

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Contrasting disease and non-disease protein aggregation by molecular simulations

Permalink

<https://escholarship.org/uc/item/7v14w7sc>

Author

Fawzi, Nicolas Lux

Publication Date

2008-01-03

Peer reviewed|Thesis/dissertation

Contrasting disease and non-disease protein aggregation by molecular simulations

by

Nicolas Lux Fawzi

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOENGINEERING

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

AND

UNIVERSITY OF CALIFORNIA, BERKELEY

Contrasting disease and non-disease protein aggregation by molecular simulations

Copyright © 2007
by
Nicolas Lux Fawzi

Dedicated to my godmother, Elfriede Van den Eeckhout

Acknowledgements

I thank Teresa Head-Gordon, my dissertation chair and research mentor, for her training over the past years. I have enjoyed hours of invaluable conversation in her office and appreciate the countless times she has spurred me to dig a little deeper. Her passion for research in proteins and chemical physics has worn off on me. The combination of stimulating research collaboration and camaraderie in the group she has and continues to build is second to none. I am grateful for the opportunity to have worked, laughed, traveled, and lunched with all the group members, including Scott Brown, Louis Clark, Rajesh Murarka, Neelan Marianayagam, Cécile Malardier-Jugroot, Jonathan Kohn, Jory Ruscio, Alex Sodt, Greg Hura, Elizabeth Verschell, Brian Carnes, Matthew Lin, Eng Hui Yap, Margaret Johnson, Kevin Kohlstedt and Yuka Okabe.

The resources of the broader Berkeley and UCSF communities as a whole have made this work possible. Harvey Blanch has always provided a friendly smile and experimental context to this work. I continue to benefit from the few years head start along this path that Troy Cellmer has. The success of the last segment of this work and my future direction in NMR spectroscopy are direct results of the overwhelmingly positive Michaela Doucleff. The advice and assistance concerning this new field from David Wemmer and Aaron Phillips is much appreciated. Ken Dill and Matt Jacobson have both taken the initiative to offer guidance and assistance at a number of critical points throughout the completion of this work. My bioengineering classmates, especially Sapun Parekh, Mike Rosenbluth and Erik Douglas, could always be counted on. The tireless efforts of the Joint Graduate Group and Bioengineering Department staff,

especially Rebecca Pauling and SarahJane Taylor have smoothed over every administrative challenge.

In my initial development as a researcher, I thank my undergraduate professors Keith Gooch whose counsel during my graduate application process I only later realized was so critical, and David Pope whose love for teaching and passion for discovery provided me an example to follow.

I thank the wisdom of providence for guiding Tom Burnett, Dean Scrimgeour, Justin Hastings and I to a house on Peralta Street at the beginning of our graduate school journeys. We share a brotherhood that will continue to shape our discoveries and adventures, academic, athletic and otherwise. The support of what has become my adopted extended family at Berkeley Covenant Church, especially that of Bruce and Leanne Wydick, has been integral to keeping everything in balance during my studies. My parents, Mahdi and Claire Fawzi, have always encouraged me to do what I thought was right and follow my own path and my sister, Susie, has been an example of hard work and dedication.

Finally and most of all, I am thankful to God for the blessing of my wife, Alice. Her love and partnership through this period in our lives makes me all the more excited for the future.

This work is composed primarily of material reprinted from three published and one submitted paper, and as a body of work is equivalent to a standard dissertation.

Teresa Head-Gordon supervised the research and manuscript writing for all the chapters in this dissertation. The author performed the majority of data collection and analysis for all studies described, and co-wrote all manuscripts. The text of the abstract, introduction and conclusion are a reprint in part of material to be submitted to *Accounts of Chemical Research*. The text of chapter 2 is a reprint of the material as it appears in *Protein Science*. Victor Chubukov assisted with data collection and preliminary analysis, Louis Clark streamlined the code used for the study of protein aggregation, and Scott Brown provided additional analysis of data from a prior study. The text of chapter 3 is a reprint of material as it appears in the *Journal of Molecular Biology*. Yuka Okabe assisted in the collection of data and automation of data analysis, and Eng Hui Yap assisted in the development and parameterization of the new hydrogen-bonding model. The text of chapter 4 is a reprint of material as it appears in *Biophysical Journal*. Kevin Kohlstedt collected some of the early data and provided useful discussion throughout the study, Yuka Okabe assisted in the collection of data and automation of data analysis. The text of chapter 5 is a reprint of material submitted to the *Journal of the American Chemical Society*. The work involved in this chapter was a joint computational and experimental study supervised by both Teresa Head-Gordon and David Wemmer. Aaron Phillips collected and analyzed the experimental data, Jory Ruscio assisted in the collection and analysis of simulation data, and Michaeleen Doucleff performed the preliminary NMR characterization.

Abstract

Contrasting disease and non-disease protein aggregation by molecular simulations

by

Nicolas Lux Fawzi

Doctor of Philosophy in Bioengineering

University of California, San Francisco and University of California, Berkeley

Professor Teresa Head-Gordon, Chair

This work describes the development and application of computational models for the investigation of disease and non-disease protein aggregation. We demonstrate how the aggregate equilibrium, formation kinetics, and structural ensembles are influenced by the structural and folding properties of the monomer units. Using a coarse-grained model, we examine the influence of folding rates and mechanisms of non-disease proteins L and G on their aggregation structure and kinetics. We demonstrate that the number and spatial distribution of contacts in the denatured state correlate with aggregation rates and the identity of inter-protein contacts, and that fast forming intermediates may inhibit aggregation through the burial of “sticky” regions. To examine the driving forces underlying the transition from amorphous aggregates to cross- β ordered amyloid fibrils, we extend this physical model and investigate the aggregation of Alzheimer’s $A\beta_{1-40}$ peptide. We find that the critical nucleus, the highest free-energy species on the aggregation pathway, is composed of ten ordered peptides, the minimum number necessary to stabilize the interfilament contacts defining a fibril axis that enable fast

growth of a fibril. Once past the critical nucleus, the model fibril elongates by efficiently incorporating monomers at only one of two asymmetric ends, connecting local structure differences to biased elongation. Familial Alzheimer's Disease (FAD) mutations represented in the model alter the number of peptides necessary to form the critical nucleus as well as the fibril stability, suggesting a molecular mechanism for the spectrum of *in vitro* aggregation kinetics and morphologies associated with dramatically different FAD clinical outcomes. The region of these mutations is examined in detail through atomistic simulations and NMR analysis of the monomeric A β_{21-30} peptide. Our simulations reproduce relaxation times and ROESY cross-peaks to interpret NMR experimental population averages, but we find no evidence of majority folded structures of this fragment as previously reported. By combining coarse-grained and atomistic simulations with experimental observables, we describe the fast-forming, poorly-populated and disordered states that drive aggregation at a level of detail unattainable by experimental techniques alone, elucidating the link between atomic level properties and aggregation outcomes for the protein L and G and the Alzheimer's A β systems.

