β_N phase is introduced to explain this lag phase. The series of β_N compartments is used as a mathematical method to represent a single phase that has heterogeneous structures with a wide range of rate constants. If a single β_N compartment is used, representing a relatively homogeneous population, there is no initial delay in change in concentrations. Mathematically, if we ignored the heterogeneity in β_N by using only a single β_N compartment with a single rate constant k_3, no lag phase occurs because β-strand/sheet structures that are included in the initial preparation as β_N are allowed to convert to toxic protofibrils (β_{TX}) with a single rate constant k_3 and the resulting β_{TX} accelerates the $RCT_0 \rightarrow$ α-helix conversion process. A single non-toxic β_N compartment is equivalent to the assumption of a homogeneous collection of β_N structures, all with the same value of k_3. However, β_N aggregates may contain a range of sizes or geometries. This structural heterogeneity[74] complicates the conversion process and produces a distribution in the transition rate from $\beta_N \rightarrow \beta_{TX}$. This structural heterogeneity is the molecular basis for explaining the lag phase and is incorporated into my model.

The effects on the kinetics of a single process due to a heterogeneous distribution in reaction rate constants can be mathematically simulated in a model by incorporating a series of 'n' fictitious steps (Fig. 3.1). The method of multiple stages, wherein a single compartment is broken into several identical

compartments in series, is widely used in epidemiological models.[75,76,77] Each sub-compartment 'i' acts as a homogeneous collection of molecules. In order for the average time for the process to remain as $\tau=1/k$, all sub-compartments have the same mean waiting time, $\tau'_i=\tau/n=1/nk$ and the same rate parameter $k'=nk$. The resulting average total time spent in all n sub-compartments is just $\tau=n\tau'_i=n(1/nk)=1/k$, the same average total time as if it was a single compartment with rate constant k. Though the average time to transition from $\beta_N \rightarrow \beta_{TX}$ is the same whether there is a single compartment with a single rate parameter k_3, or a series of n fictitious compartments (as in the model), each with rate parameter nk_3, the probability distribution of individual molecular reaction times is not the same.

The probability that a random variable is between two values is the definite integral of the probability function between these two values. For the simple process of entities transforming from compartment A to compartment B with a single rate parameter, the probability density function for molecular reaction times is exponential, as shown in Fig 2.8.

For a series of compartments in my model, the probability density distribution is described by a gamma function. The sub-compartmentalization method generates the same distribution in reaction times as expected from a heterogeneous collection of structures, each performing a single step conversion at a different rate. The probability density function for $\beta_N \rightarrow \beta_{TX}$ reaction times $p(t)$ is given by $p(t)=t^{n-1}e^{-nkt} [(nk)^n/\Gamma(n)]$, where $\Gamma(n)$ is the gamma function, and n is the number of sub-compartments. The variance in reaction rates of this

distribution is given by $1/(nk^2)$. The variance of this distribution is manifested on a macroscopic experimental level by producing a lag phase in the aggregation process. The number of sub-compartments in the model can be estimated by fitting the experimentally observed lag phase.

The gamma distribution of reaction times for the β_N structures is more in alignment with what is known about the molecular species. The β_N species undergoes aggregation as it transforms into the toxic β_{TX} species. The dynamics of aggregation are complex with many different molecular routes, and therefore it is expected that aggregation of β-strands/β-sheets into protofibrils is not a homogeneous process. In light of this, we can view the mean reaction time as a mean aggregation time. Interestingly, the lack of lag phase with the use of a single compartment, homogeneous β_N resembles the faster nucleation and polymerization processes observed for Aβ(1-42)[78,28] which displays a significantly shorter lag time as compared to Aβ(1-40). I focus on Aβ(1-40) because there are more experimental data points than for Aβ(1-42).

3.3.3 Modeling the Rapid Drop-Off at Long Times in α-Helix Concentration

Another important aspect incorporated into my model is the ability of β_{TX} to accelerate the process of α-helix→β-strand conversion in order to have the theoretical α-helix curve drop off at long times as quickly as the experimental data in the tail after the peak. In order to match the quick drop observed experimentally, the degree of this dependence had to be enhanced by

51

introducing the exponent q in the β_{TX} activity so that the rate of conversion from

α-structure is given by $k_2[\beta_{TX}]^q$. Because of the complex nature of the

aggregation process, a higher-order concentration dependence ($q>1$) is not

unusual.

3.4 Numerical Results Using My Model to Fit the Experimental Data

I developed my kinetic model to explain the experimental data on the Aβ

fibrillogenesis process from several studies. In one study, Kirkitadze et al.[28]

experimentally followed the time course of the fibril formation process and measured

the percentages of various structural elements for both Aβ(1-40) and Aβ(1-42) peptides.

I fit my theoretical curves to the experimental data for Aβ(1-40) fibrillogenesis

presented in Fig. 1b of Kirkitadze et al.[28] The Kirkitadze data are reproduced as the

points in Fig. 3.2.

In order to monitor the development of various secondary structural

motifs during fibrillogenesis, they initialized their Aβ peptide samples by

subjecting them to a dissolution process. This dissolution process produced a

starting solution of mostly unstructured random coil and β-turn peptides, along

with a small amount of structured β-strands and β-sheets. Over a period of three

weeks, the solution was subjected to circular dichroism (CD) analysis to track

the time evolution of random coil, β-turn, α-helix, and β-strand concentrations.

Accurate estimation of the β-turn content was relatively more difficult than

accurate estimation of α-helix and β-strand concentrations as a result of the β-

turn's structural variations. Two different deconvolution algorithms for spectral analyses (CDANAL[79] and CONTIN/LL[80]) yielded the same results for the concentrations of both α-helix and β-strand but yielded quite different β-turn content. However, the sum of the random coil and the β-turn concentrations remained the same using either method. Since both methods give the same results for the combined concentrations of random coil and β-turn, my model treats them together.

The rate parameters for the aggregation kinetics depend on factors such as pH, initial concentrations, temperature,[81] and methods of preparation for the initial protein sample.[28,3,81] Therefore, lag times vary greatly among experiments.[13,81,82] Lag time is highly dependent on initial concentrations among the different species as it is sensitive to various assembly states of the β-aggregates.[83]
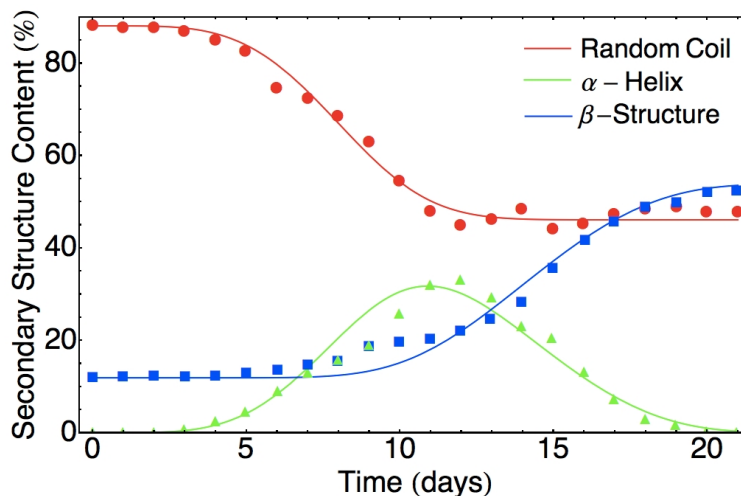
Figure 3.2 Experimental data (points) from Kirkitadze et al.[28] fit with theoretical curves derived from the model of Fig. 3.1.

53

Following the experimental conditions, I set the initial contents to be $RCT_0(t=0)$=88.1%, $\beta_{N1}(t=0)$=11.9%, and zero for all other compartments. The concentration of the monomers/dimers is difficult to measure, and quantitative information is not available.[7] Therefore I set ε=0 in my work. The parameter ε can be appropriately adjusted when experimental data are available. The model includes the reaction pathway with sub-compartments that generate a time lag that results from nucleation and the resulting curves accurately fit the data.

In Figure 3.2, I display my fit to the experimental data presented in Kirkitadze et al.[28] The following values of the rate parameters generated the curves: k_0=0.590/day, k_1=0.672/day, k_2=0.678/day, k_3=0.0392/day, and k_4=0.554/day. The number of sub-compartments is n=4, and the exponent is q=2. From Fig. 3.2, it is clear that my model accurately fits the experimental data. It successfully reproduces the lag phase kinetics involved in the fibril formation process. My model also reproduces the sharp drop in the experimental α-helix curve near the end of the process.

I further use my model to fit other experimental data. Specifically, I fit the experimental data by Walsh et al.[37] and by Fezoui and Teplow.[36] Walsh et al.[37] performed experiments with the low molecular weight Aβ(1–40) peptide and followed the conformational changes during the fibril formation process with CD spectra for over 30 days. The data points in Fig. 3.3a show the secondary structure content plotted against time as given in Table I of Walsh et al.[37] Their data are also consistent with similar observations that fibrillogenesis involves formation of intermediate α-helix secondary structure during the conformational transition from random coil to β-aggregates. The curves in Fig. 3.3a show the fit with my model. Following the

experiments, I set the initial contents to be $RCT_0(t=0)=76\%$, $\alpha(t=0)=11\%$,

$\beta_{N1}(t=0)=13\%$, and zero for all other species and compartments. The transition rates and

other parameters that generated my fitting curves were $k_0=0.0154$/day, $k_1=0.0318$/day,

$k_2=0.0282$/day, $k_3=0.0565$/day, $k_4=0.520$/day, and the number of sub-compartments

$n=6$. I note that different experimental conditions will affect the rates of different steps

in the process. For example, though the temporal trends are similar, the data from

Walsh et al.[37] display much smaller changes in concentrations than the data from

Kirkitadze et al.[28] because of differences in experimental conditions. Therefore, the

values for the rate parameters used in the fit in Figure 3.3a are different than those used

for the curves in Figure 3.2. The exponent $q=2$ remains the same as used for the curves

in Fig. 3.2.
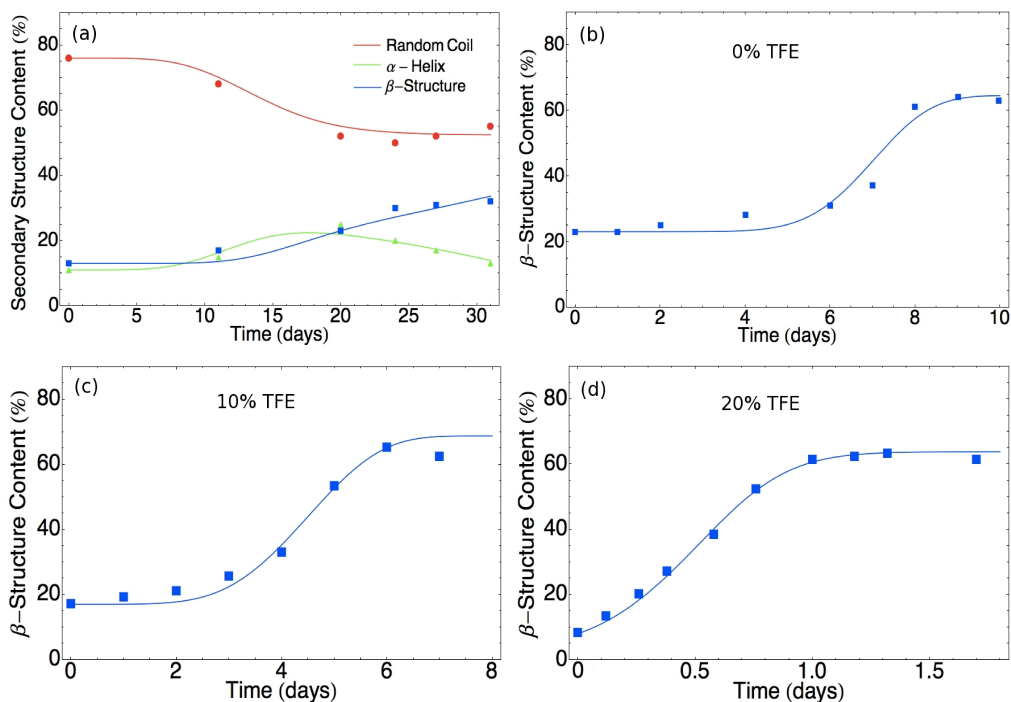
Figure 3.3 (a) The experimental data points from Walsh et al.[37] showing the secondary structure content obtained by deconvolution of CD spectra. The curves are the numerical results from fits obtained using my model. (b–d) Data points from Fezoui and Teplow[36] showing the temporal changes in the β-structure content for Aβ(1-40) fibrillogenesis at different TFE concentrations.

3.4.1 Co-Solvents

Experimentalists often use co-solvents with water, which changes the fibril formation dynamics as compared to a sytem of proteins in pure water. Urea is commonly used to denature proteins.[84] Alcohols are often used as co-solvents.[84] A good review of alcohol co-solvents is given by Buck.[85]

Trifluoroethanol (TFE) is a fluorinated alcohol. Proteins in TFE/water solution having different secondary structure dynamics than when in a pure water solvent.[86] The effects of TFE seem to depend on amino acid sequence.[85] It is well established that TFE promotes formation and stabilzation of α-helices and hairpins.[87] A theory as to why

TFE favors the formation of intrachain hydrogen bonds is that TFE molecules coat the outside of a protein.[86] The coating displaces water molecules[86] which removes hydrogen bonding partners to which backbone atoms could hydrogen bond.[86] The displacement of water also creates a lower dielectric environment which promotes electrostatic interactions.[86] Water has one of the largest dielectric constants among fluids, 78 at 25^0C.[88] Trifluoroethanol does not does not disrupt severely the hydrophobic interactions within a protein[86] or may even strengthen them, which is in contrast with most other organic solvents.

However, it is not well established whether TFE promotes or inhibits formation of β-sheets. Trifluoroethanol has been known to disrupt and also enhance quaternary interactions. The hydrophobic interactions between proteins tend to be weakened by TFE. The TFE molecule is approximately nine times larger than the water molecule.

Circular Dichroism spectropolarimetry was done on T4 LYS (11-36), a peptide dervied from a region of the T4 lysozyme molecule, by Anderson et al.[89] It was found that as the TFE content rose to 20%, the β-sheet content grew from 26 to 38%.[89] However, as the TFE content then rose to 50%, the β-sheet content decreased to 15%.[89]

In terms of the energy landscape picture of Fig. 3.4, TFE lowers the energy of the α-helix state. In addition, the height of the barriers for the RC→α-helix transition and the α-helix→β-structure transition can be affected by TFE, as evidenced by the temporal data in Fig. 2a of Fezoui and Teplow.[36]

Fezoui and Teplow[36] show that the amount of Aβ, the protein whose fibrils cause AD, goes down with increasing TFE concentration and is 0 for 30% and higher TFE concentration.[36] It also shows increasing amounts of α-helix with increasing TFE

concentrations.[36] However, the fibrils actually form faster when the TFE concentration is raised up to about 20%. Experiments have found that the random coil to fibril formation process involves significant α-helix intermediates for Aβ and that the formation of these α-helix intermediates is "on pathway" for fibril formation.[90,28] Insight into the role of the α-helix intermediate in the fibrillization process is provided by experiments[36] that use TFE which facilitates and stabilizes helical structure. At TFE concentrations below 20%, the fibril formation process is accelerated by increasing the concentration of TFE. However, the addition of higher concentrations of TFE can potentially slow fibril formation, implying that the α-helix intermediate[91] becomes too stable and does not allow the progression to the β-structures necessary for fibril formation. The dependence of fibril formation on the concentration of TFE can be explained thermodynamically with the help of an energy landscape diagram. A schematic energy landscape representing these structural stages is shown in Fig. 3.4. Increasing the TFE concentration deepens the energy minimum corresponding to the α-helix structure, which can facilitate the transition from random coil→α-helix, a necessary step in fibrillogenesis. At low TFE concentrations, fibrillogenesis continues with the transition from α-helix→β-strand. However, at high TFE concentrations, the α-helix structure becomes stabilized so strongly that the α-helix energy minimum is difficult to escape. The α-helix→β-strand conversion is impeded and fibril formation is hindered.
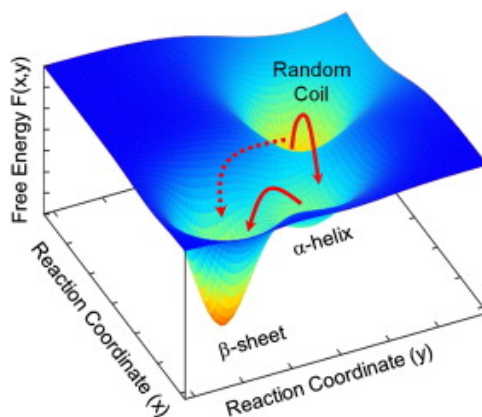
Fig. 3.4 Schematic energy landscape of Aβ secondary structure conversion as a function of arbitrary structural reaction coordinates *x* and *y*. TFE stabilization of α-helix structure is represented by increasing the depth of the α-helix energy minimum. If the α-helix minimum is too deep, the peptide may find it difficult to undergo the conversion of α-helix→β-strand. The dotted arrow represents a possible pathway in which the random coil converts directly to β-structure, bypassing α-helical structure.

Figure 3.3b-d shows the experimental data points from Fezoui and Teplow,[36] which investigated the effects of α-helix stabilization on Aβ fibrillogenesis. In that work, Fezoui and Teplow used several different concentrations of helix stabilizing TFE solution and followed the rates of fibril formation and fibril elongation. They concluded that the helical intermediate structures play a crucial role in peptide oligomerization, which is followed by the conformational reorganization into extended β-sheets. Without TFE (0% concentration), the time course of β-structure content as shown in Figure 3.3b is similar to that displayed in Figure 3.2.

Since Figure 3.3b-d only contains data for one molecular structural component, it may be possible to fit the data using different parameter sets because the fits are not constrained as strongly as for the data in Figure 3.2 and Figure 3.3a. Since the experimental conditions were significantly different in Figure 3.3b-d, I expect that the fit parameters may also be different. For example, the data in Figure 3.3b-d result from

different initial concentrations of α-helix. Differences in initial conditions, temperature, and TFE concentration lead to important differences in the time required before the β-structure content starts to dramatically increase and this will have a strong effect on the values of some of my model's parameters.

When the TFE concentration is increased up to 20%, the fibril formation process is accelerated and the lag phase is decreased. For the condition of 0% TFE in Figure 3.3b, the fit parameter values are k_0=3.00/day, k_1=0.448/day, k_2=0.310/day, k_3=0.0512/day, k_4=0.107/day, q=2, and n=4. The fit parameters for 10% TFE in Fig. 3.3c are k_0=1.23/day, k_1=0.463/day, k_2=0.350/day, k_3=0.100/day, k_4=0.733/day, q=2, and n=3. For 20% TFE in Fig. 3.3d, they are k_0=0.771/day, k_1=0.0116/day, k_2=0.467/day, k_3=0.774/day, k_4=1.00/day, q=2, and n=1. For Fig. 3.3b-d the initial content of RCT_0 and α-helix were set to the values given in Table 1 of Fezoui and Teplow.[36] The initial β content that I used was determined from the data plots in Figure 2a of Fezoui and Teplow[36] and divided as follows for the calculations in this paper: Figure 3.3b (0% TFE): $\beta_{N1}(t=0)$=18.1%, $\beta_{N2}(t=0)$=5.0%; Fig. 3.3c(10% TFE): $\beta_{N1}(t=0)$=15%, $\beta_{N2}(t=0)$=2.0%; Fig. 3.3d (20% TFE): $\beta_{N1}(t=0)$=7.0%, $\beta_{TX}(t=0)$=1.0%. All other species and compartments were assigned initial contents of 0%.

My model fits the lag phase for various TFE concentrations up to 20% by adjusting the number n of sub-compartments. The data for 0% TFE concentration have the longest lag phase (~7 days) and the largest number of sub-compartments (n=4). Similarly, 10% TFE requires n=3, and 20% TFE which has the smallest lag phase and fastest fibril formation has n=1. Increasing the TFE concentration enhances α-helix

60

stability as seen in Table 1 of Fezoui and Teplow[36] and Buck et al.[85] Interestingly, the

change in energy barrier heights causes an increase in the speed of the overall fibril

formation process up to 20% TFE, but above this TFE concentration the increased

stability of the α-helical structure acts as a kinetic trap and slows down the overall

process. In addition to the TFE concentration dependence, the lag phase also depends

on pH of the solution,[65,14,28] and my model can be adapted to these biochemical

environments as well.

3.5 Discussion of Results

It is important to note that though my results provide a good fit to a variety of

experimental data analyzed in this paper, my model may need modification for other

experimental conditions. Both the α-helix→β-structure sequence of steps as well as the

possibility of a process involving α-helix→random coil→β-sheet are discussed in the

literature.[92,93] Velez-Vega and Escobedo[94] discuss that the α-helix stage may be more

important for some peptides than for others, while Soto et al.[91] and Szabo et al.[95]

discuss the possibility that the α-helix secondary structure is not necessary for

fibrillogenesis. Additional pathways for aggregates are also proposed.[96] My model can

be appropriately modified to reflect different pathways.

Important correlations can be made between my model and the kinetic model of

fibril nucleation and elongation developed by Lomakin et al.[2] In that work, one

pathway for the fibrillogenesis mechanism for high protein concentration was described

as: Monomers ←→ Micelles→Nuclei→Fibrils. The pathway of

Monomers←→Micelles→Nuclei→Fibrils is dominant when protein concentration

exceeds a critical value above which pre-existing "seeds" are not necessary for fibril

formation. Instead, high local protein concentration leads to formation of micelles that

later form the domains of fibril nuclei. Experiments performed by Kirkitadze et al.[28]

and Lomakin et al.[2] exhibit similar lag times even though they were performed at

different peptide concentrations, 25 μM and >100 μM, respectively. The similarity may

be due to their peptide concentrations being above their respective critical micellar

concentrations at which the lag times become independent of the peptide concentration

as observed.[2,97] Other experiments[13,14,15] as well as the nucleation theoretical

treatment[98] show a more complicated dependence of aggregation rate and lag time on

peptide concentrations. The Aβ monomers and the micelles were considered to be in

rapid equilibrium so that this step was not the rate-limiting step. The two possible rate

limiting steps in the model of Lomakin et al.[2] are a micelle to nuclei formation process

with a kinetic rate parameter they denoted as k_n, and a fibril elongation process with

rate parameter k_e that incorporates free monomers that are present with concentration c.

Different nuclei sizes have been reported for different conditions. For example, Pellarin

and Calfisch[99] found the sizes of the nuclei to vary between 4 and 18 monomers

corresponding to different micellar oligomeric conditions. This number can be >30

under certain conditions, as suggested by Garai et al.[74] Here I use the nuclei size to

have 10 monomers as considered by Lomakin et al.[2]

The rate constant k_n is the number of nuclei spontaneously formed out of

micelles.[2] By fitting the experimentally observed temporal evolution of the

hydrodynamic radius of fibrils, Lomakin et al.[2] estimated the value of k_n to be

2.4×10^{-6} s^{-1} or 0.207 d^{-1}. Having k_n be 0.207 d^{-1} represents one new nucleation every

five days per micelle. My rate limiting step in Fig. 3.1, used to generate the curves in Fig. 3.2 and Fig. 3.3, involves a similar structural transformation. In order to compare with my model, I convert k_n in order to represent the number rate at which monomers contained in micelles transition to monomers contained in nuclei, rather than the number of micelles that transition to nuclei. Since each micelle on average has an aggregation number of 25 monomers, whereas there is an average number of 10 monomers in a nucleus, I make the conversion as follows: k_n=(0.207 nuclei micelles^{-1} d^{-1})×(10 monomers/nucleus)/(25 monomers/micelle)= 0.207 d^{-1}×(10/25)=0.083 d^{-1}. The comparable step in my model corresponds to amino acid residues in β_N species transforming into residues in β_{TX} species, and the rate is represented by k_3. Strikingly, the value of k_3=0.039 d^{-1} ($1.8×10^{-6}$ s^{-1}) that I determined compares nicely (factor of 2) with Lomakin's value for k_n. Also, my value for k_3 falls nicely within the range of k_n found experimentally by Wang et al.[100]

The agreement between my k_3 and the k_n is further enhanced when we take into account that my value of k_3 was determined by fitting to experimental data obtained at 22 °C, whereas experiments[2] for k_n were performed at 25 °C. Since these reactions would likely occur faster[27] at higher temperatures, we would expect my k_3 to be larger if it was determined by analyzing experiments at higher temperature, if available. This would bring my k_3 even closer to k_n.

My k_3 and the nucleation rate k_n refer to similar, but not identical processes. The observation that $k_n \geq k_3$ is consistent with my proposed fibrillogenesis mechanism. The parameter k_n refers to nucleation, whereas my k_3 encompasses both nucleation and partial aggregation. Since nuclei must form before the formation of any toxic

protofibrils, the rate of nucleation k_n can be expected to be faster than k_3. Comparing

my k_3 to Lomakin's k_n directly connects Lomakin's model using macroscopic structures

such as micelles and nuclei to my model, which deals with kinetics of microscopic

secondary structural elements. I would also like to compare my rate parameters with the

rate parameter for elongation of k_e.[2] However, comparing my rate parameters with the

rate parameter for elongation of k_e is especially complicated and requires additional

work since fibril formation in my model finishes with the step $\beta_{TX} \rightarrow \beta_M$, but also

includes a part of the $\beta_N \rightarrow \beta_{TX}$ conversion process.

My model also allows for the calculation of the time-dependent increase in

aggregate size. I now show that the calculated results agree well with experimental

results, and allow differences in experimental conditions to be included in the

calculations. Fig. 2 of Lomakin et al.[3] shows the hydrodynamic radius as a function of

time. Using formulas (1–3) from that reference, I can convert the hydrodynamic radius,

R_h, for aggregates into the number of monomers n per aggregate:

$$R_h = \frac{L\sqrt{1-x^2}}{2\ln\dfrac{1+\sqrt{1-x^2}}{x}} \qquad (3.9)$$

$$x = \frac{d}{L} + \frac{.37d(L-d)}{L^2} \qquad (3.10)$$

$$n = \lambda L \qquad (3.11)$$

where L is the length of an aggregate and λ is the number of monomers per unit length.

I set λ equal to 1.6 as used in Lomakin et al.,[3] Kusumoto et al.,[101] and Tomski and

Murphy.[102] The other parameters are: n is the number of monomers per aggregate, and

d is the diameter of an aggregate. I assume that all of the aggregates could be treated as

cylinders, as in Yong et al.[103] which says that micelles have a spherocylindrical shape. I set d=8 nm for all the aggregates[3,101] independent of L. I used Mathematica to invert equation (3.9) and find a value of L for each R_h in Lomakin et al.[3] The L was then used to calculate n for each value of R_h.

I calculate a theoretical curve from my model for the time evolution of the number of monomers per aggregate and compare with the experimental results of Lomakin et al.[3] and Kusumoto et al.[101] The two experiments were similar, but were done at different temperatures and this may have a significant effect on k_n but not k_e.[27] Though the steps in my model remain the same, differences in experimental conditions give different numerical values for the parameters in my model to fit each experiment. Differences in experimental conditions giving different numerical values for the parameters in my model is consistent with Lee[104] in which it is reported that the average length of fibrils is sensitive to pressure and temperature. I define $[\beta_{NT}]$, $[\beta_{Tx}]$, and $[\beta_M]$ as the percentages of the total number of monomers in non-toxic aggregates (micelles), toxic aggregates (nuclei), and mature aggregates, respectively. Similarly, m_{NT} and m_{Tx} are the average numbers of monomers in a micelle and nucleus, respectively. Literature supports a micelle having 25 monomers on average and a nucleus having an average of 10 monomers.[2,3,99,27,103] I calculated the maximum number of monomers per aggregate, defined as parameter m^*, which varies depending on the experimental conditions. The value of m^* is determined from the long-time, asymptotic value of the experimental data. The parameters a_{NT}, a_{Tx}, and a_M are the total numbers of non-toxic aggregates,

toxic aggregates, and mature aggregates, respectively, and N is the total number of

monomers in the entire system.

$$[\beta_{NT}] = \frac{m_{NT} \times a_{NT}}{N} \times 100\%$$ (3.12)

$$[\beta_{Tx}] = \frac{m_{Tx} \times a_{Tx}}{N} \times 100\%$$ (3.13)

$$[\beta_M] = \frac{m* \times a_M}{N} \times 100\%$$ (3.14)

$$n = \frac{[\beta_{NT}] + [\beta_{Tx}] + [\beta_M]}{\dfrac{[\beta_{NT}]}{m_{NT}} + \dfrac{[\beta_{Tx}]}{m_{Tx}} + \dfrac{[\beta_M]}{m*}}$$ (3.15)

Using eq. (3.15) I was able to obtain good fits to both sets of experimental data.