Contents

Contents	ix
List of Tables	x
List of Figures	xii
Chapter 1: Introduction	1
Chapter 2: Influence of denatured and intermediate states of folding on protein aggregation.....	7
2.1 Introduction	8
2.2 Results	12
2.3 Discussion and Conclusion	25
2.4 Methods.....	29
2.5 Acknowledgments.	35
Chapter 3: Determining the Critical Nucleus and Mechanism of Fibril Elongation of the Alzheimer's A β_{1-40} Peptide	36
3.1 Introduction	37
3.2 Results	46
3.3 Discussion and Conclusions.....	61
3.4 Methods.....	68
3.5 Acknowledgements.....	74
Chapter 4: Protofibril Assemblies of the Arctic, Dutch and Flemish Mutants of the Alzheimer's A β_{1-40} Peptide.....	76
4.1 Introduction	77
4.2 Methods.....	83
4.3 Results	90
4.4 Discussion	98
4.5 Conclusions	102
4.6 Acknowledgements.....	104
Chapter 5: Structure and dynamics of the A β_{21-30} peptide from the interplay of NMR experiments and molecular simulations.....	105
5.1 Introduction	106
5.2 Methods.....	111
5.3 Results	120
5.4 Discussion	140
5.5 Conclusion.....	142
5.6 Acknowledgements.....	144
Conclusion	145
Bibliography	147

List of Tables

Table 2.1: Parameters obtained from fits to aggregation kinetic data of protein L and G from this study. The data is fit to the equation: $A_0 \exp(-t/\tau_0) + (1-A_0) \exp(-t/\tau_1)$. We also include the kinetic parameters fit to the folding of proteins L and G as reported in Table 1 of [7] and an estimate of the timescale for the formation of protein G's intermediate from Figure 8 in [7]. The χ^2 values for these data indicate a good fit.	17
Table 3.1: Equilibrium populations of ordered fibrils, C_n and populations with free monomer, C_{n-1} , and calculated changes in free energy, ΔG , per unit monomer based on Equation 3.1.....	53
Table 3.2: Mapping 20-letter (20) amino acid code to coarse-grained four-letter (4) code	69
Table 3.3: Parameters for Dihedral Types.....	70
Table 5.1: ^{13}C NMR Spin Relaxation times T_1 and T_2 for non-glycine C_α positions from experiment and TIP4P-Ew, TIP3P and time scaled TIP4P-Ew, in milliseconds. Glu22 and Lys28 resonances overlapped such that T_1 and T_2 could not be independently measured. The relaxation times calculated from the overlapped peaks are indicated with an asterisk.	124
Table 5.2: Peak volume predictions of each water model, TIP3P and TIP4P-Ew for 900MHz H_2O . Peaks are labeled with the predicted volume based on r^{-6} average distances computed from a simulation ensemble combining structures from both independent simulations. The last column shows the experimental peak intensity	

(marked with a # if the intensity displayed in the H₂O column is from the ²H₂O experiment) or a “U” if some evidence of a peak is present, but the peak is so weak that no definite peak can be assigned. 133

Table 5.3: Merged peak volume predictions from simulations of each water model, TIP3P and TIP4P-Ew for 800MHz ²H₂O. See Table 5.2 caption for further detail..... 134

List of Figures

Figure 2.1: Ribbon drawing of the protein L (and protein G) model used in this study. Figure created by PyMOL[42].	11
Figure 2.2: Free energy projection of inter-chain contact number, χ_{inter} , vs. the number of intra-chain contacts, χ_{intra} for (a) protein L and (b) protein G at their respective folding temperature, T_f .	15
Figure 2.3: Fraction of unaggregated states $P_{unaggregated}$ as a function of unitless time t/τ for protein L (squares) and protein G (circles) at their respective folding temperature. (a) Based on the aggregation definition of $\chi_{inter}=38$ for protein G and $\chi_{inter}=28$ for protein L. The fit is shown as a solid line, and the best fit parameters are given in Table 2.1. (b) Based on the same definition of aggregation: $\chi_{inter}=25$ for both proteins. This shows that differences in definition of χ_{inter} for the two proteins used in (a) above is not biased toward slower aggregation rates for protein G.	18
Figure 2.4: Contact maps comparing native conformation to intermediate states and denatured states. Native state contacts (represented by the area lying within the black contours) compared to contacts that are present in at least 60% of the ensemble of denatured state structures (contoured in red) for (a) protein L and (b) protein G. (c) Contact map comparing native state (black) and contacts that are present across at least 60% of the intermediate state ensemble (red) for protein G.	20
Figure 2.5: Contact map comparing native state (black) and <i>intra-chain</i> contacts made in at least 60 % of the aggregated ensemble (green) for (a) protein L's fast aggregation	

pathway, (b) protein L's slow aggregation pathway, (c) protein G's fast aggregating pathway, and (d) protein G's slow aggregation pathway. For reference, contacts that are present in at least 60% of the denatured state ensemble of protein L are contoured in red in (a) and (b), and in at least 60% of the intermediate ensemble of protein G are contoured in red in (c) and (d).22

Figure 2.6: Contact map comparing native state (black) and *inter-chain* contacts made in at least 15 % (green) and 8% (blue) of the aggregated ensemble for (a) protein L's fast aggregation pathway, (b) protein L's slow aggregation pathway, (c) protein G's fast aggregating pathway, and (d) protein G's slow aggregation pathway, respectively. For reference, *intra-chain* contacts that are present in at least 60% of the aggregated ensemble of protein L are contoured in red in (a) and (b), and in at least 60% of the aggregated ensemble of protein G are contoured in red in (c) and (d).....24

Figure 2.7: Ribbon diagram of a snapshot for the aggregation simulation of protein G that illustrates both native intra-chain and inter-chain contacts made. Figure created by PyMOL[42].....25

Figure 2.8: Contact map comparing native state (black), the denatured state ensemble (red), and the transition state ensemble (blue) for (a) protein L and (b) protein G...27

Figure 3.1: Examples of starting structures for C_{2x} and C_{2z} symmetry forms. Protofibril seeds composed of 4, 6, 8, 10, and 12 (14-20 not shown) monomers were simulated for protofibril stability.39

Figure 3.2: Interdigitation of the N- and C- terminal β -strands to form side-chain contacts between different monomer chains introduces a stagger in the strand alignments. Side chain contacts between the C-termini of monomer i with the N-termini of monomer $i+1$ and $i+2$ (STAG(-2)) or between the N-termini of monomer i with the C-termini of monomer $i+1$ and $i+2$ (STAG(+2))[13]. Our model naturally relaxes to the STAG (-1) definition, although we show the STAG (+1) for completeness. 41

Figure 3.3: Free energy profile for the nucleation-polymerization reactions. Typical free energy (ΔG) profile and slope of ΔG ($\Delta\Delta G$) vs. number of chains in protofibril for fibril formation by a nucleation-dependent polymerization mechanism. At high number of chains, the protofibril is stable and free-energetically favorable, and the free-energy benefit to adding chains is constant, as seen in a constant slope of ΔG . Since the slope of ΔG is constant in this regime, the free-energy benefit to adding a chain or free-energy cost for removing a chain is the same as in an infinite fibril. As the number of chains decreases, the free-energy change for removing chains decreases, indicating that the fibril is approaching the number of chains in the critical nucleus. At the critical nucleus, the least free-energetically favorable species, the slope of ΔG is zero. (Typical ΔG data adapted from Ferrone[86]) 43

Figure 3.4: Effect of internal stagger on terminating ends of fibril. A schematic of 16 chain C_{2x} and C_{2z} fibrils are shown for internal staggers STAG (-1), STAG (+1) and mixed STAG(+1/-1).with N-terminal region colored in teal and C-terminal region colored in orange. (a) STAG (-1) C_{2x} has superimposable, symmetric ends. End A can be approximately superimposed on End B by a simple rotation of 180° about the

x-axis (hence C_{2x}). STAG (-1) C_{2z} has distinct, asymmetric ends. End A exposes the C-terminal β -strands, and End B exposes the N-terminal β -strands. Ends A and B of the C_{2z} fibril cannot be superimposed on End A by any rotation. (b) C_{2x} STAG (+1), like C_{2x} STAG(-1), has superimposable, symmetric ends. C_{2z} STAG (+1), like C_{2z} STAG (-1) above it, has distinct, asymmetric ends. (c) C_{2x} STAG (-1/+1) has the top peptide STAG (+1) and bottom peptide STAG (-1). Mixing staggers in C_{2x} desymmetrizes the C_{2x} ends. Mixing staggers in C_{2z} symmetrizes the C_{2z} ends so that each end has one subunit with an exposed N-terminal β -strand, and the other with an exposed C-terminal β -strand, unlike the two asymmetric ends in “pure” C_{2z} STAG (-1) or C_{2z} STAG (+1) models. 48

Figure 3.5: Time course for protofibril stability measured by χ_f . The metric χ_f measures the pair distances between the residues on both sides of the fibril, and thus is more sensitive to rotation of one subunit with respect to the other, and thus measures fibril disorder of the quaternary structure. The time course data averaged over all trajectories of C_{2z} fibrils for lengths 4 to 16 chains for fibril end B..... 50

Figure 3.6: Protofibril stability measured by $\langle\chi_f\rangle$ vs. number of chains. $\langle\chi_f\rangle$ is an average measure of the fibril order of the edge chains for stable quaternary structure for (a) C_{2x} form and (b) C_{2z} for the two ends of the protofibril: end A (black) and end B (red). Error bars are standard deviation. 51

Figure 3.7: Free energy profile for free monomer and protofibril equilibrium. The free energy vs number of ordered chains in an oligomer is plotted for C_{2x} (X, black) and C_{2z} forms. The free energy shows a clear maximum at 10 chains for C_{2z} and 10-12

chains for C_{2x} , indicating the region of the critical nucleus. A constant, negative slope at ~ 16 chains and above is indicative of reaching a stable fibril regime..... 53

Figure 3.8: Hydrophobic residue density vs number of chains. Hydrophobic density (number of hydrophobic residues per unit volume) vs number of chains for the C_{2x} and C_{2z} forms after initial equilibration. Error bars are standard deviation for the 24 structures created from the 40 chain equilibration runs. The hydrophobic density for C_{2z} is higher than C_{2x} for all oligomer sizes. 55

Figure 3.9: Monomer additions to protofibrils for C_{2x} and C_{2z} fibrils. (a) Fraction of trajectories resulting in partial parallel (black) and anti-parallel (red) additions to the N-terminal (\square) and C-terminal (\diamond) β -sheets. Error bars are standard deviation approximated from distributions with binary outcomes. (b) Ratio of partial parallel to anti-parallel additions to the N-terminal (\square) and C-terminal (\diamond) β -sheets. Error bars are 95% confidence interval for “relative risk” measure comparing binary outcomes. 59

Figure 3.10: Example addition to fibril seed by free peptide. A peptide (yellow) with a random initial configuration without contacts with the seed is shown with partial in-register parallel addition to both N- and C- terminal β -sheets of the fibril seed..... 60

Figure 3.11: Comparing structural stability of example structures of varying oligomer length. Representative oligomer structures after 5000τ constant temperature simulations depicting greater structural stability as number of chains increases. (a) 4 chain simulation shows complete loss of fibril structure (b) 10 chain simulation shows that although a significant fraction of intermolecular β -sheet is retained, the

fibril subunits rotate with respect to one another, leading to disorder and loss of contacts in the edge chains (c) 16 chain simulations shows retention of fibril order, and a clear fibril axis..... 63

Figure 4.1: Ideal cross-section of agitated fibril morphology. The two monomer cross-section of the bead model with C_{2z} symmetry based on SS-NMR data [17, 31-33] after equilibration within a large WT fibril (40 chain). Ala 21, site of the Flemish mutation, is pictured in orange on one monomer. Glu 22, site of Arctic and Dutch mutations, is pictured in red..... 81

Figure 4.2: Population histograms with respect to protofilament order (P_f) for 4, 8, and 20 chains. The histograms emphasize that protofilament order increases for the FAD mutants as: Flemish (red) < Arctic (green) < WT (black) < Dutch (blue), at any oligomer size. 92

Figure 4.3: Representative protofibril structure of the Arctic and Flemish mutants. While both FAD mutants show disorder in the turn region, the Arctic mutant (green) retains much better β -strand structure over the whole cross-section at the end of the 5000τ trajectories, while the Flemish mutant (red) has almost lost a monomer after the same amount of time..... 93

Figure 4.4: Representative protofilament structure of the Dutch mutant compared to the WT sequence. Already after only initial equilibration from the model build, the Dutch mutant (blue) shows a greater twist of the intermolecular β -sheet down the protofilaments with respect to WT (black). 94

Figure 4.5: Population histograms with respect to protofibril order (χ_f) for 4, 8, and 20 chains. While no protofibrils are present for the 4 chain assemblies for any sequences, the level of protofibril structure is different among the sequences at 8 chains to yield the following trend for FAD mutants: Flemish (red) < Dutch (aqua) < WT (black) < Arctic (green). Even with the addition of the Dutch polymorph as a reference state (blue), there is slightly more disorder for the Dutch with respect to WT.....95

Figure 4.6: Representative protofibril structure of the Dutch mutant with respect to WT. A comparison of the Dutch polymorph (blue) with respect to the agitated fibril morphology [17, 31-33] favored by the WT sequence (black) at the end of the 5000 τ trajectories. The yellow spheres represent amino acid 33 on each monomer chain, which shows how it is displaced due to a register shift of the C-terminal β -strands at the interface under the Dutch polymorph.....98

Figure 4.7: (a) Free energy profile for free monomer and protofibril (χ_f) equilibrium for the WT and FAD mutants. The free energy shows a maximum corresponding to the critical nucleus size of 6 chains for Arctic (green), 10 chains for WT (black), and no preferred order for either Dutch (aqua) or Flemish (red) mutants. The constant, negative slope beyond the critical nucleus is indicative of reaching a stable fibril regime in which the equilibrium shifts decidedly away from the monomer form. When the alternative polymorph for the Dutch mutant (Figure 4.6) is added as a reference for fibril order (blue), there is now a critical nucleus of 10 chains for the E22Q mutant but with a larger free energy barrier and shallow slope indicating a

slower approach to protofibril order for the Dutch mutant. (b) Free energy profile for free monomer and protofilament (P_f) equilibrium for the WT and FAD mutants. The free energy shows a maximum corresponding to the critical nucleus size of 6 chains for Dutch (blue), 6-8 chains for WT (black), and no preferred order for either Arctic (green) or Flemish (red) mutants. The constant, negative slope beyond the critical nucleus is indicative of reaching a stable regime in which the equilibrium favors the protofilament form, which is more strongly evident for Dutch over WT..... 101

Figure 5.1: ^{13}C and ^1H chemical shifts from experiment and simulation for $A\beta_{21-30}$.

Experimental NMR shifts are calculated as difference from the peptide measurements and tabulated random coil values. For ^{13}C we used the tabulated 25°C random coil shifts, while the reference amide proton shifts for each residue are adjusted to take into account the difference in temperature (25°C vs 10°C). The predicted chemical shifts are SHIFTS calculations averaged over the full ensemble after subtracting SHIFTS calculations averaged over the unstructured subpopulations (as defined by DSSP to represent our calculated reference random coil population)..... 122

Figure 5.2: Scalar coupling constant $^3J_{\text{HNH}\alpha}$ calculated from the self consistently

converged individual replica exchange ensembles for different empirical force fields and compared to experimentally determined coupling constants. Error bars are experimental uncertainty for NMR values as well as simulated standard deviations calculated for converged trajectories split into three sections. The scalar coupling

- constants for the glycines are added for the redundant C_{α} hydrogens in the simulation to connect to experiment, which is unable to distinguish them. 123
- Figure 5.3: Normalized average vector time correlation function for Val24 C_{α} to H_{α} position for constant energy trajectories solvated with TIP4P-Ew and TIP3P. The TIP3P model (red) shows a dramatically faster decay for all vector time correlation functions relative to TIP4P-Ew, and we present this as an example. 125
- Figure 5.4: Fingerprint region of ROESY spectrum in H_2O of $A\beta_{21-30}$ demonstrating the cross peak interpreted by Lazo et al. and Grant et al. as H_{α} Glu22 with HN Ala30 in their 500MHz experiments is clearly resolved as only H_{α} Lys28 to HN Ala30 in our 900MHz experiment. 126
- Figure 5.5: HB3 Lys28 has a nearly overlapping chemical shift with HB2 Glu22, potentially leading to cross peak misassignment in previous studies at lower field. 127
- Figure 5.6: Representative structure from restraint energy minimized ensemble, simultaneously applying all observed ROE interactions. 136
- Figure 5.7: Representative structures showing A) the most populated N-terminal hydrogen bond (dotted red) and B) the hydrogen bonds and electrostatic interactions (dotted red) stabilizing the type I β -turn (cyan) centered at Val24 and Gly25. 138
- Figure 5.8: Asp23 to Lys28 salt-bridge (shown by red dotted line between aqua amino acids) found in 7% of the TIP4P-Ew ensemble, and Ala21 HB to Val24 HG (both in light blue) van der Waals contacts which give rise to the strongest observed ROE interaction. 140