The fit in Fig. 3.5a for the data from Lomakin et al.[3] was done with $[\beta_{NT}]$ initially set at

50%, $[\beta_{Tx}]$ initially set at 50%, and $[\beta_M]=0$. The parameter $m*$ used for this fit was set to

340. This fit showed significant deviation from the data for small times. I will do

further work to achieve a better fit. A fit was produced for the Kusumoto data using

$[\beta_{NT}]$, $[\beta_{Tx}]$, and $[\beta_M]$ all initially at 33.3% and $m*$ of 600.
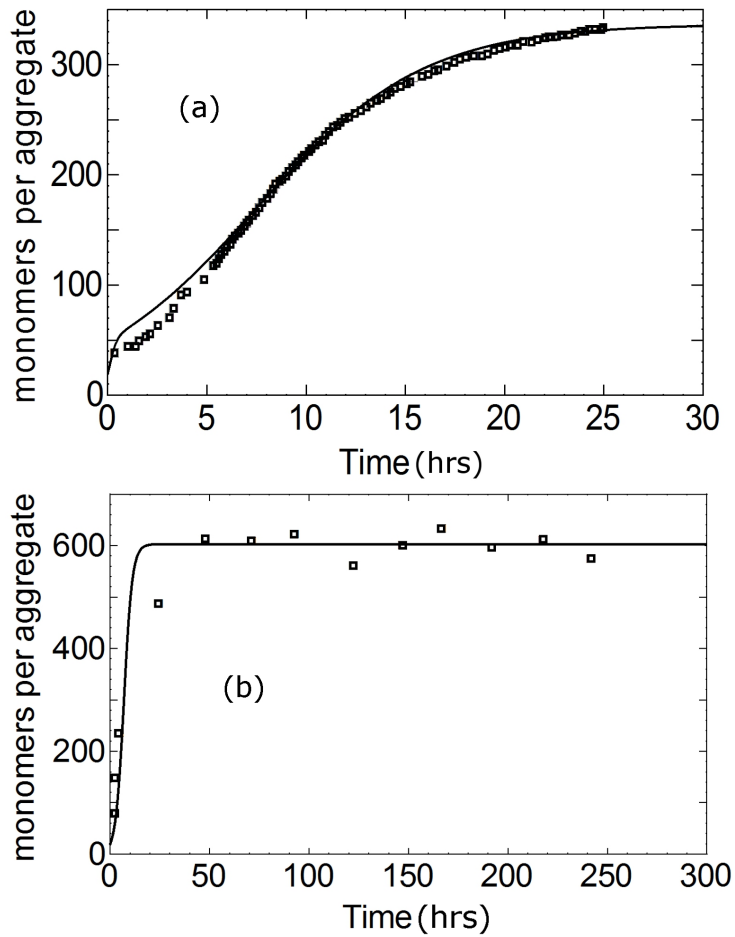
Fig. 3.5 Theoretical fits (-------) to experimental data (□) for the number of monomers *n* per aggregate for two different experiments. (a) Lomakin et al.,[3] (b) Kusumoto et al.[101]

One thing that is gained from Figure 3.5 is that for large times a state of equilibrium is reached in which the average size of aggregates is stabilized. Another thing learned from Figure 3.5 is the approximate number of proteins in a mature Aβ fibril. Since aggregates that are smaller than mature fibrils can be detected when it comes to getting the radius of hydration and some aggregates may never turn to mature fibrils, the average number of proteins in mature fibrils is likely to be larger than the asymptotic values in the two graphs in Figure 3.4. A question is why the asymptotic

value for the Figure 3.4b, about 600 proteins, is approximately twice the asymptotic value of the Figure 3.4a. Both experiments were done at similar experimental conditions and by the same group. Slight variations in experimental preparation procedures such as purification and synthesis may significantly change the peptide system's fibrillogenic properties. A study showed that Aβ fibrils are composed of five or six protofilaments,[9] whereas another study showed Aβ mature fibrils consist of two or three protofilaments.[105] Experiments done by Pellarin showed mature fibrils having four protofilaments.[106]

4. RESULTS OF MD COMPUTER SIMULATIONS FOR PROTOFIBRILLAR FORMATION OF CCβ

In this chapter I provide results of my MD computer simulations that provide molecular details of the dynamics of protofibrillar formation by the peptide ccβ. I explain the steps that were taken to remove aphysical, computational artifacts so that the results would be physically valid. I also show how the MD results reinforce the idea that molecular heterogeneity is an important factor in understanding amyloid fibrillization.

4.1 Computational Details

The initial configuration of the 12 chain system in my REMD simulation consisted of four ccβ coiled-coil α-helix trimers arranged at the corners of a square, as shown in the first panel of Figure 4.1. Each trimer was created from the coordinates in the PDB 1S9Z file. The PDB file contains the atomic coordinates for a single ccβ helix along with the translation and rotation coordinates for adding two additional ccβ peptides to form a triplet. The four trimers at the corners of a square were separated by enough distance to ensure that there would be no van der Waals overlap between atoms from different trimers, with at least 24 Å between the center of mass of one trimer and the center of mass of another trimer. The EEF1 force field with implicit solvent was used for modeling the solvation.

I performed the REMD simulation for 300 ns using 20 replicas with exponential spacing of the temperatures in a range of 380K-600K: 380.00K, 389.25K, 398.72K, 408.42K, 418.36K, 428.53K, 438.96K, 449.64K, 460.58K, 471.79K, 483.27K,

495.03K, 507.07K, 519.41K, 532.05K, 544.99K, 558.25K, 571.83K, 585.75K, 600.00K. Each MD step was 2fs and I saved the configuration every 500 steps (making each frame 1ps apart). My temperature swap test was performed every 500 MD steps. I performed 20 ps of REMD initialization and 20 ps of REMD equilibration. The average acceptance rate for temperature swapping was 24%, which is close to the ideal of 20%.[107,108]

In addition, I performed MD simulations at a constant temperature (CTMD). The parameters were the same in the CTMD as in the REMD simulations, except for the temperature. Also, in some of the CTMD simulations, there was a *DROFF* (confining potential radius) of 100Å as opposed to 600Å in the REMD simulations.

4.2 Description of β-Sheet (Nucleus) Formation

Figure 4.1 shows a sequence of frames taken at various times for a replica that formed two β-sheets, an important structural stage for the eventual fibrillization that could occur in a much larger system than I investigated. During the 300 ns covered in this simulation, this replica changed its temperature many times.

helix trimers

helix collapse

β-strand/hairpin
formation

fibril

growth/hairpin straightening

nucleation
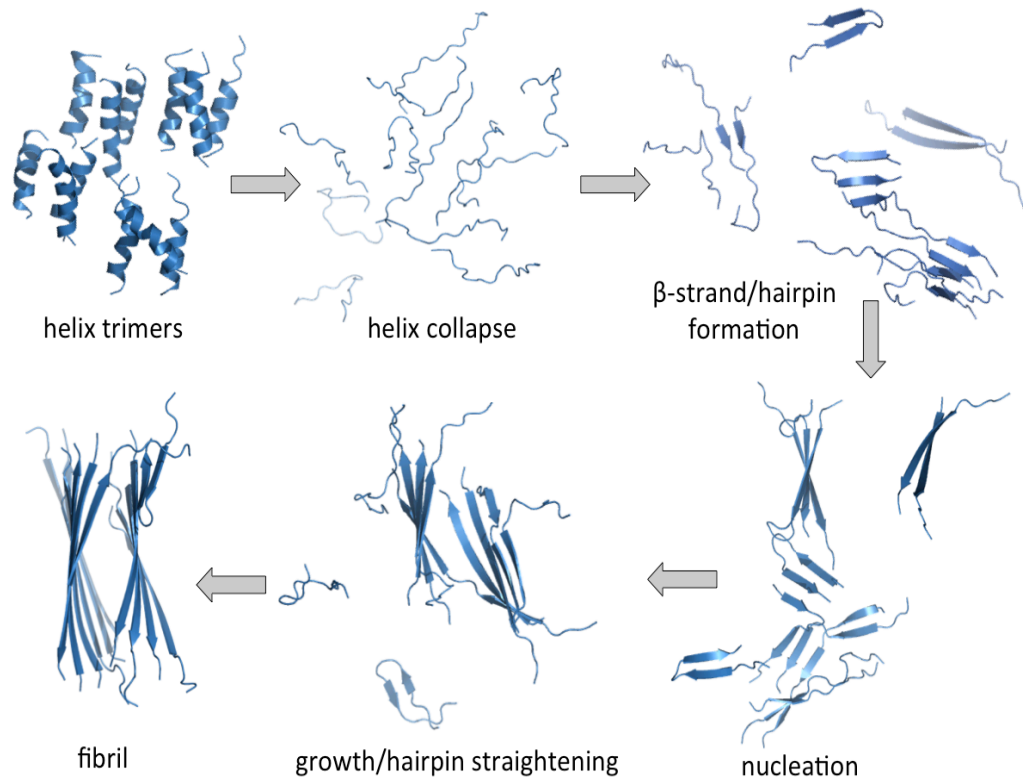
Fig. 4.1 Frames from the REMD simulations showing steps in the helix structural transition to β-strands and the formation of β-sheets.

My REMD simulation was performed in a temperature range from 380K to 600K in order to investigate β-chain formation and aggregation. These relatively high temperatures cause the initial α-helical trimers to rapidly unfold so that the system quickly progresses to the random coil/β-hairpin/β-aggregate stages that I am interested in investigating. Experimental observations on ccβ found that at 310K, the α-helical configuration is unstable.[42] Consistent with this, all of my replicas lost their initial α-helical conformations within 1 ns. I focus attention on one replica, replica 3, which created the structure that most resembles a protofilament. The process of fibrillation for replica 3 is summarized in Figure 4.1. The last frame in Fig. 4.1 (from replica 3) looks

like a protofilament steric zipper. It has similarities to the transthyretin single protein.[109] For amyloids in general, some aggregates have parallel β-sheets, some have antiparallel β-sheets,[110] and some have both parallel and antiparallel β-sheets.[111,112]

For the replicas that created a β-aggregate, after the proteins get out of the α-helix secondary structure, some of the segments of the chain are in random coil and some are in β-structure. The proteins associate with each other in a disorganized aggregate, like a micelle. They then reorganize to form the β-aggregate.

To quantify the organization of the last frame of replica 3 (t-300 ns), which is shown in Figure 4.1, I calculated the nematic order parameter S for the larger of the two β-sheets, which contains six ccβ chains. The nematic order parameter[113] is a measure of the alignment of rods in a system. I set S to be the largest eigenvalue of the matrix

$$Q_{\alpha\beta} = \frac{1}{N_{protein}} \sum_{j=1}^{N_{protein}} \left(\frac{3}{2} a_{j\alpha} a_{j\beta} - \frac{1}{2} \delta_{\alpha\beta} \right) \tag{4.1}$$

where j is an index that identifies each rod, α and β represent laboratory Cartesian coordinates, and $\vec{a_j}$ is the unit direction vector for rod j. In my ccβ system, the protein chains are not perfectly rigid rods. I chose the direction vector, $\vec{a_j}$ for each protein to be the principal axis of the C_α inertia tensor $I_{\alpha\beta} = \sum_{i=1}^{N_{atoms}} m_i \left(r_i^2 \delta_{\alpha\beta} - r_{i\alpha} r_{i\beta} \right)$ corresponding to the smallest value of the moment of inertia for that chain. This provides a direction that is an average for the C_α atoms in the chain. r_i is the distance from the center of mass of the protein to each of the middle eleven C_α atoms in the protein. I used only the middle eleven of the 17 residues because a β-strand protein,

even when attached to other proteins, often has considerable flexibility for the amino

acids near either end. Figure 4.2 shows for replica 3 the nematic order parameter as a

function of time for the proteins that make up the largest β-sheet. A nematic order

parameter of one means that there is perfect alignment of the protein direction vectors

and zero means total isotropy in the arrangement of the protein direction vectors.

Characteristic structures for the six chains are also displayed at various times during the

simulation. Since these results are from an REMD simulation, a single replica such as

replica 3 changes temperature often. The temperature of replica 3 at any time is shown

by the color of the curve. The order of colors from high temperature to low temperature

is: red→yellow→green→cyan→blue.
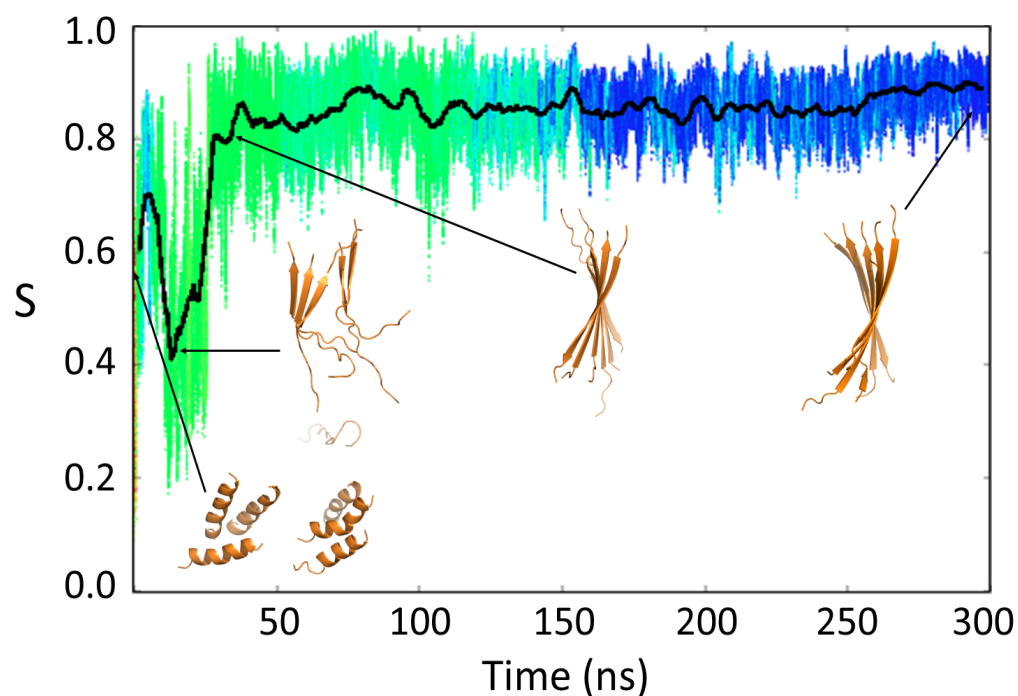
Figure 4.2 The nematic order parameter *S* for the proteins that make up the largest β-sheet in replica 3. Characteristic structures for the six chains are displayed at various times during the simulation. The temperature of replica 3 at any time is shown by the color of the curve. The order of colors from high temperature to low temperature is: red→yellow→green→cyan→blue.

Figure 4.2 displays large changes in *S* during the first 40 ns while the six β-strands undergo major rearrangements relative to each other. By 50 ns, the six chains have settled into a low energy, stable steric zipper β-sheet configuration with much smaller fluctuations in *S*. In the β-sheets formed in the simulations, each β-strand has a twist. The average twist angle for the larger of the two β-sheets for the last 200 ns is 10°. This is a similar twist angle to that calculated in Knowles et al.[114]

4.3 Phase Transition

Experimental observations[42] find that at a temperature of 277K, ccβ exists as a stable coiled-coil α-helix trimer, whereas at 310K the helical structure converts to β-strands. REMD investigations on a single trimer by Strodel et al.[40] found that a β-sheet composed of three ccβ chains is stable up to a temperature of 440K.

In order to investigate the thermodynamics of the process, I calculated various parameters from my REMD simulations. Figure 4.3 displays the heat capacity as a function of temperature for my system of 12 chains. The heat capacity at each temperature was calculated from the fluctuations in the energy using equation (4.2)

$$C_v = \frac{\overline{E^2} - \overline{E}^{\,2}}{RT^2} \qquad (4.2)$$

I found that for all temperatures, the system had settled into its final configuration, whether ordered into an aggregate or not, by 100 ns. Therefore, I calculated the

equilibrium C_V at all temperatures using the final 200 ns of the simulations, which

covers the period from 100 ns to 300 ns.
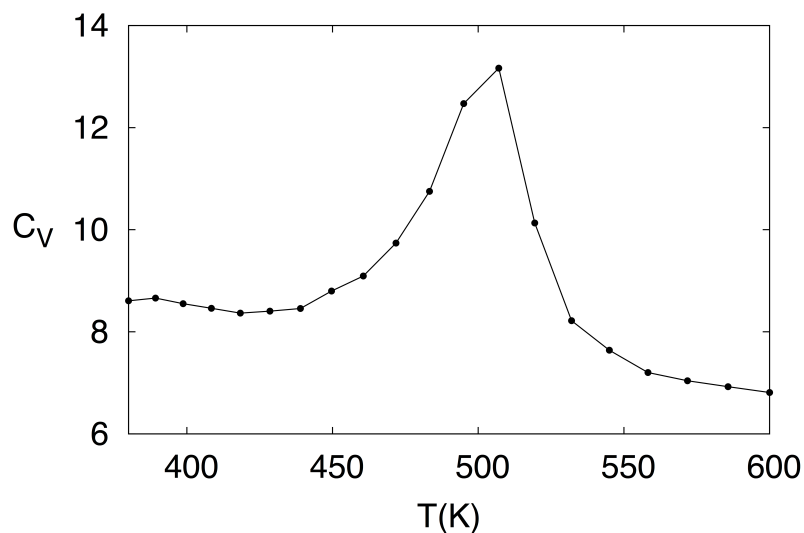
Figure 4.3 The heat capacity of my 12 chain system as a function of temperature.

The process in my system fits the thermodynamic characteristics of a phase

transition, as can be seen by the peak in the heat capacity curve at the temperature point

507K. In order to obtain more information on the dynamics that produce this heat

capacity peak, I calculated additional parameters. The criterion to determine if two

chains are a part of the same aggregate is if the center of mass of any residue of one

chain is within a certain distance of the center of mass of any residue of the other chain.

This distance is $1/2(d^{res1}+d^{res2})$, in which $d^{res}=3.5Å+2*|R_{Cα}-R_{COM}|$. $R_{Cα}$ and R_{COM} are

the positions of the Carbon α atom and the center of mass of the residue respectively.

The reason for the 3.5Å in the formula for d^{res} is if the Carbon α atom and the center of

mass of the residue have the same position. 3.5Å is the distance between two covalently bonded Carbon atoms. Figure 4.4 displays the number of chains in the largest aggregate as a function of temperature, averaged over the last 200 ns. It is clear from Figure 4.4 that significant aggregation occurs only for temperatures below 500K.
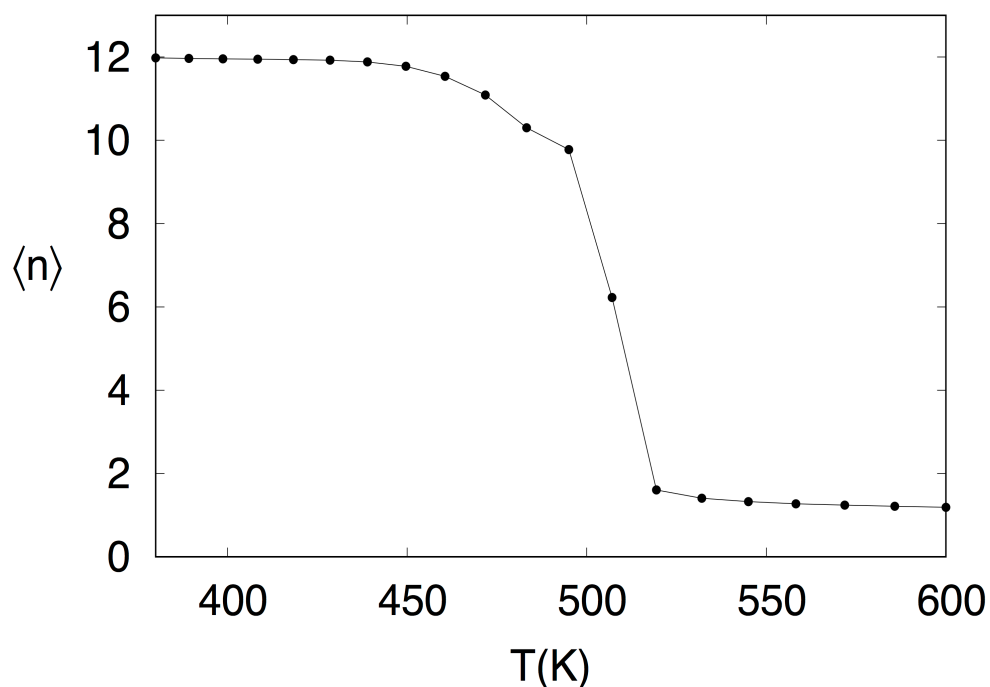
Figure 4.4 The number of chains in the largest aggregate as a function of temperature, averaged over the last 200 ns.

4.4 Validity of Results

I examined the two β-sheet structure formed in replica 3. This type of structure may be important in allowing ccβ to form amyloid-like protofilaments. As can be seen in Figure 4.5a, the two β-sheets are held together by a variety of sidechain interactions.

The blue (positively charged arginine) and red (negatively charged glutamic acid) sidechains are oppositely charged sidechains which provide electrostatic forces between the β sheets. Figure 4.5b shows that similar sidechains extend out on both sides of each sheet, providing the ability for additional sheets to be added to increase the size of the β aggregate. Each sheet also contains white spheres representing hydrophobic sidechains (isoleucine and leucine) on each side, both in the middle between the two sheets (Figure 4.5a) and on the outside (Figure 4.5b). The specific location of these hydrophobic regions may provide specificity that directs the proper alignment[115] of sheets as they are added to the aggregate.

I found that ccβ formed β-sheets that are held together by electrostatic forces between charged sidechains. Charged sidechains are located on both sides of each β-sheet, providing the opportunity for additional β-sheets to be added in layers to increase the size of the aggregate. Regions of hydrophobic sidechains may provide specificity that facilitates alignment of charged sidechains as the β-aggregate grows in size

.
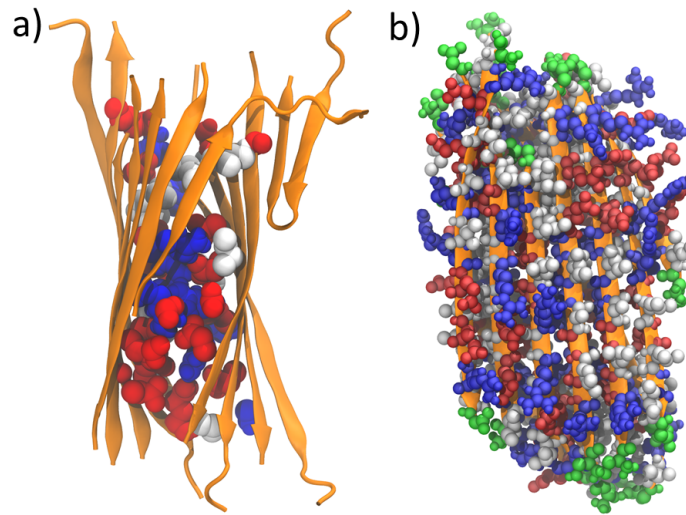
Figure 4.5 From replica 3, positively charged (blue) and negatively charged (red) sidechains provide strong binding interactions: (a) between two β-sheets to hold the sheets together (b) on the outside of the two β-sheets to allow the addition of more β-sheets to the aggregate. Hydrophobic sidechains (white spheres) may provide specificity to facilitate alignment.

In addition to determining the average number of chains in the largest aggregate, I also investigated the amount of order or stability in the size of the largest aggregate. In Figure 4.6 I plot the normalized number fluctuations χ_n for the chains in the largest aggregate

$$\chi_n = \frac{<n^2>-<n>^2}{<n>^2} \qquad (4.3)$$
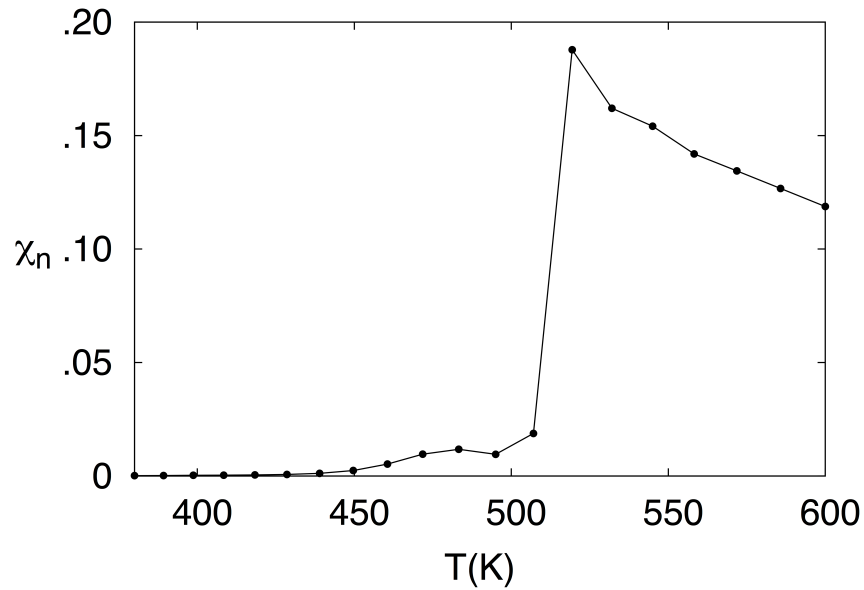
Figure 4.6 Stability of the largest aggregate based upon a parameter χ_n that measures the fluctuations in the number of chains in the largest aggregate.

Even though there are a much larger number of chains in the largest aggregate for temperatures below 500K compared to temperatures above 500K, Fig. 4.6 shows that below 500K the structure of the system is much more stable with very small fluctuations. The isothermal compressibility can be calculated using the expression[35] $\kappa_T = \chi_n(V/KT)$. Figure 4.6 for χ_n shows that the structurally organized aggregate below 500K has a significantly smaller κ_T compared to the disorganized system above 500K, which makes physical sense. Experimenal determination of κ_T is a possibilty for validating my results.

4.5 Temperature Window of β-Sheet Formation

The figures shown above reflect properties of the multichain system and interactions between the chains. I also investigated the behavior and dynamics of individual chains within the system.

Hydrogen bond rearrangement is an important factor accompanying the structural transition from α helix to β sheets. At all of the temperatures that I used in the simulations, the α helices fall apart too quickly (< 1 ns) for me to monitor them and their intra helix hydrogen bonds. However, I was able to monitor the formation of interchain hydrogen bonds as β strands form and aggregate into β sheets. Figure 4.7 shows the time development of interchain hydrogen bonds for different temperatures. Since these results are from REMD simulations, the characteristic times are not reliable. However, the plots show that below 507K, the number of interchain hydrogen bonds increases as the replicas form a β sheet, whereas above 507K the number of interchain hydrogen bonds never stabilizes at a high number, implying that a β sheet never forms.
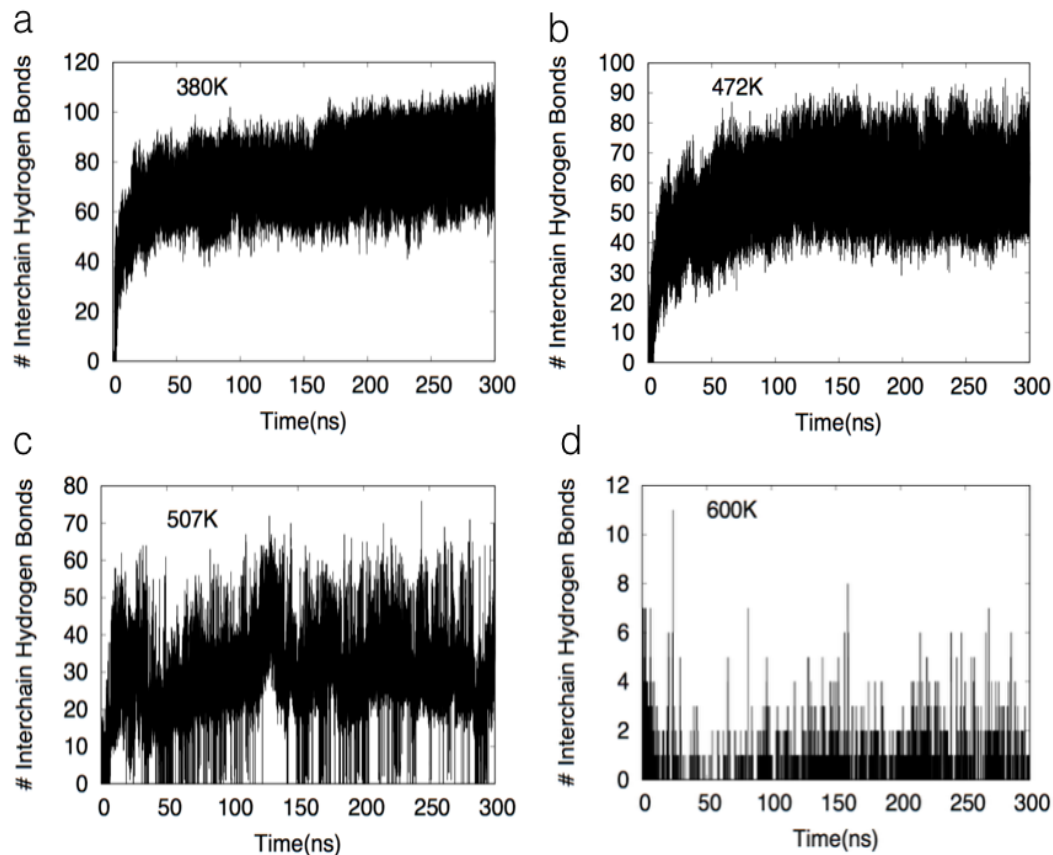
Fig 4.7 Number of interchain hydrogen bonds versus time. (a) 380K (b) 472K (c) 507K (d) 600K.