Introduction

Evolution has guided the design of the amino acid sequence of globular proteins such that proteins reliably assume a specific functional native state. This native state carries out a particular function by precisely bringing together residues to form, for example, catalytic sites in enzymes or specific binding site architectures for protein complexation and signaling. The ability of the protein to find and maintain the native state is therefore dependent on an amino acid sequence that gives rise to a structural ensemble that is thermodynamically stable at the physiological pressures and temperatures and solution conditions in the normal cellular or extracellular environment of the protein. Destabilizing sequence mutations[1], chemical cleavage or modification[2], or changes in protein concentration and solution environment of the protein[3] can shift the equilibrium from the native state in favor of aggregates, misfolded states with inter-protein contacts with other monomers. These protein aggregates range from unstructured amorphous collections of misfolded proteins often found in inclusion bodies when proteins are overexpressed in bacterial hosts[4], to fibrils with regular and repeating structure associated with a number of human diseases[5].

In order to develop strategies to change aggregation outcomes in protein expression and disease, it is therefore of critical importance to develop an understanding of the factors that protect normal, non-disease proteins from aggregation, as well as the driving forces for forming disease protein aggregates that nucleate more ordered and thermodynamically stable structure. What common features of the sequence, structural or folding properties distinguish non-disease proteins from disease proteins vulnerable to

pathological aggregation and the formation of structured aggregates at physiological conditions? In this work, we study in detail how the amino acid sequence and folding properties of the monomeric peptide and protein units affect structural stability of aggregates, aggregation rates, and mechanisms for aggregate assembly. Though the gross morphology of aggregates can be investigated with current experimental techniques[5], early events and small oligomers that determine aggregation propensities, pathways and rates are too highly disordered to probe by most current biochemical or protein structural techniques such as fiber and crystal x-ray diffraction, nuclear magnetic resonance (NMR) spectroscopy, and electron and atomic force microscopy. Although experimental studies are rapidly developing more sophisticated approaches[6], only molecular simulations currently offer the promise of directly observing the entire aggregation process in molecular detail. By employing both all-atom and coarse-grained models, validated against appropriate experimental observables, we characterize the aggregation structures, pathways and precursors at a level of detail and confidence not possible with either experiment alone.

We use a thoroughly validated coarse-grained model of the homologous pair of proteins L and G [7] to examine the sequence and structural factors that favor normal folding over aggregation. We have previously characterized the folding characteristics of these proteins, including their folding rates [8] as well as the structure of their denatured and intermediate states along the folding pathway[7]. We begin in Chapter 2 by characterizing *in silico* the aggregation of these non-disease proteins at high concentration, akin to the environment of overexpressed proteins where many of the

mechanisms that protect non-disease proteins from pathological aggregation breakdown[4]. By monitoring their aggregation rates and structures, we observe that rapid intermediate formation and spatially distributed contacts in the denatured state provide protein G more protection from aggregation than protein L. We characterize structures of the aggregates formed and observe how the different folding rates and mechanisms of each protein results in different aggregate structures. We conclude this section by noting that these oligomeric aggregates we observe in this study may represent the nuclei for larger aggregates. This insight extends not only to large amorphous inclusion bodies but also to the seeds of ordered fibrillar aggregates known as amyloid fibrils, since most non-disease proteins, including proteins L and G, can form amyloid fibrils under conditions which destabilize the native state[9, 10].

Amyloid fibrils are ordered fibrillar structures that are micrometers in length and five to twenty nanometers in diameter and are associated with a number of human diseases including Alzheimer's Disease (AD)[5]. Although the presence of amyloid fibrils clearly correlates with disease, recent studies have raised the possibility that the main toxic species in amyloid diseases may not be the fibrils themselves, but rather the pre-fibrillar aggregates [11, 12], underscoring the need to develop a more detailed understanding of the aggregation process, not simply the structure of the mature fibril. Due to their transient, disordered, and aggregation-prone nature, experimentally trapping and characterizing these critical early oligomers of amyloid peptides has proved a significant challenge.

Following from the first section of our study where we examined the factors that protect non-disease proteins from aggregation, we turn our attention now to the properties of the monomers, small early oligomers and fibril state of the A β peptides associated with the neurodegenerative Alzheimer's Disease. Although an atomic model for the fibril form has been confirmed based on solid-state NMR constraints [13], the structural ensemble of the potentially toxic early oligomers and the critical nucleus (the purported transition state from unstructured oligomers to the fibril form), have not yet been determined.

In chapter 3, we extend the applicability of the coarse grained model to capture the cooperativity of formation of β -sheet structure and use it to examine the transition from disordered oligomers to ordered fibrils as well as the elongation of fibrils of the Alzheimer's A β_{1-40} peptide. The free energy barrier is a function of the number of peptides needed to order not just the extended structures that form the cross- β structure, but also the contacts along the hydrophobic interface between the two subunits that make up the full fibril. We find that the critical nucleus for WT A β_{1-40} is composed of ten peptide oligomers with ordered cross- β structure protofilaments, but the interface between protofilaments do not align in such a way to define order along the fibril axis. Beyond the critical nucleus corresponding to the stabilization of a well-defined fibril axis, the fibril structures have distinct ends, which we show elongate at different rates, leading to our predictions of biased fibril elongation.

In chapter 4, we use the same model to examine three Familial AD (FAD) mutants of the A β peptide. Referred to as Arctic (E22G), Dutch (E22Q), and Flemish (A21G) for the region from which the kindred originated [2], the FAD mutant sequences

cause dramatically different disease progression, dementia symptom severity, and cerebral plaque deposition site [14-18]. These clinical differences are most likely due in part to the biophysical differences in the kinetics of formation of fibrils and protofibrils, partially ordered oligomers observed in vitro. The Dutch mutation results in faster fibrillization, the Arctic mutation in faster formation of partially ordered oligomer, and the Flemish mutation in much slower fibrillization [19-22]. Representing these mutations in our coarse-grained model, we demonstrate how the local changes in side chain chemistry give rise to global biophysical changes in the stability and critical nuclei of the amyloid fibril form. The Dutch mutation increases the free energy benefit for the extension of the fibril and the Flemish mutation dramatically decreases the benefit, suggesting a mechanism for the observed accelerated or inhibited fibril formation, respectively. The Arctic mutation results in a shift in the critical nucleus to smaller oligomers, stabilizing protofibril formation, while at the same time decreasing order in the fibril state.

Finally, in chapter 5, we return to the question of whether the folding properties of the single chain drives aggregation behavior by examining in detail the equilibrium structure of the aggregation resistant A β_{21-30} subpeptide, previously thought to comprise the “folding nucleus” of the full length peptide [23, 24]. To accomplish this characterization, we employ fully atomic molecular dynamics simulations using recent and critical updates to the empirical forcefield parameters [25, 26] combined with the replica exchange accelerated sampling technique [27] to calculate the average structure and dynamics of the peptide. We find excellent agreement with our NMR experiments, validating the simulations in which we find a diverse ensemble of conformations with a

minority population of a β -turn centered at Val24 and Gly25. This finding is in contrast to the conclusions of previously published NMR studies, which we demonstrate were likely the result of errors in assignment leading to an inappropriate model.

Introduction and Conclusion reproduced with permission from Accounts of Chemical Research, material to be submitted for publication. Unpublished work copyright 2007 American Chemical Society.