In order to better understand the role of hydrogen bonds in the structural

transition, I examined how the number of hydrogen bonds varied depended on

temperature at the end of my simulations. Figure 4.8 displays the number of interchain

hydrogen bonds, as well as the number of intrachain hydrogen bonds, averaged over the

final 200 ns. It can be seen that the number of interchain hydrogen bonds drops

suddenly to almost zero above 500K, which matches the temperature dependence

displayed in Figure 4.4 for the number of chains in the largest aggregate. In contrast,

the number of intrachain bonds begins to drop at a temperature of 450K, a lower

temperature than the dramatic drop that occurs for the interchain hydrogen bonds at 500K.
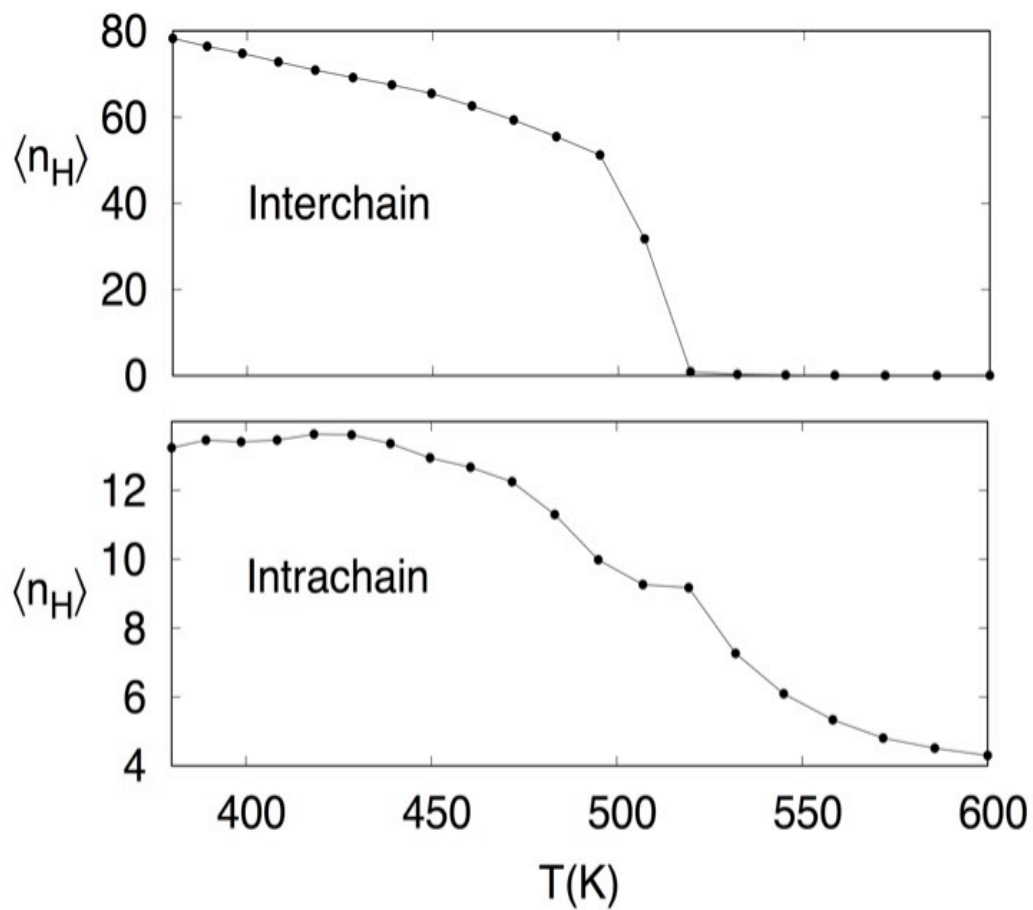
Figure 4.8 Average number of hydrogen bonds as a function of temperature. Top: interchain hydrogen bonds, Bottom: intrachain hydrogen bonds.
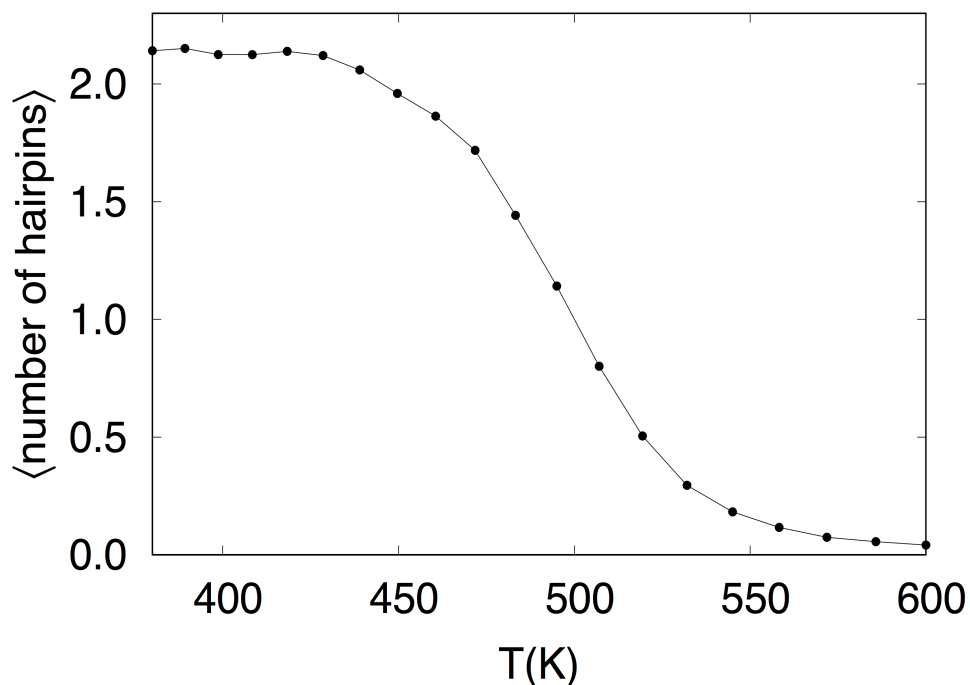
Figure 4.9 The average number of chains that are in a hairpin structure as a function of temperature.

The graphs above that reflect the behavior of the multichain system show a clear structural transition involving β-structure that occurs around 500K (Figure 4.3). Above 500K, the system is highly disordered with negligible aggregation (Figure 4.4) and large fluctuations (Figure 4.6). Below 500K there is almost complete aggregation of all β-chains into stable aggregates with little fluctuation.

In my REMD simulation, if the system does not spend significant time above 450K, the individual chains remain in β-hairpin conformation and form non-amyloid structures. However, if the system spends significant time above 450K, Figure 4.9

shows that the probability for an individual chain to be in a hairpin structure drops significantly. This allows individual chains to straighten-out into β-strands, a process that is facilitated by hairpin structures forming hydrogen bonds with other chains. The straightened β-strands are in a conformation that allows them to bind into β-sheets and form protofibrils. However, Figure 4.8 shows that above 500K, the probability for forming the interchain hydrogen bonds necessary to form β-sheets drops dramatically. As shown in Figure 4.4, this prevents the formation of aggregates above 500K.

The CTMD simulations showed that there tends to be hairpins at low temperatures and straightened-out chains at high temperatures.

To better understand the relevance of interchain and intrachain hydrogen bonds on the structure of individual protein chains, I determined the average number of chains that are in a hairpin structure as a function of temperature. This is plotted in Figure 4.9. A chain was counted as being in a hairpin configuration if it contained four or more intrachain hydrogen bonds. The drop in Figure 4.9 above 450K closely matches the temperature dependence of the number of intrachain bonds displayed in Figure 4.8.

In general for amyloid proteins, the intermediate aggregates are composed of β-strands in the hairpin state along with β-strands that are extended. The smaller oligomers tend to have a higher percentage of β-strands in the hairpin state than the mature fibrils. Hairpins seem more likely to straighten out as an aggregate gets bigger[26] because there are more interactions between chains that can dominate over the intrachain hydrophobic interactions that stabilize the hairpin conformation of a chain. However, even in large β-sheets, proteins in a hairpin conformation can serve as the growing end of the sheet.[26,22,116,117]

Two different structural transition processes involving hydrogen bonds are required for aggregation into fibrils: the breaking of intrachain hydrogen bonds to allow β-hairpin proteins to straighten, and the subsequent formation of interchain hydrogen bonds during β aggregation into amyloid fibrils. For my REMD investigations, I found that the temperature dependence of these two different β structural transition processes results in the existence of a temperature window that the protein experiences during the process of forming protofibrillar structures. Appearance of the temperature window allows me to investigate the dynamics on a molecular level. If the protein system in my simulations spends all of its time at either lower temperatures or higher temperatures relative to this window, fibrillization will not occur.

4.6 Free Energy and Heterogeneity
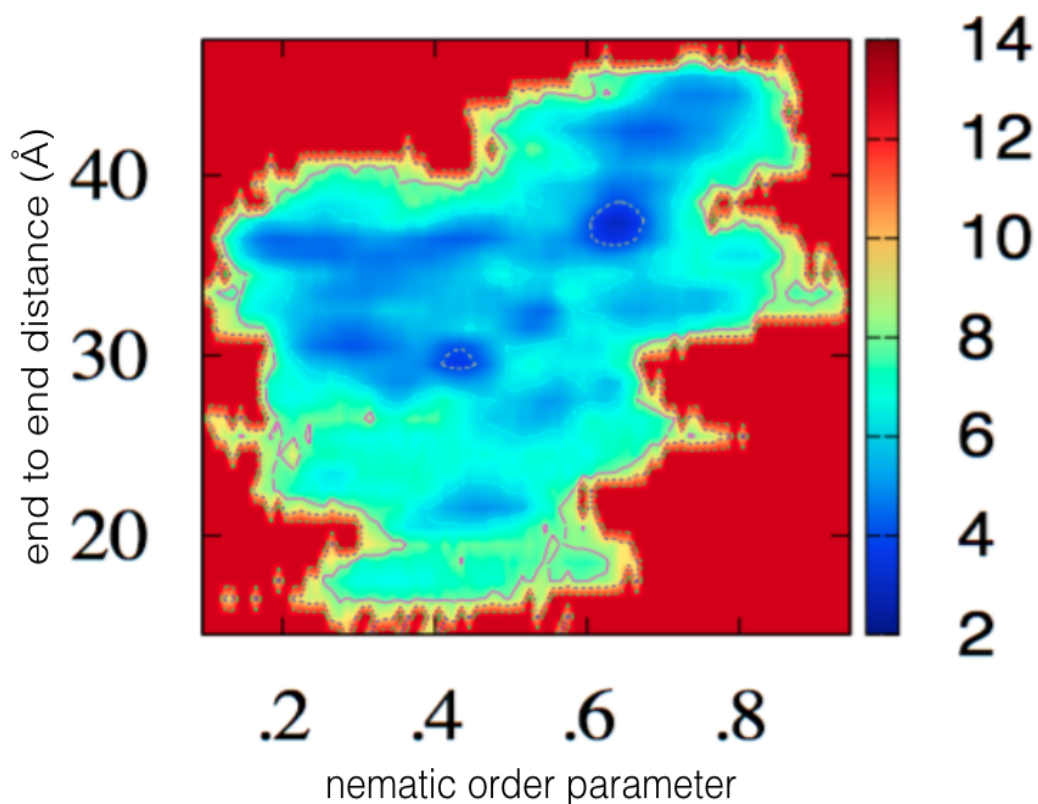
Figure 4.10 Free energy at 380K for all 300ns (minus 20 ps from initialization and 20 ps from equilibration). The end to end distance(d_{ee}) is the distance, averaged over all 12 proteins, from the C_α at one end to the C_α at the other end of the protein.

The free energy for 380K for the last 300 ns (minus 20 ps from initialization and 20 ps from equilibration) is shown in Figure 4.10. The reason I included the free energy for the entire simulation is to include the structures going from a disordered aggregate to an ordered β-aggregate. The free energy was calculated by first creating bins. Each bin included a range of nematic order parameter values and a range of d_{ee} values. For each bin, the number of frames within the S and d_{ee} ranges of that bin for a certain temperature (for whatever replica at the frame that fits the criteria) was counted. The

formula for the free energy, $G=-RTln(P)$, was used, with P the fraction of frames that are within a certain bin.

The graph in Figure 4.10 shows that there is a lot of heterogeneity for the earlier structures as well as for the β-aggregates that form later in the simulation. There are many places of dark blue, and each dark blue island can have many significantly different structures. The structure of the system is not defined totally by d_{ee} and S.
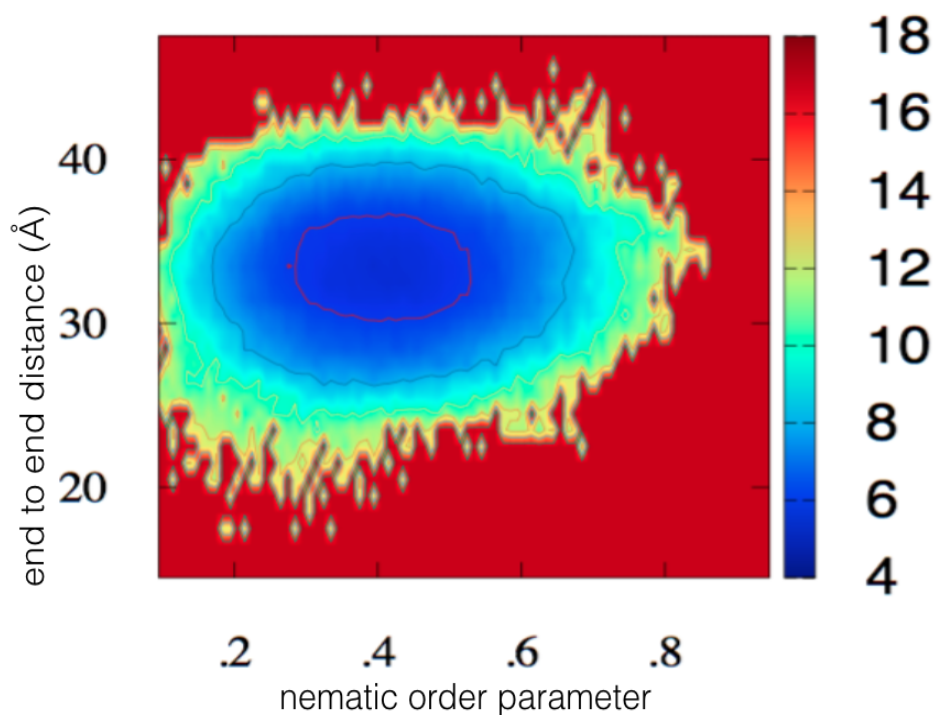
Figure 4.11 Free energy at 507K as a function of d_{ee} and S for the last 200 ns.

The free energy in Figure 4.11 is at the temperature 507K. (The transition temperature 500K was not one of the specific temperatures in my REMD simulation.) There is only one free energy well. If this was a first-order transition, there would likely be two free energy wells at the transition temperature. According to polymer theory, a first-order phase transition happens for a homogenous polymerizing system.[118]

However, a glass-like transition happens for a polymerizing system that is heterogenous.[118] A glass-like transition is a continuous transition.[118] It has similarities to a second-order phase transition, but is not strictly a second-order phase transition.[118]

The last frame of replica 3 is shown as the last panel of in Figure 4.1. The last frame in replica 3 looks more like a steric protofilament than the last frame of any of the other replicas. Replicas 14 and 16 also look like steric zipper protofilaments. However, you can see large structural differences among replica 3, 14, and 16. Figure 4.12 shows the last frame of replicas 3, 14, and 16 next to each other.
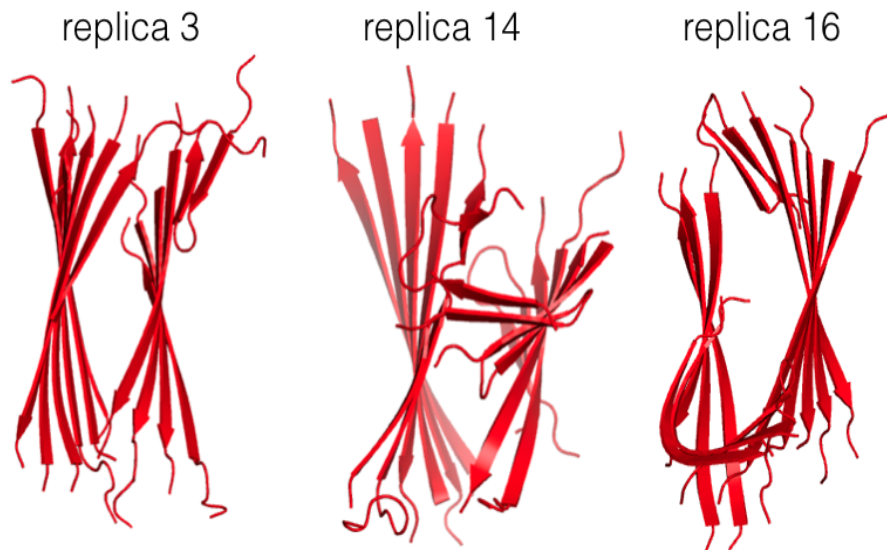
Figure 4.12 Last frame of replicas 3, 14, and 16 next to each other.

4.7 Protein to Protein Interactions

The CTMD showed that the fibril formation occurs much quicker with a smaller confining potential radius as opposed to a larger confining potential radius. With a smaller confining potential radius, there is more protein to protein contact and hence more chance for proteins to catalyze each other to change other proteins' secondary

structure. Increasing the concentration in amyloid structures has the effect of speeding up the aggregation process.

5. CONCLUSIONS

To investigate the kinetics and thermodynamics of amyloid formation in amyloidogenic proteins, I used various mathematical and computational tools. Specifically, I developed a mathematical model to investigate the population kinetics of structural conversions leading to Aβ fibril formation. I then used molecular dynamics (MD) computational simulations to investigate the molecular details of the structural transformations and calculate thermodynamic parameters for amyloid formation. I employed a special form of molecular dynamics, Replica Exchange Molecular Dynamics (REMD), which is described in section 2.2.5, to simulate a system of 12 ccβ proteins. I found that the results of both the mathematical modeling of the kinetics and the molecular dynamics computational simulations could be explained in terms of two important molecular-level aspects. Section 5.1 emphasizes how hydrogen bond rearrangements correlate to aggregation phase transitions and population dynamics of secondary structure. In section 5.2 I describe how my results show the prevalence of the heterogeneity in the systems I studied. Section 5.3 discusses more ideas for MD simulations and suggests experimental ideas based on my results.

5.1 Hydrogen Bonds

The model I developed to fit the population of secondary structure kinetics data that was available from experiments is described in Fig. 3.1 and Table 3.1. For my fitting of the Kirkitadze data[28], Fig. 3.2, I fit changes of secondary structure with time. Changes in secondary structure with time correlate with changes in the number of

interchain and intrachain hydrogen bonds. α helices have a lot of intrachain hydrogen bonds, whereas β-sheets have a lot of interchain hydrogen bonds.

Two different structural transition processes involving hydrogen bonds are required for aggregation into fibrils: the breaking of intrachain hydrogen bonds to allow hairpin proteins to straighten, and the subsequent formation of interchain hydrogen bonds during β aggregation into amyloid fibrils. For my REMD investigations, I found that the temperature dependence of these two different β structural transition processes results in the existence of a temperature window that my system experiences during the process of forming protofibrillar structures. If a replica in my simulation spends all of its time at either lower temperatures or higher temperatures relative to the temperature window, fibrillization will not occur.

My MD simulations look at molecular-level details of some of the stages whose transition kinetics I mathematically modeled. A summary of the steps from the MD simulations of the ccβ replica that created the aggregate that most resembled a steric zipper protofilament is shown in Figure 4.1.

In Figure 4.7, the number of interchain hydrogen bonds versus time for four different temperatures is displayed. For the temperatures 380K and 472K, the number of interchain bonds increases as time advances. The reason for the increase in the number of interchain hydrogen bonds is the formation of β-sheets. For temperature 507K, the number of interchain hydrogen bonds fluctuates rapidly. The reason for the rapid fluctuation is that this is the transition temperature, in which some replicas form β-sheets and some don't. At 600K, there are few interchain hydrogen bonds throughout the simulation, since the temperature is too high to allow β-sheets.

I found that the hydrogen bond rearrangement is an important factor in the structural transition and helps elucidate the temperature dependency. The initial configuration of the 12 chains was all α-helix, which means many intrachain hydrogen bonds. At all temperatures, the α helices fell apart too quickly (<1 ns) for me to be able to monitor the helical intrachain hydrogen bonds. I was able to monitor the increase of interchain hydrogen bonds as β sheets form, as shown in Figure 4.7. Since my simulations used REMD, the time scales in Fig. 4.7 are not reliable, but the temperature dependencies are. I examined the temperature dependencies more carefully. Figure 4.8 displays the number of interchain hydrogen bonds, as well as the number of intrachain hydrogen bonds, averaged over the final 200 ns. It can be seen that the number of interchain hydrogen bonds drops suddenly to almost zero above 500K, which matches the temperature dependence displayed in Figure 4.4 for the number of chains in the largest aggregate. In contrast, the number of intrachain hydrogen bonds begins to drop at a temperature of 450K, a lower temperature than the dramatic drop that occurs for the interchain hydrogen bonds at 500K.

At high temperature the system being entropically favored makes it so that random coils are the preferred secondary structure. Hydrogen bonds don't last long in proteins in random coil because each protein can sample many different configurations. At low temperature the system being enthalpically favored makes the proteins have many hydrogen bonds. The proteins having many hydrogen bonds lowers the energy, but also makes it a low entropy configuration.

The graphs that reflect the behavior of the multichain system show a clear structural transition involving β-structure that occurs around 500K (Figure 4.3). Above

500K, the preferred configuration of the system is highly disordered with negligible aggregation (Figure 4.4) and large fluctuations (Figure 4.6). Below 500K there is almost complete aggregation of all β-chains into stable aggregates with little fluctuation.

Interestingly, my REMD simulation displayed dynamics defined by a temperature window for the formation of organized β-sheets in the region of 450K to 500K. If the system does not spend significant time above 450K, the individual chains remain in hairpin conformation and form non-amyloid structures. If the system spends significant time above 450K, Figure 4.9 shows that the probability for an individual chain to be in a hairpin structure drops significantly. This allows individual chains to straighten-out into β-strands, a process that is facilitated by hairpin structures forming hydrogen bonds with other chains. The straightened β-strands are in a conformation that allows them to form interchain hydrogen bonds and organize into β-sheets that can form protofibrils. However, Fig. 4.8 shows that above 500K, the probability for forming the interchain hydrogen bonds necessary to form β-sheets drops dramatically. As shown in Fig. 4.4, this prevents the formation of any type of aggregate above 500K.

To better understand the relevance of interchain and intrachain hydrogen bonds on the structure of individual protein chains, I investigated more closely the hairpin intermediate conformation that prevents chains from organizing into β-sheets. I determined the average number of chains that are individually in a hairpin structure as a function of temperature. This is plotted in Fig. 4.9. A chain was counted as being in a hairpin configuration if it contained four or more intrachain hydrogen bonds. The drop

93

in Fig. 4.9 above 450K closely matches the temperature dependence of the number of intrachain bonds displayed in Fig. 4.8.

5.2 Heterogeneity

The ccβ protein system I simulated had 12 chains, each with 17 amino acid residues. Based on simulations performed by other researchers of two short Aβ(16-22) peptides that showed significant heterogeneity in forming dimers[18] we can expect significantly more heterogeneity in my system of ccβ aggregation. The reasons for the greater heterogeneity are that my system has more chains and each ccβ chain is longer than each Aβ(16-22) peptide. We can expect even more heterogeneity in systems of Kirkitadze et al.[28] and Walsh et al.[37], having a large number of Aβ(1-40) proteins.

The mathematical model I used to fit population dynamics of secondary structure is shown in Figure 3.1. To represent the heterogeneity in the system, in the mathemaitcal model I used sub-compartments which are mathematically equivalent to heterogeneous distributions. My fit to the kinetic data in the article by Kirkitadze[28], Fig 3.2, had four sub-compartments which shows a significant amount of heterogeneity. I spent a lot of time working to get the parameters just right for the best fit, including the number of sub-compartments. The more sub-compartments, the more heterogeneity.[77] The shapes in the population dynamics curves are significantly different with four sub-compartments than with no sub-compartments, as shown in Figure 3.15 of in the book *Modeling Infectious Diseases in Humans and Animlas* by Keeling.[77] My fitting to the data from the Walsh article[37], Fig 3.3a, had six sub-compartments, displaying even more heterogeneity than in the data from the Kirkitadze article.[28]

94

In addition to the heterogeneity in my modeling of the kinetics, my MD simulations also manifested significant heterogeneity. From the MD results, I calculated the free energy of various configurations of the 12 chain system. The amount of heterogeneity in my ccβ simulations can be seen in the free energy plot for a temperature of 380K, Figure 4.10. There are a many different regions on the free energy plot that have low free energy. In addition, each region of low free energy on that diagram contains different structures.

According to polymer theory, a first-order phase transition can occur in a homogenous polymerizing system.[118] However, a glass-like transition happens for a polymerizing system that is heterogenous.[118] A glass-like transition is a continuous transition.[118] It has similarities to a second-order phase transition, but is not strictly a second-order phase transition.[118] There weren't enough data points in the heat capacity curve, Figure 4.3, to decide whether it was a first-order phase transition. However, I deduced from the free energy diagram at the transition temperature, 507K, Figure 4.11 that my system did not display a first-order phase-transition. There is only one free energy well that spans the different large-scale configurations, showing a continuous phase transition.

The heterogeneity in my REMD simulations was also shown by the different replicas. Not all replicas formed β-aggregates. Those replicas that produced β-aggregates all produced a different β-aggregate or β-aggregates. The last frame of the three replicas that most resembled steric zipper protofilaments are shown in Figure 4.12. They look significantly different from each other.

Figure 4.2 shows the nematic order parameter *S* for the proteins that make up the largest β-sheet for the replica that most resembled a steric zipper protofilament. A value of one for *S* means there is perfect alignment of all the protein direction vectors and zero means total isotropy in the protein direction vectors. Figure 4.2 displays large changes in *S* during the first 40 ns while the six β-strands undergo major rearrangements relative to each other. This shows a lot of heterogeneity in the early micelle-like aggregates.

5.3 Directions for Future Work

If I had more time, I would have wanted to do simulated annealing constant temperature MD to help verify the temperature window hypothesized from my REMD simulation. The procedure for simulated annealing constant temperature MD is as follows: (1) perform constant temperature MD with the system of 12 ccβ proteins at 600K for 15ns; (2) take the last frame of the simulation at 600K and use it for another constant temperature simulation at 590K for 15ns; (3) take the last frame of the simulation at 590K and use it for the first frame of a simulation at 580K; (4) continue this process down to 380K.

Other simulations of ccβ would be around 20 different 150 ns constant temperature MD simulations of my ccβ system at the same temperature and other conditions. As mentioned in Chapter 4, 100 ns is about the time it takes a system to equilibrate. Because MD has some random numbers in it, each simulation is likely to create unique structures and dynamics for each simulation, even though they all would

have the same conditions. Seeing how different each simulation is from the other ones will help to understand how heterogenous the aggregation process is.

Both KFFEAAAKKFFE and KFFEYNGKKFFE have four amno acids flanked on both sides by the de novo peptide KFFE. KFFEAAAKKFFE does a better job of forming amyloid fibrils than KFFEYNGKKFFE.[119] KFFEYNGKKFFE folds into a stable β-hairpin.[119] MD simulations on a system of several KFFEYNGKKFFE peptides and on a system of several KFFEAAAKKFFE peptides could help investigate why KFFEYNGKKFFE has a higher propensity than KFFEAAAKKFFE to form β-hairpins.

For ccβ, temperature is the control parameter for its structural transitions. One idea for experiments on ccβ is to have the experiment start with high temperature and slowly lower the temperature. This change in the temperature during the experiment can help facilitate the process of fibrillization.

For the REMD simulations on ccβ, I found hydrogen bonds between backbone atoms. This made my results generalizable to any protein because every protein has the same backbone. The two different structural transitions involving hydrogen bonds that are controlled by temperature in ccβ may also create a window for other amyloid forming proteins, but the window-defining control parameter may not be temperature. For other amyloid forming proteins in which temperature is not the structural transition control parameter, perhaps another control parameter may be used based on the knowledge gained from the current work on ccβ for which there is a temperature window.

Figure 3.4 helps explain the energetics of how Trifluoroethanol (TFE) affects the secondary structure. Intermolecular bonds are formed at low TFE concentrations,

97

whereas, at high TFE concentrations, intramolecular hydrogen bonds are preferred.[85] The temperature window found for ccβ can help explain why too much TFE for Aβ mentioned in section 3.4.1 impedes fibrillization. A condition is necessary in which there can be breaking of intrachain hydrogen bonds, not only for the α helices to break once they've formed but also for the hairpins to straighten out.