Chapter 2: Influence of denatured and intermediate states of folding on protein aggregation

Reproduced with permission from *Protein Science*. 2005 Apr;14(4):993-1003. Copyright 2005 Protein Society.

Nicolas J. Fawzi¹, Victor Chubukov¹, Louis A. Clark³, Scott Brown⁴, Teresa Head-Gordon^{1,2}

¹UCSF/UCB Joint Graduate Group in Bioengineering and ²Department of Bioengineering, University of California at Berkeley, Berkeley, CA 94720

³Current address: Biogen Idec, 14 Cambridge Center, Cambridge, Massachusetts 02142

⁴Current address: Abbott Laboratories, 1401 Sheridan Road, North Chicago, Illinois 60064-400

Abstract

We simulate the aggregation thermodynamics and kinetics of proteins L and G, each of which self-assembles through distinctly different folding mechanisms. We find that the aggregation kinetics of both proteins at an experimentally relevant concentration exhibit both fast and slow aggregation pathways, although a greater proportion of protein G aggregation events are slow relative to that found for protein L. The kinetic differences in aggregation pathways are correlated with the amount and distribution of intra-chain contacts formed in the denatured state ensemble (DSE), or an intermediate state ensemble (ISE) if it exists, as well as the folding timescales of the two proteins. Protein G aggregates more slowly than L due to its rapidly formed folding intermediate, which exhibits native intra-chain contacts spread across the protein, suggesting that certain early

folding intermediates may be evolutionary selected for their protective role against unwanted aggregation. Protein L shows more localized native structure in the DSE with timescales of folding that are commensurate with the aggregation timescale, leaving it vulnerable to domain swapping or non-native interactions with other chains that increase the aggregation rate. Folding experiments that characterize the structural signatures of the DSE, ISE or the transition state ensemble under non-aggregating conditions, should be able to predict where inter-chain contacts will be made in the aggregate, and to correlate slower aggregation rates for proteins with contacts that are dispersed across the fold. Because protein L and G can both form amyloid fibrils, this work also provides mechanistic and structural insight into the formation of prefibrillar species.

2.1 Introduction

In order to perform their biological function, proteins adopt a three-dimensional structure that represents a global or very low-lying minimum on their free-energy surface. Through molecular events not fully understood, proteins can sacrifice these stabilizing intra-chain contacts in favor of configurations that promote inter-molecular interactions leading to the formation of aggregates. These aggregates range from amorphous structures without order to highly structured fibrils, each arising by distinct aggregation mechanisms. The resulting structure of protein aggregates and the kinetics of their formation will depend on protein sequence, protein concentration and solution conditions.

For example, aggregates can take on a common morphology of un-branched fibrils that are several micrometers in length, ~ 10 nm in diameter, and rich in β -sheets orthogonal to the fibril axis [5]. These aggregates, termed amyloid fibrils, are believed to

be responsible for a number of diseases, including Alzheimer's and Huntington's disease [5]. The first experimentally identifiable nucleating species are known as prefibrillar aggregates which resemble small bead structures, and these prefibrillar species then go on to further organize into "protofilaments", which are thought to later organize into the mature amyloid fibrils. It is also noteworthy that many non-disease proteins can be induced to form amyloid fibrils. Alternatively, when engineered proteins are expressed in a bacterial host for industrial production, the product often accumulates in the form of inclusion bodies and must be solubilized and subsequently refolded to the native state while avoiding aggregation [4]. The morphologies of inclusion body aggregates are less structurally distinct, and this seeming non-specificity makes it unclear whether inclusion body formation shares common molecular origins of aggregation as that of amyloid formation.

Increased aggregation propensity has been noted for proteins that fold through both obligatory and non-obligatory kinetic intermediates [28] or through molten globule states [29]. This is attributed to attraction between inter-chain hydrophobic patches that resemble the folded monomer intrachain contacts, and promote further interchain stabilization upon aggregation. It is thought that random coils are less susceptible to this association because of the reduction in native-like hydrophobic patterns [30, 31]. However, it has also been shown that there is competition between the refolding of protein monomers with the formation of transient oligomeric protein aggregates, that instead derive from association of random coil states of the protein [32]. While the aggregates are short-lived and do dissociate so that individual chains fold, their folding

rate is fractionally slower than the normal folding reaction [32]. Chiti and co-workers were able to show that stabilizing local elements of secondary structure in the denatured state ensemble of the small α/β protein AcP prevented fibril formation [33].

This work examines the question of aggregation kinetics and mechanism using simulation of coarse-grained models of two α/β proteins, Ig-binding proteins L and G (Figure 2.1). Proteins L and G make excellent targets for theoretical study of aggregation since their folding characteristics have been extensively studied by experiment [34-41]. Experimental evidence indicates that protein L folds two-state through a transition state ensemble involving a native-like β -hairpin 1. Protein G on the other hand, folds through an early intermediate, followed by a rate-limiting step that involves formation of β -hairpin 2. Our most recent theoretical studies also differentiate the folding mechanism of proteins L and G as seen experimentally [7]. We find that their folding is consistent with a nucleation-condensation mechanism, each of which is described as helix-assisted β -1 and β -2 hairpin formation, respectively [7]. We determine that protein G exhibits an early intermediate which draws together misaligned secondary structure elements that are stabilized by hydrophobic core contacts involving the third β -strand, and the later transition state ensemble (TSE) corrects the strand alignment of these same secondary structure elements [7]. The kinetic data for protein G folding was fit to a two-step first order reversible reaction, proving that protein G folding involves an on-pathway early intermediate, and should be populated and therefore observable by experiment [7].

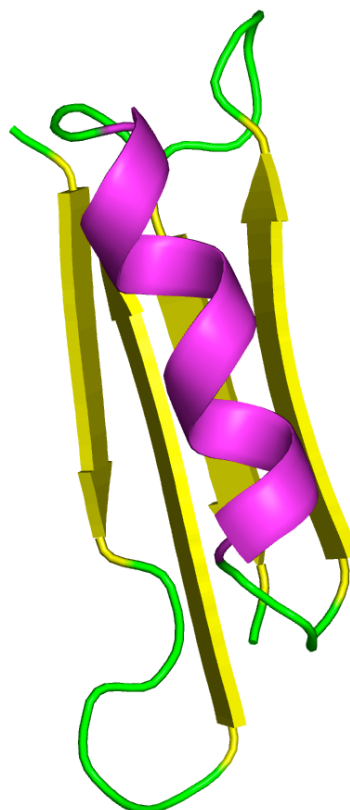


Figure 2.1: Ribbon drawing of the protein L (and protein G) model used in this study. Figure created by PyMOL[42].

The purpose of our study is to demonstrate that the different folding characteristics of protein L and G, both which form the same native topology, explain their different rates of aggregation. Furthermore, given the concentration and temperature used in our study corresponding to partially denaturing conditions, and the fact that protein L and protein G form amyloid fibrils, we would argue that these simulations are most directly relevant to the earliest aggregation events involving prefibrillar formation. We show that proteins L and G provide a good contrast for understanding features of aggregation that arise from different mechanisms of folding and/or due to protein folding

intermediates, while controlling for size, topology, and stability. We find that protein G aggregates more slowly than protein L due to its rapidly formed folding intermediate, which exhibits native intra-chain contacts spread across the protein. Protein L shows more localized native structure in the denatured state ensemble (DSE) with timescales of folding that are commensurate with the aggregation timescale, leaving it vulnerable to domain swapping or non-native interactions with other chains that increase the aggregation rate.

This computational study suggests that experiments that can characterize the structural signatures of the DSE (or transition state ensemble) under non-aggregating conditions, should be able to predict where inter-chain contacts will be made in the aggregate, and to predict slower aggregation rates for proteins with contacts that are dispersed across the protein fold. A corollary of this work is that early intermediates in folding may be evolutionary selected for their protective role against unwanted aggregation, and thus could be useful to employ in reengineered sequences to slow aggregation and increase refolding yield. Finally, given that protein L and G can both form amyloid fibrils under certain solution conditions [43, 44], this work also provides mechanistic and structural insight into the formation of the earliest prefibrillar species.