One idea is to experimentally start out with a high concentration of TFE for Aβ. This can stabilize the α helices. Then the experimentalist can lower the concentration of TFE after many α helices have been formed. Lowering the concentration of TFE helps not only the α helices to break but also helps straighten out the β-strands.

An idea for therapies to prevent amyloid diseases is to stabilize the proteins in their hairpin state. One idea for a control parameter to be put into medications are affibodies.[120,121] Affibodies are binding proteins derived from the Z domain of staphylococcal protein A.[121] Affibodies are able to stabilize the hairpin state of proteins by clamping down on them.[122]

REFERENCES

1.   Schnabel, J., Protein folding: The dark side of proteins. *Nature* **2010,** *464* (7290), 828-9.

2.   Lomakin, A.; Teplow, D. B.; Kirschner, D. A.; Benedek, G. B., Kinetic theory of fibrillogenesis of amyloid beta-protein. *Proc Natl Acad Sci U S A* **1997,** *94* (15), 7942-7.

3.   Lomakin, A.; Chung, D. S.; Benedek, G. B.; Kirschner, D. A.; Teplow, D. B., On the nucleation and growth of amyloid beta-protein fibrils: detection of nuclei and quantitation of rate constants. *Proc Natl Acad Sci U S A* **1996,** *93* (3), 1125-9.

4.   Stefani, M.; Dobson, C. M., Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* **2003,** *81* (11), 678-99.

5.   Keshet, B.; Yang, I. H.; Good, T. A., Can size alone explain some of the differences in toxicity between beta-amyloid oligomers and fibrils? *Biotechnol Bioeng* **2010,** *106* (2), 333-7.

6.   Mousseau, N.; Derreumaux, P., Exploring the early steps of amyloid peptide aggregation by computers. *Acc Chem Res* **2005,** *38* (11), 885-91.

7.   Pallitto, M. M.; Murphy, R. M., A mathematical model of the kinetics of beta-amyloid fibril growth from the denatured state. *Biophys J* **2001,** *81* (3), 1805-22.

8.   Zerovnik, E., Amyloid-fibril formation. Proposed mechanisms and relevance to conformational disease. *Eur J Biochem* **2002,** *269* (14), 3362-71.

9.   Sipe, J. D., *Amyloid proteins : the beta sheet conformation and disease*. Wiley-VCH: Weinheim, 2005.

10.   Kelly, J. W.; Balch, W. E., Amyloid as a natural product. *J Cell Biol* **2003,** *161* (3), 461-2.

11.   Otzen, D. E., *Amyloid fibrils and prefibrillar aggregates : molecular and biological properties*. Wiley-VCH: Weinheim, Germany, 2013; p xxviii, 481 p.

12.   Giurleo, J. T.; He, X.; Talaga, D. S., Beta-lactoglobulin assembles into amyloid through sequential aggregated intermediates. *J Mol Biol* **2008,** *381* (5), 1332-48.

13.   Hellstrand, E.; Boland, B.; Walsh, D. M.; Linse, S., Amyloid beta-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem Neurosci* **2010,** *1* (1), 13-8.

14.   Jarrett, J. T.; Berger, E. P.; Lansbury, P. T., Jr., The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. *Biochemistry* **1993,** *32* (18), 4693-7.

15.   Xue, W. F.; Homans, S. W.; Radford, S. E., Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc Natl Acad Sci U S A* **2008,** *105* (26), 8926-31.

16.   Rahimi, F.; Bitan, G.; SpringerLink (Online service), Non-fibrillar Amyloidogenic Protein Assemblies - Common Cytotoxins Underlying Degenerative Diseases. Springer Netherlands,: Dordrecht, 2012; p. 1 online resource. http://dx.doi.org/10.1007/978-94-007-2774-8.

17.   Dobson, C. M., Protein misfolding, evolution and disease. *Trends Biochem Sci* **1999,** *24* (9), 329-32.

18.   Santini, S.; Wei, G.; Mousseau, N.; Derreumaux, P., Pathway complexity of Alzheimer's beta-amyloid Abeta16-22 peptide assembly. *Structure* **2004,** *12* (7), 1245-55.

19.   Hurshman, A. R.; White, J. T.; Powers, E. T.; Kelly, J. W., Transthyretin aggregation under partially denaturing conditions is a downhill polymerization. *Biochemistry* **2004,** *43* (23), 7365-81.

20.   Steckmann, T.; Awan, Z.; Gerstman, B. S.; Chapagain, P. P., Kinetics of peptide secondary structure conversion during amyloid beta-protein fibrillogenesis. *J Theor Biol* **2012,** *301*, 95-102.

21.   Cheon, M.; Chang, I.; Mohanty, S.; Luheshi, L. M.; Dobson, C. M.; Vendruscolo, M.; Favrin, G., Structural reorganisation and potential toxicity of oligomeric species formed during the assembly of amyloid fibrils. *PLoS Comput Biol* **2007,** *3* (9), 1727-38.

22.   Matthes, D.; Gapsys, V.; Daebel, V.; de Groot, B. L., Mapping the Conformational Dynamics and Pathways of Spontaneous Steric Zipper Peptide Oligomerization. *PLoS ONE* **2011,** *6* (5), e19129.

23.   Nelson, R.; Eisenberg, D., Structural models of amyloid-like fibrils. *Adv Protein Chem* **2006,** *73*, 235-82.

24.   Eisenberg, D.; Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Sambashivan, S.; Ivanova, M. I.; Madsen, A. O.; Riekel, C., The structural biology of protein aggregation diseases: Fundamental questions and some answers. *Acc Chem Res* **2006,** *39* (9), 568-75.

25.   Zheng, J.; Jang, H.; Ma, B.; Tsai, C. J.; Nussinov, R., Modeling the Alzheimer Abeta17-42 fibril architecture: tight intermolecular sheet-sheet association and intramolecular hydrated cavities. *Biophys J* **2007,** *93* (9), 3046-57.

26.   Zhou, R., *Molecular modeling at the atomic scale : methods and applications in quantitative biology*. CRC Press, Taylor & Francis Group: Boca Raton, 2015; p xix, 360 pages.

27.   Sabate, R.; Gallardo, M.; Estelrich, J., Temperature dependence of the nucleation constant rate in beta amyloid fibrillogenesis. *Int J Biol Macromol* **2005,** *35* (1-2), 9-13.

28.   Kirkitadze, M. D.; Condron, M. M.; Teplow, D. B., Identification and characterization of key kinetic intermediates in amyloid beta-protein fibrillogenesis. *J Mol Biol* **2001,** *312* (5), 1103-1119.

29.   Ramirez-Alvarado, M.; Kelly, J. W.; Dobson, C. M., *Protein misfolding diseases : current and emerging principles and therapies*. Wiley: Hoboken, N.J., 2010; p xxx, 1045 p.

30.   Serpell, L. C., Alzheimer's amyloid fibrils: structure and assembly. *Biochim Biophys Acta* **2000,** *1502* (1), 16-30.

31.   Petkova, A. T.; Buntkowsky, G.; Dyda, F.; Leapman, R. D.; Yau, W. M.; Tycko, R., Solid state NMR reveals a pH-dependent antiparallel beta-sheet registry in fibrils formed by a beta-amyloid peptide. *J Mol Biol* **2004,** *335* (1), 247-60.

32.   Roher, A. E.; Baudry, J.; Chaney, M. O.; Kuo, Y. M.; Stine, W. B.; Emmerling, M. R., Oligomerizaiton and fibril asssembly of the amyloid-beta protein. *Biochim Biophys Acta* **2000,** *1502* (1), 31-43.

33.   Chimon, S.; Shaibat, M. A.; Jones, C. R.; Calero, D. C.; Aizezi, B.; Ishii, Y., Evidence of fibril-like beta-sheet structures in a neurotoxic amyloid intermediate of Alzheimer's beta-amyloid. *Nat Struct Mol Biol* **2007,** *14* (12), 1157-64.

34.   Murphy, R. M.; Pallitto, M. M., Probing the kinetics of beta-amyloid self-association. *J Struct Biol* **2000,** *130* (2-3), 109-22.

35.   Abney, J. R.; Scalettar, B. A.; Hackenbrock, C. R., On the measurement of particle number and mobility in nonideal solutions by fluorescence correlation spectroscopy. *Biophysical Journal* **1990,** *58* (1), 261-265.

36.   Fezoui, Y.; Teplow, D. B., Kinetic studies of amyloid beta-protein fibril assembly. Differential effects of alpha-helix stabilization. *J Biol Chem* **2002,** *277* (40), 36948-54.

37.   Walsh, D. M.; Hartley, D. M.; Kusumoto, Y.; Fezoui, Y.; Condron, M. M.; Lomakin, A.; Benedek, G. B.; Selkoe, D. J.; Teplow, D. B., Amyloid beta-protein fibrillogenesis. Structure and biological activity of protofibrillar intermediates. *J Biol Chem* **1999,** *274* (36), 25945-52.

38.   Ding, F.; LaRocque, J. J.; Dokholyan, N. V., Direct observation of protein folding, aggregation, and a prion-like conformational conversion. *J Biol Chem* **2005,** *280* (48), 40235-40.

39.   Steinmetz, M. O.; García-Echeverría, C.; Kammerer, R. A., Design of a Coiled-Coil-based Model Peptide System toExplore the Fundamentals of Amyloid Fibril Formation. *International Journal of Peptide Research and Therapeutics* **2005,** *11* (1), 43-52.

40.   Strodel, B.; Fitzpatrick, A. W.; Vendruscolo, M.; Dobson, C. M.; Wales, D. J., Characterizing the first steps of amyloid formation for the ccbeta peptide. *J Phys Chem B* **2008,** *112* (32), 9998-10004.

41.   Missimer, J. H.; Steinmetz, M. O.; Baron, R.; Winkler, F. K.; Kammerer, R. A.; Daura, X.; van Gunsteren, W. F., Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Sci* **2007,** *16* (7), 1349-59.

42.   Kammerer, R. A.; Kostrewa, D.; Zurdo, J.; Detken, A.; Garcia-Echeverria, C.; Green, J. D.; Muller, S. A.; Meier, B. H.; Winkler, F. K.; Dobson, C. M.; Steinmetz, M. O., Exploring amyloid formation by a de novo design. *Proc Natl Acad Sci U S A* **2004,** *101* (13), 4435-40.

43.   Kammerer, R. A.; Steinmetz, M. O., De novo design of a two-stranded coiled-coil switch peptide. *J Struct Biol* **2006,** *155* (2), 146-53.

44.   Nölting, B.; SpringerLink (Online service), Protein Folding Kinetics Biophysical Methods. Second Edition. ed.; Springer-Verlag Berlin Heidelberg,: Berlin, Heidelberg, 2006. http://dx.doi.org/10.1007/b138868.

45.   Newman, M. E. J., *Computational physics*. Rev. and expanded [ed.] ed.; Createspace: Charleston, SC?, 2013; p xi, 549 p.

46.   Landau, R. H.; Páez Mejía, M. J.; Bordeianu, C. C., *Computational physics : problem solving with computers*. 2nd revised and enlarged edition. ed.; Wiley-VCH ; John Wiley distributor: Weinheim
Chichester, 2007; p xxi, 593 p.

47.   Schlick, T.; SpringerLink (Online service), Molecular Modeling and Simulation: An Interdisciplinary Guide An Interdisciplinary Guide. In *Interdisciplinary Applied Mathematics,* [Online] Second. ed.; Springer Science+Business Media, LLC,: New York, NY, 2010; p. 1 online resource. http://dx.doi.org/10.1007/978-1-4419-6351-2.

48.   Becker, O. M., *Computational biochemistry and biophysics*. M. Dekker: New York, 2001; p xii, 512 p.

49.   Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: the biomolecular simulation program. *J Comput Chem* **2009,** *30* (10), 1545-614.

50.   Finkelstein, A. V.; Ptitsyn, O. Bebrary Inc., Protein physics a course of lectures. In *Soft condensed matter, complex fluids and biomaterials series* [Online] Academic Press,: Amsterdam ; Boston, 2002; pp. xix, 354 p. http://site.ebrary.com/lib/yale/Doc?id=10186414.

51.   Frenkel, D.; Smit, B., *Understanding molecular simulation : from algorithms to applications*. 2nd ed.; Academic Press: San Diego, Calif. London, 2002; p xxii, 638 p.

52.   Rao, F.; Caflisch, A., Replica exchange molecular dynamics simulations of reversible folding. *J Chem Phys* **2003,** *119* (7), 4035-4042.

53.   Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999,** *314* (1-2), 141-151.

54.   Gc, J. B.; Bhandari, Y. R.; Gerstman, B. S.; Chapagain, P. P., Molecular dynamics investigations of the alpha-helix to beta-barrel conformational transformation in the RfaH transcription factor. *J Phys Chem B* **2014,** *118* (19), 5101-8.

55.   Pu, M.; Garrahan, J. P.; Hirst, J. D., Comparison of implicit solvent models and force fields in molecular dynamics simulations of the PB1 domain. *Chemical Physics Letters* **2011,** *515* (4–6), 283-289.

56.   Ferrara, P.; Apostolakis, J.; Caflisch, A., Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins* **2002,** *46* (1), 24-33.

57.   Lazaridis, T.; Karplus, M., Effective energy function for proteins in solution. *Proteins* **1999,** *35* (2), 133-52.

58.   Feig, M.; Karanicolas, J.; Brooks, C. L., 3rd, MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph Model* **2004,** *22* (5), 377-95.

59.   Khandogin, J.; Brooks, C. L., 3rd, Linking folding with aggregation in Alzheimer's beta-amyloid peptides. *Proc Natl Acad Sci U S A* **2007,** *104* (43), 16880-5.

60.   Antzutkin, O. N.; Balbach, J. J.; Leapman, R. D.; Rizzo, N. W.; Reed, J.; Tycko, R., Multiple quantum solid-state NMR indicates a parallel, not antiparallel, organization of beta-sheets in Alzheimer's beta-amyloid fibrils. *Proc Natl Acad Sci U S A* **2000,** *97* (24), 13045-50.

61.   Benzinger, T. L.; Gregory, D. M.; Burkoth, T. S.; Miller-Auer, H.; Lynn, D. G.; Botto, R. E.; Meredith, S. C., Two-dimensional structure of beta-amyloid(10-35) fibrils. *Biochemistry* **2000,** *39* (12), 3491-9.

62.   Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R., A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A* **2002,** *99* (26), 16742-7.

63.   Murphy, R. M., Peptide aggregation in neurodegenerative disease. *Annu Rev Biomed Eng* **2002,** *4*, 155-74.

64.   Inouye, H.; Kirschner, D. A., A beta fibrillogenesis: kinetic parameters for fibril formation from congo red binding. *J Struct Biol* **2000,** *130* (2-3), 123-9.