2.2 Results

All aggregation simulations we report for protein L and G were done at the midpoint of their temperature denaturant curve. This corresponds to a folding temperature for protein L of $T_f=0.42$, while the aggregation simulations for protein G were done at its folding temperature of $T_f=0.41$, as determined in previous work [7]. Aggregation

simulations performed at the proteins folding temperature correspond to partially denaturing conditions typically used by experimentalists to promote aggregation [45]. The use of the two (but only slightly different) temperatures for the aggregation studies allows us to remove the trivial effect of greater or lesser stability of the native state of one sequence as a factor in the comparison of their aggregation rates. By simulating aggregation at each of their folding temperatures, the native and unfolded states are of equal stability.

Because protein aggregation is a second order or higher reaction process, aggregation rates will be concentration dependent. All simulations were performed with 3 identical chains, of either the L or G sequence, in a periodically replicated box of length 32 distance units, set at maximal distances apart in their unfolded state. A box length of 32 units with 3 chains corresponds to a protein solution concentration of about 20 mg/mL (assuming a length scale of a bead-bead distance 3.8Å). This concentration is on the same order of magnitude as 5mg/mL for experimental aggregation kinetics studies performed by Dobson and coworkers [46] for SH3, another small protein. We also performed comparable simulations at a higher concentration (~80mg/ml) and determined the same qualitative features of the mechanism of aggregation (NF, VC, TH-G, unpublished observations) we uncover and discuss more fully below.

Instead of defining an aggregation event through a defined oligomer state, we chose to define it in terms of a critical number of inter-chain contacts, χ_{inter} , and number of intra-chain contacts, χ_{intra} (see Methods). This had the benefit, from our view, of no a priori assumptions about aggregation structure. We quantified these contact numbers by

evaluating free energy surfaces as projected onto χ_{inter} and χ_{intra} for protein L and protein G at their respective folding temperatures (Figure 2.2). The free energy projection shows a free energy minimum corresponding to aggregated states for each protein, and while their χ_{intra} position is the same, they differ in their position with respect to the χ_{inter} reaction coordinate. Based on these free energy projections, we use $\chi_{inter}=28$ for protein L and $\chi_{inter}=38$ for protein G as the measure for when three protein chains have aggregated. Using these definitions, aggregates of L and G are comprised of both dimers and trimers.

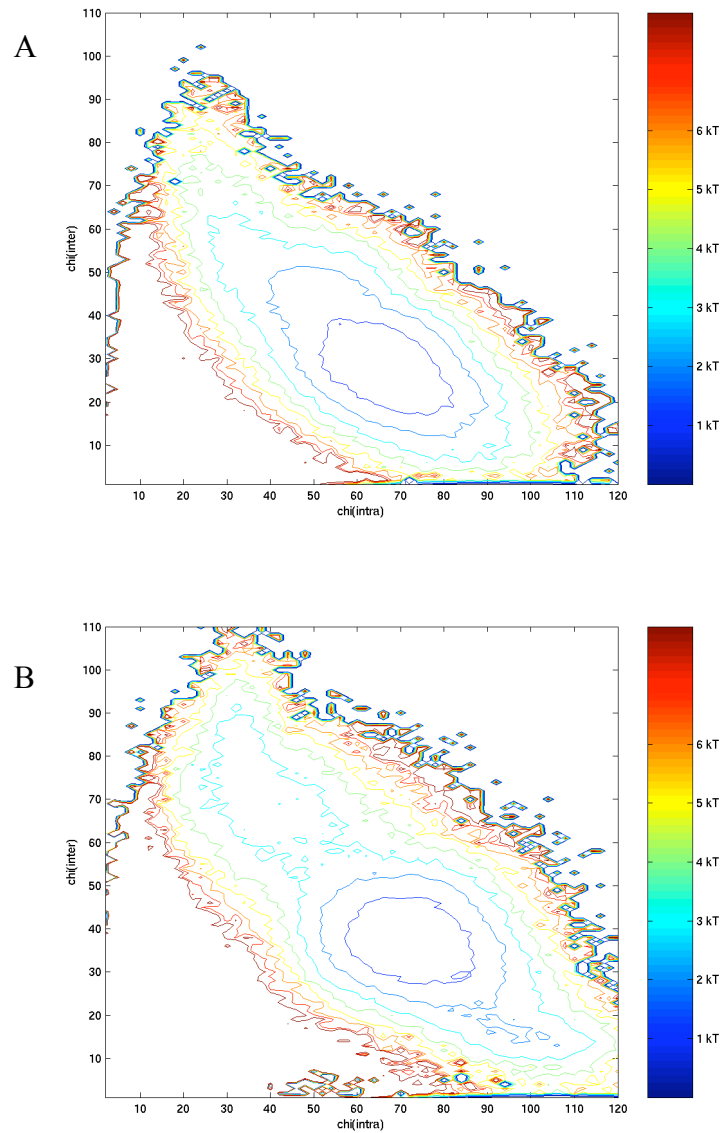


Figure 2.2: Free energy projection of inter-chain contact number, χ_{inter} , vs. the number of intra-chain contacts, χ_{intra} for (a) protein L and (b) protein G at their respective folding temperature, T_f .

In Figure 2.3a we plot the unaggregated population, $P_{unaggregated}$, vs. unitless time t/τ (where τ is the Langevin timestep), for protein L and protein G at their respective folding temperature. The kinetic data for both proteins is fit by a double exponential (parameters shown in Table 2.1). It is clear from the figure and parameter fit that protein

G aggregates more slowly than protein L. Just for clarification, Figure 2.3b compares the kinetic profiles for both proteins at $\chi_{inter}=25$, and we can see that protein G still aggregates more slowly, so that the differences in definition of χ_{inter} for the two proteins is not biased toward slower aggregation rates for protein G. Therefore, we focus on the aggregation kinetics using the χ_{inter} definition based on the free energy projections for protein L and protein G, which corresponds to the minimum in each basin. In Table 2.1 we also report timescales of folding for protein L and for protein G, including an estimate for the timescale for forming the intermediate, which will be important for later analysis. For protein L and the fast aggregation pathway for protein G, the timescales for folding are comparable to the aggregation timescale, whereas the protein G intermediate forms on timescales that are an order of magnitude faster than the fastest timescale for early aggregation (Table 2.1).

Aggregation Rates						
	T	A_0	$1-A_0$	τ_0	τ_1	$\chi^2 / 10^{-2}$
L	0.42	0.43	0.57	2175	33768	5.0
G	0.41	0.27	0.73	3554	42750	2.0
Folding Rates						
L	0.42	1.0	0.0	15700	0	0.034
G	0.41	0.81	0.19	13700	46400	0.035
G_{intermediate}	0.41	0.5	0.0	600	0	

Table 2.1: Parameters obtained from fits to aggregation kinetic data of protein L and G from this study. The data is fit to the equation: $A_0 \exp(-t/\tau_0) + (1-A_0) \exp(-t/\tau_1)$. We also include the kinetic parameters fit to the folding of proteins L and G as reported in Table 1 of [7] and an estimate of the timescale for the formation of protein G's intermediate from Figure 8 in [7]. The χ^2 values for these data indicate a good fit.

From the fit to the aggregation data, we find that a greater proportion of protein G chains (73%) than protein L chains (57%) aggregate through their slow aggregation pathways. From these kinetic fits, we can separate the raw aggregation trajectories into two sub-populations- fast and slow aggregation- for each protein (see Methods). We then can analyze these kinetic sub-populations for proteins L and G by evaluating contact maps of the native state, intermediate state (if it exists) and denatured state ensembles of individual folding protein chains, as well as contact maps for the intra-chain and inter-chain aggregated ensembles, to explain differences in aggregation rates. The simulation and numerical procedures for obtaining these data are described in the Methods section, and in previous work [7].