65.   Hortschansky, P.; Schroeckh, V.; Christopeit, T.; Zandomeneghi, G.; Fandrich, M., The aggregation kinetics of Alzheimer's beta-amyloid peptide is controlled by stochastic nucleation. *Protein Sci* **2005,** *14* (7), 1753-9.

66.   Bitan, G.; Vollers, S. S.; Teplow, D. B., Elucidation of primary structure elements controlling early amyloid beta-protein oligomerization. *J Biol Chem* **2003,** *278* (37), 34882-9.

67.   Hortschansky, P.; Christopeit, T.; Schroeckh, V.; Fandrich, M., Thermodynamic analysis of the aggregation propensity of oxidized Alzheimer's beta-amyloid variants. *Protein Sci* **2005,** *14* (11), 2915-8.

68.   Auer, S.; Dobson, C. M.; Vendruscolo, M.; Maritan, A., Self-templated nucleation in peptide and protein aggregation. *Phys Rev Lett* **2008,** *101* (25), 258101.

69.   Gursky, O.; Aleshkov, S., Temperature-dependent beta-sheet formation in beta-amyloid Abeta(1-40) peptide in water: uncoupling beta-structure folding from aggregation. *Biochim Biophys Acta* **2000,** *1476* (1), 93-102.

70.   Lee, J. P.; Stimson, E. R.; Ghilardi, J. R.; Mantyh, P. W.; Lu, Y. A.; Felix, A. M.; Llanos, W.; Behbin, A.; Cummings, M.; Van Criekinge, M.; et al., 1H NMR of A beta amyloid peptide congeners in water solution. Conformational changes correlate with plaque competence. *Biochemistry* **1995,** *34* (15), 5191-200.

71. Kheterpal, I.; Williams, A.; Murphy, C.; Bledsoe, B.; Wetzel, R., Structural features of the Abeta amyloid fibril elucidated by limited proteolysis. *Biochemistry* **2001,** *40* (39), 11757-67.

72. Takano, K., Amyloid beta conformation in aqueous environment. *Curr Alzheimer Res* **2008,** *5* (6), 540-7.

73. Baumketner, A.; Bernstein, S. L.; Wyttenbach, T.; Bitan, G.; Teplow, D. B.; Bowers, M. T.; Shea, J. E., Amyloid beta-protein monomer structure: a computational and experimental study. *Protein Sci* **2006,** *15* (3), 420-8.

74. Garai, K.; Sahoo, B.; Sengupta, P.; Maiti, S., Quasihomogeneous nucleation of amyloid beta yields numerical bounds for the critical radius, the surface tension, and the free energy barrier for nucleus formation. *J Chem Phys* **2008,** *128* (4), 045102.

75. Chapagain, P. P.; van Kessel, J. S.; Karns, J. S.; Wolfgang, D. R.; Hovingh, E.; Nelen, K. A.; Schukken, Y. H.; Grohn, Y. T., A mathematical model of the dynamics of Salmonella Cerro infection in a US dairy herd. *Epidemiol Infect* **2008,** *136* (2), 263-72.

76. Lloyd, A. L., Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor Popul Biol* **2001,** *60* (1), 59-71.

77. Keeling, M. J.; Rohani, P., *Modeling infectious diseases in humans and animals*. Princeton University Press: Princeton, 2008; p xi, 366 p.

78. Lansbury, P. T., A Reductionist View of Alzheimer's Disease. *Accounts of Chemical Research* **1996,** *29* (7), 317-321.

79. Perczel, A.; Park, K.; Fasman, G. D., Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: a practical guide. *Anal Biochem* **1992,** *203* (1), 83-93.

80. Sreerama, N.; Venyaminov, S. Y.; Woody, R. W., Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal Biochem* **2000,** *287* (2), 243-51.

81. Sengupta, P.; Garai, K.; Sahoo, B.; Shi, Y.; Callaway, D. J.; Maiti, S., The amyloid beta peptide (Abeta(1-40)) is thermodynamically soluble at physiological concentrations. *Biochemistry* **2003,** *42* (35), 10506-13.

82. Williams, A. D.; Portelius, E.; Kheterpal, I.; Guo, J. T.; Cook, K. D.; Xu, Y.; Wetzel, R., Mapping abeta amyloid fibril secondary structure using scanning proline mutagenesis. *J Mol Biol* **2004,** *335* (3), 833-42.

83.   Zagorski, M. G.; Yang, J.; Shao, H.; Ma, K.; Zeng, H.; Hong, A., Methodological and chemical factors affecting amyloid beta peptide amyloidogenicity. *Methods Enzymol* **1999,** *309*, 189-204.

84.   Feig, M., *Modeling solvent environments : applications to simulations of biomolecules*. Wiley-VCH: Weinheim, 2010; p xviii, 316 p.

85.   Buck, M., Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Q Rev Biophys* **1998,** *31* (3), 297-355.

86.   Roccatano, D.; Colombo, G.; Fioroni, M.; Mark, A. E., Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: a molecular dynamics study. *Proc Natl Acad Sci U S A* **2002,** *99* (19), 12179-84.

87.   Santiveri, C. M.; Pantoja-Uceda, D.; Rico, M.; Jimenez, M. A., Beta-hairpin formation in aqueous solution and in the presence of trifluoroethanol: a (1)H and (13)C nuclear magnetic resonance conformational study of designed peptides. *Biopolymers* **2005,** *79* (3), 150-62.

88.   Karshikoff, A., *Non-covalent interactions in proteins*. Imperial College Press : World Scientific: London ; Hackensack, 2006; p xii, 333 p.

89.   Anderson, C. O.; Spiegelberg, S.; Prausnitz, J. M.; Blanch, H. W., Effect of secondary structure on the interactions of peptide T4 LYS (11-36) in mixtures of aqueous sodium chloride and 2,2,2,-Trifluoroethanol. 2001.

90.   Chu, H. L.; Lin, S. Y., Temperature-induced conformational changes in amyloid beta(1-40) peptide investigated by simultaneous FT-IR microspectroscopy with thermal system. *Biophys Chem* **2001,** *89* (2-3), 173-80.

91.   Soto, C.; Castano, E. M.; Frangione, B.; Inestrosa, N. C., The alpha-helical to beta-strand transition in the amino-terminal fragment of the amyloid beta-peptide modulates amyloid formation. *J Biol Chem* **1995,** *270* (7), 3063-7.

92.   Lazo, N. D.; Grant, M. A.; Condron, M. C.; Rigby, A. C.; Teplow, D. B., On the nucleation of amyloid beta-protein monomer folding. *Protein Sci* **2005,** *14* (6), 1581-96.

93.   Tomaselli, S.; Esposito, V.; Vangone, P.; van Nuland, N. A.; Bonvin, A. M.; Guerrini, R.; Tancredi, T.; Temussi, P. A.; Picone, D., The alpha-to-beta conformational transition of Alzheimer's Abeta-(1-42) peptide in aqueous media is reversible: a step by step conformational analysis suggests the location of beta conformation seeding. *Chembiochem* **2006,** *7* (2), 257-67.

94.   Velez-Vega, C.; Escobedo, F. A., Characterizing the structural behavior of selected Abeta-42 monomers with different solubilities. *J Phys Chem B* **2011,** *115* (17), 4900-10.

95.   Szabo, Z.; Klement, E.; Jost, K.; Zarandi, M.; Soos, K.; Penke, B., An FT-IR study of the beta-amyloid conformation: standardization of aggregation grade. *Biochem Biophys Res Commun* **1999,** *265* (2), 297-300.

96.   Ma, B.; Nussinov, R., Polymorphic C-terminal beta-sheet interactions determine the formation of fibril or amyloid beta-derived diffusible ligand-like globulomer for the Alzheimer Abeta42 dodecamer. *J Biol Chem* **2010,** *285* (47), 37102-10.

97.   Sabate, R.; Estelrich, J., Evidence of the existence of micelles in the fibrillogenesis of beta-amyloid peptide. *J Phys Chem B* **2005,** *109* (21), 11027-32.

98.   Perutz, M. F.; Windle, A. H., Cause of neural death in neurodegenerative diseases attributable to expansion of glutamine repeats. *Nature* **2001,** *412* (6843), 143-4.

99.   Pellarin, R.; Caflisch, A., Interpreting the aggregation kinetics of amyloid peptides. *J Mol Biol* **2006,** *360* (4), 882-92.

100.   Wang, S. S. S.; Chen, Y.-T.; Chen, P.-H.; Liu, K.-N., A kinetic study on the aggregation behavior of β-amyloid peptides in different initial solvent environments. *Biochemical Engineering Journal* **2006,** *29* (1–2), 129-138.

101.   Kusumoto, Y.; Lomakin, A.; Teplow, D. B.; Benedek, G. B., Temperature dependence of amyloid beta-protein fibrillization. *Proc Natl Acad Sci U S A* **1998,** *95* (21), 12277-82.

102.   Tomski, S. J.; Murphy, R. M., Kinetics of aggregation of synthetic beta-amyloid peptide. *Arch Biochem Biophys* **1992,** *294* (2), 630-8.

103.   Yong, W.; Lomakin, A.; Kirkitadze, M. D.; Teplow, D. B.; Chen, S. H.; Benedek, G. B., Structure determination of micelle-like intermediates in amyloid beta -protein fibril assembly by using small angle neutron scattering. *Proc Natl Acad Sci U S A* **2002,** *99* (1), 150-4.

104.   Lee, C. F., Self-assembly of protein amyloids: a competition between amorphous and ordered aggregation. *Phys Rev E Stat Nonlin Soft Matter Phys* **2009,** *80* (3 Pt 1), 031922.

105.   Sachse, C.; Xu, C.; Wieligmann, K.; Diekmann, S.; Grigorieff, N.; Fandrich, M., Quaternary structure of a mature amyloid fibril from Alzheimer's Abeta(1-40) peptide. *J Mol Biol* **2006,** *362* (2), 347-54.

106. Derreumaux, P., *Alzheimer's disease : insights into low molecular weight and cytotoxic aggregates from in vitro and computer experiments : molecular basis of amyloid-beta protein aggregation and fibril formation*. Imperial College Press ; Distributed by World Scientific Publishing Co. Pte. Ltd.: London Singapore, 2013; p xix, 444 pages.

107. Rathore, N.; Chopra, M.; de Pablo, J. J., Optimal allocation of replicas in parallel tempering simulations. *J Chem Phys* **2005,** *122* (2), 024111.

108. Kone, A.; Kofke, D. A., Selection of temperature intervals for parallel-tempering simulations. *J Chem Phys* **2005,** *122* (20), 206101.

109. Isobe, T., *Amyloid and amyloidosis*. Plenum Press: New York, 1988; p xx, 885 p.

110. Tycko, R., Molecular structure of amyloid fibrils: insights from solid-state NMR. *Q Rev Biophys* **2006,** *39* (1), 1-55.

111. Li, D. W.; Mohanty, S.; Irback, A.; Huo, S., Formation and growth of oligomers: a Monte Carlo study of an amyloid tau fragment. *PLoS Comput Biol* **2008,** *4* (12), e1000238.

112. Yu, L.; Edalji, R.; Harlan, J. E.; Holzman, T. F.; Lopez, A. P.; Labkovsky, B.; Hillen, H.; Barghorn, S.; Ebert, U.; Richardson, P. L.; Miesbauer, L.; Solomon, L.; Bartley, D.; Walter, K.; Johnson, R. W.; Hajduk, P. J.; Olejniczak, E. T., Structural characterization of a soluble amyloid beta-peptide oligomer. *Biochemistry* **2009,** *48* (9), 1870-7.

113. Demus, D., *Physical properties of liquid crystals*. Wiley-VCH: Weinheim ; New York, 1999; p xix, 503 p.

114. Knowles, T. P.; De Simone, A.; Fitzpatrick, A. W.; Baldwin, A.; Meehan, S.; Rajah, L.; Vendruscolo, M.; Welland, M. E.; Dobson, C. M.; Terentjev, E. M., Twisting transition between crystalline and fibrillar phases of aggregated peptides. *Phys Rev Lett* **2012,** *109* (15), 158101.

115. Chapagain, P. P.; Gerstman, B. S., Removal of kinetic traps and enhanced protein folding by strategic substitution of amino acids in a model alpha-helical hairpin peptide. *Biopolymers* **2006,** *81* (3), 167-78.

116. Larini, L.; Shea, J. E., Role of beta-hairpin formation in aggregation: the self-assembly of the amyloid-beta(25-35) peptide. *Biophys J* **2012,** *103* (3), 576-86.

117. Gill, A. C., beta-hairpin-mediated formation of structurally distinct multimers of neurotoxic prion peptides. *PLoS One* **2014,** *9* (1), e87354.

118. Odian, G. G., *Principles of polymerization*. 4th ed.; Wiley: Hoboken, N.J., 2004; p xxiv, 812 p.

119. Hosia, W.; Bark, N.; Liepinsh, E.; Tjernberg, A.; Persson, B.; Hallen, D.; Thyberg, J.; Johansson, J.; Tjernberg, L., Folding into a beta-hairpin can prevent amyloid fibril formation. *Biochemistry* **2004,** *43* (16), 4655-61.

120. Hoyer, W.; Hard, T., Interaction of Alzheimer's A beta peptide with an engineered binding protein--thermodynamics and kinetics of coupled folding-binding. *J Mol Biol* **2008,** *378* (2), 398-411.

121. Nord, K.; Gunneriusson, E.; Ringdahl, J.; Stahl, S.; Uhlen, M.; Nygren, P. A., Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat Biotechnol* **1997,** *15* (8), 772-7.

122. Hoyer, W.; Gronwall, C.; Jonsson, A.; Stahl, S.; Hard, T., Stabilization of a beta-hairpin in monomeric Alzheimer's amyloid-beta peptide inhibits amyloid formation. *Proc Natl Acad Sci U S A* **2008,** *105* (13), 5099-104.

VITA

TIMOTHY M. STECKMANN

2009-2015   Teaching Assistant
Florida International University
Miami, Florida, USA

2006-2007   Teaching English as a Second Language
English Friends School
Busan, South Korea

2004-2005   Teaching English as a Second Language
Ding Ding Dong English School
Pohang, South Korea

2001-2004   Master of Science in Physics
Oregon State University
Corvallis, Oregon, USA

1992-1999   Bachelor of Science in Physics
Oregon State University
Corvallis, Oregon, USA

PUBLICATIONS AND PRESENTATIONS

Steckmann, T.; Awan, Z.; Gerstman, B. S.; Chapagain, P. P., Kinetics of peptide secondary structure conversion during amyloid beta-protein fibrillogenesis. *J Theor Biol* 2012, *301*, 95-102.

American Physical Society March Meeting, Denver, Colorado March 2014 "Structural Transitions and Aggregation in Amyloidogenic Proteins"