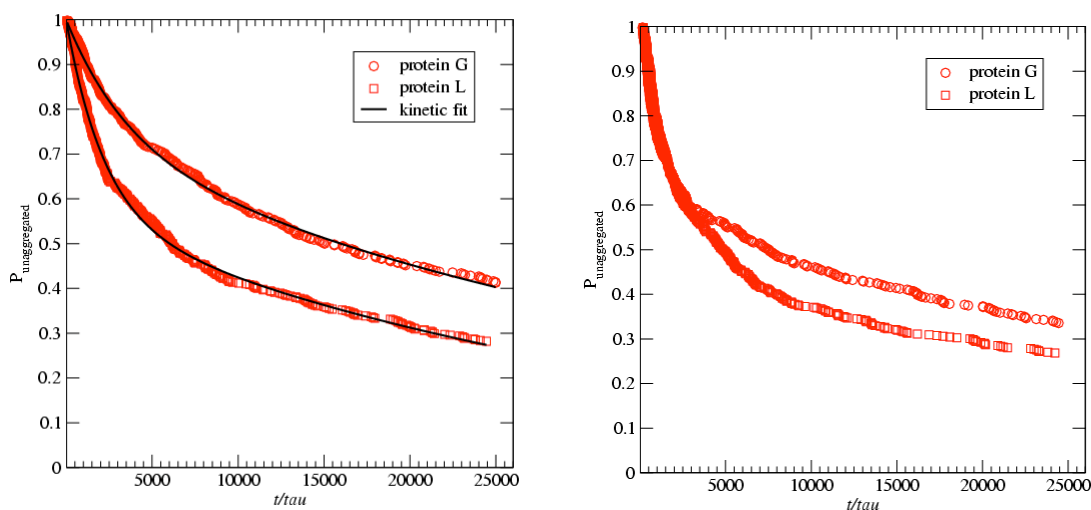


Figure 2.3: Fraction of unaggregated states $P_{unaggregated}$ as a function of unitless time t/τ for protein L (squares) and protein G (circles) at their respective folding temperature. (a) Based on the aggregation definition of $\chi_{inter}=38$ for protein G and $\chi_{inter}=28$ for protein L. The fit is shown as a solid line, and the best fit parameters are given in Table 2.1. (b) Based on the same definition of aggregation: $\chi_{inter}=25$ for both proteins. This shows that differences in definition of χ_{inter} for the two proteins used in (a) above is not biased toward slower aggregation rates for protein G.

The contact maps presented in Figures 2.4 through 2.6 illustrate which areas of the protein chain are in contact in the various states examined. The secondary structural arrangement of protein L and G (β strand 1, β strand 2, helix, β strand 3, β strand 4) are presented along the bottom and side of each map. Diagonal arrangements of contacts going up and to the right in the β regions correspond to a parallel β sheet, down and to the right correspond to anti-parallel β sheet. For all contact maps, a contact is formed if two beads are within 2.5 distance units (corresponding to about 9\AA), roughly the center of mass distance between sidechains. In all Figures, reference lines indicating native state contacts of folded chains are given in black. Unlike contacts in the native state that are always formed, contacts in the DSE, ISE and aggregated states that are significant are formed with a certain probability. Therefore, these states are contoured at various percentage levels described below to bring out significant contacts. For example, a contour level of 60% was chose for the intrachain contacts, to bring out contacts that are formed by 60% of the chains, more than half the chains in the aggregates. The frequency of the most populated interchain contacts are lower than intrachain contacts, due to greater number of possibilities for a particular contact to form involving three chains. Trivial contacts between neighboring residues found along the main diagonal are ignored in intrachain contours (native state, DSE, ISE, TSE, and aggregated intrachain contacts).

Interchain contours have no such trivial contacts and the area along the main diagonal is treated no differently from contacts in other regions. Interchain contour levels for protein G are adjusted by a factor of $(38/28)$ to correct for the increased number of contacts by definition included in the protein G aggregates.

Figures 2.4a and 2.4b provide a comparison of the DSE of protein L and protein G, respectively. In Figure 2.4c we display the intermediate ensemble for the slow pathway of protein G at the same contour level, which in previous work we have shown is characterized as an assembly of misaligned β -strands which are corrected in the later TSE [7]. These figures show that *native* contacts made in the DSE of protein L are more localized relative to that exhibited in the DSE of protein G, which exhibits stable native structural elements dispersed over the entire protein chain. In addition, there is a greater population of *non-native* elements in the DSE of protein L relative to DSE of protein G, while the ISE of protein G is the most native like. The non-native element of protein L involves parallel association of β -strands 2 and 4, but does contribute to a greater delocalization of intra-chain contacts. As we will see, the structural signatures of the DSE of each protein, and the ISE for protein G and timescale for its formation, provide complete insight into the aggregation pathways and kinetics.

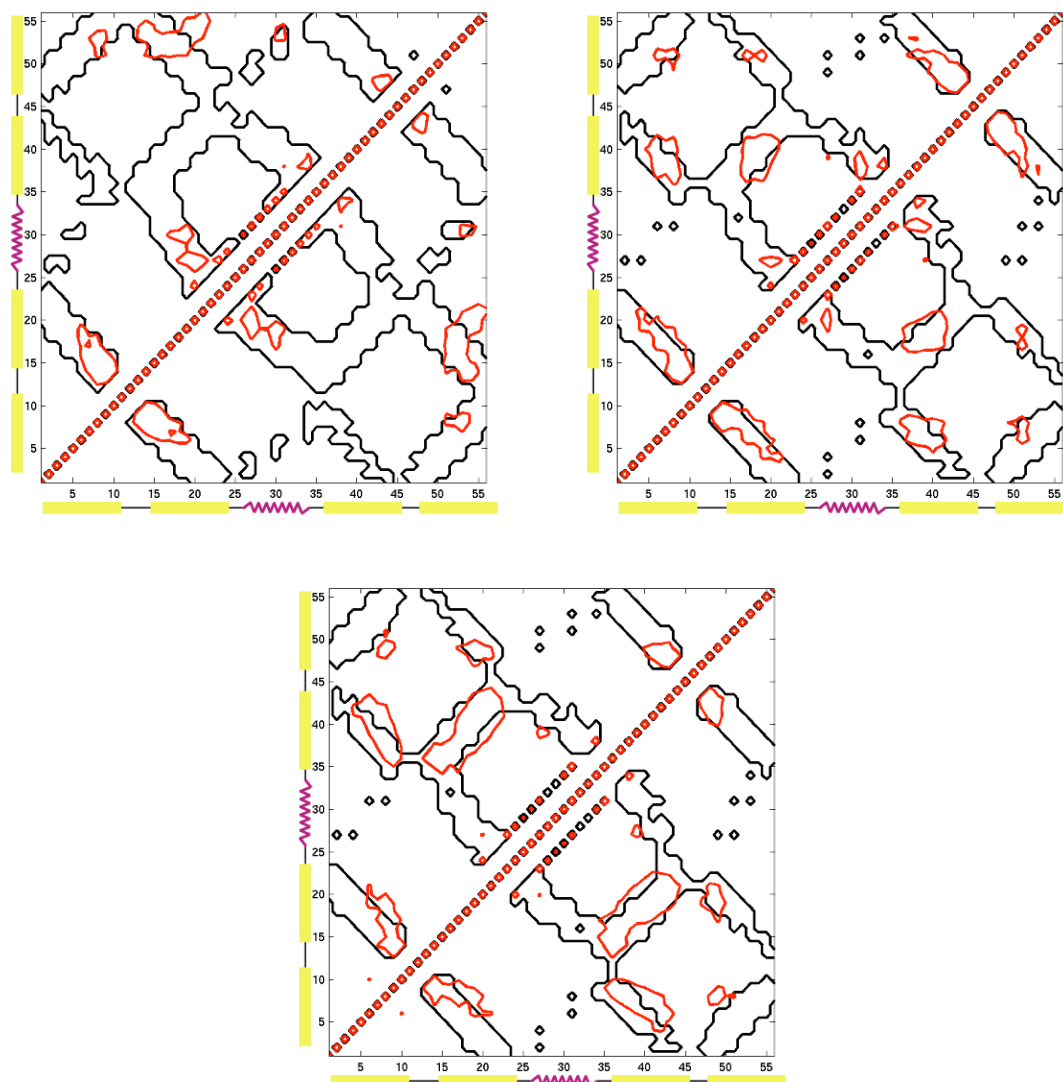


Figure 2.4: Contact maps comparing native conformation to intermediate states and denatured states. Native state contacts (represented by the area lying within the black contours) compared to contacts that are present in at least 60% of the ensemble of denatured state structures (contoured in red) for (a) protein L and (b) protein G. (c) Contact map comparing native state (black) and contacts that are present across at least 60% of the intermediate state ensemble (red) for protein G.

In Figure 2.5 we display the *intra*-chain contacts made by 60% of the population (i.e. more than one and a half full chains per three chain aggregate) in the aggregated

ensemble against the native and DSE reference for the fast and slow aggregation pathways, for protein L (Figures 2.5a and 2.5b, respectively) and protein G (Figures 2.5c and 2.5d, respectively). For each protein it is evident that the intra-chain contacts of the aggregated ensemble resembles contacts formed in the denatured state ensemble. The fast aggregation pathway for protein L (Figure 2.5a) protects only the localized first β -hairpin region, consistent with its folding pathway, but which leaves a majority of the residues vulnerable to entanglement with other chains. The aggregation is slowed down by chains that exploit both a more extensively formed first β -hairpin and contacts more greatly dispersed across the fold arising from association of β -strands 2 and 4 in the DSE (Figure 2.5b). Baker and coworkers' experiments confirm the localized structure in hairpin 1 of denatured state of protein L [47], and, although no long range associations like that between strands 2 and 4 were detected, the authors note that their use of 2M guanidine denaturant might disrupt long range structure. The fast aggregation pathway for protein G protects some of the first β -hairpin and more extensively the second β -hairpin region, consistent with the folding pathway of protein G, but these two regions are still relatively localized, i.e. they do not provide sufficient pinning sites throughout the fold (Figure 2.5c). By contrast the slowest aggregation pathway for protein G has a more extensive network of stabilizing native contacts across the protein, more consistent with the ISE that protects the sticky third β -strand (Figure 2.5d).

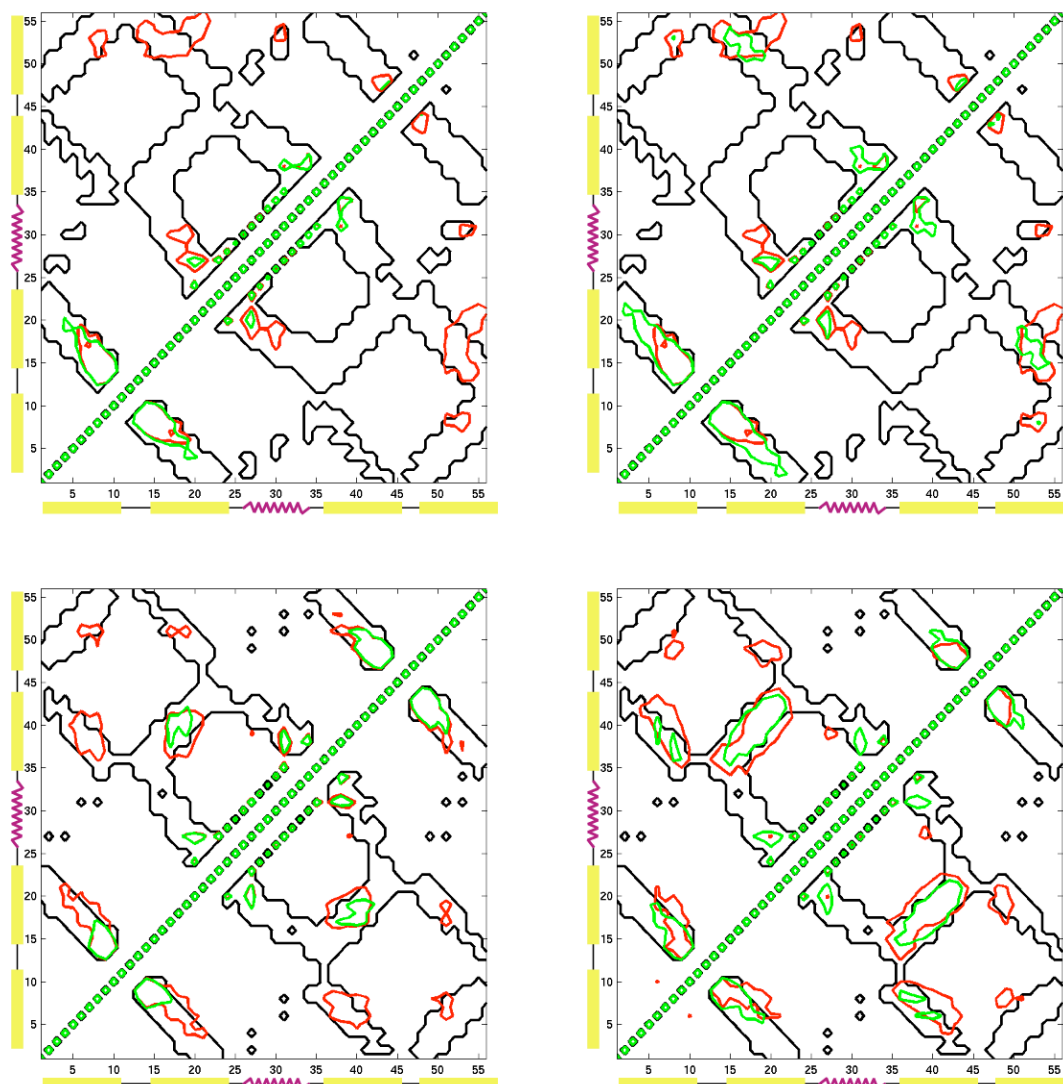


Figure 2.5: Contact map comparing native state (black) and *intra-chain* contacts made in at least 60 % of the aggregated ensemble (green) for (a) protein L's fast aggregation pathway, (b) protein L's slow aggregation pathway, (c) protein G's fast aggregating pathway, and (d) protein G's slow aggregation pathway. For reference, contacts that are present in at least 60% of the denatured state ensemble of protein L are contoured in red in (a) and (b), and in at least 60% of the intermediate ensemble of protein G are contoured in red in (c) and (d).

Finally, in Figure 2.6 we display contact maps for *inter-chain* contacts made in the aggregated ensemble for 15% (green) and 8% (blue) of the proteins for the fast

aggregating pathway and slow aggregation pathway, for protein L (Figure 2.6a and 2.6b) and protein G (Figure 2.6c and 2.6d), respectively. We also show a snapshot of aggregated chains for protein G in Figure 2.7. The 15% contour is comparable to level of significance of the intra-chain contact maps, which is the reference ensemble used in this comparison (red). The point of this comparison is to show that protection afforded by the intra-chain contacts reduces their representation in the inter-chain contacts that can be made. The more permissive contact level of 8% emphasizes that protein L gives rise to aggregates with more inter-chain contacts, and exhibits a greater degree of domain swapping, especially between strands 2 and 4' (2' and 4), as well as inter-chain association of same strands, i.e. 2 and 2' as well as 4 and 4'. It is clear that the greater protection factor afforded by the stable structural elements dispersed over the entirety of the DSE of protein G, with the ISE viewed as an especially structured DSE, results in a much sparser inter-chain contact map. Protein G has a much reduced propensity for domain swapping, and largely exhibits only inter-chain association of same strands 3 and 3'. In fact the third β -strand is the stickiest region of protein G, and therefore potentially more harmful with respect to unwanted aggregation, but its rapid protection in the folding mechanism as an early intermediate potentially minimizes this destructive tendency.

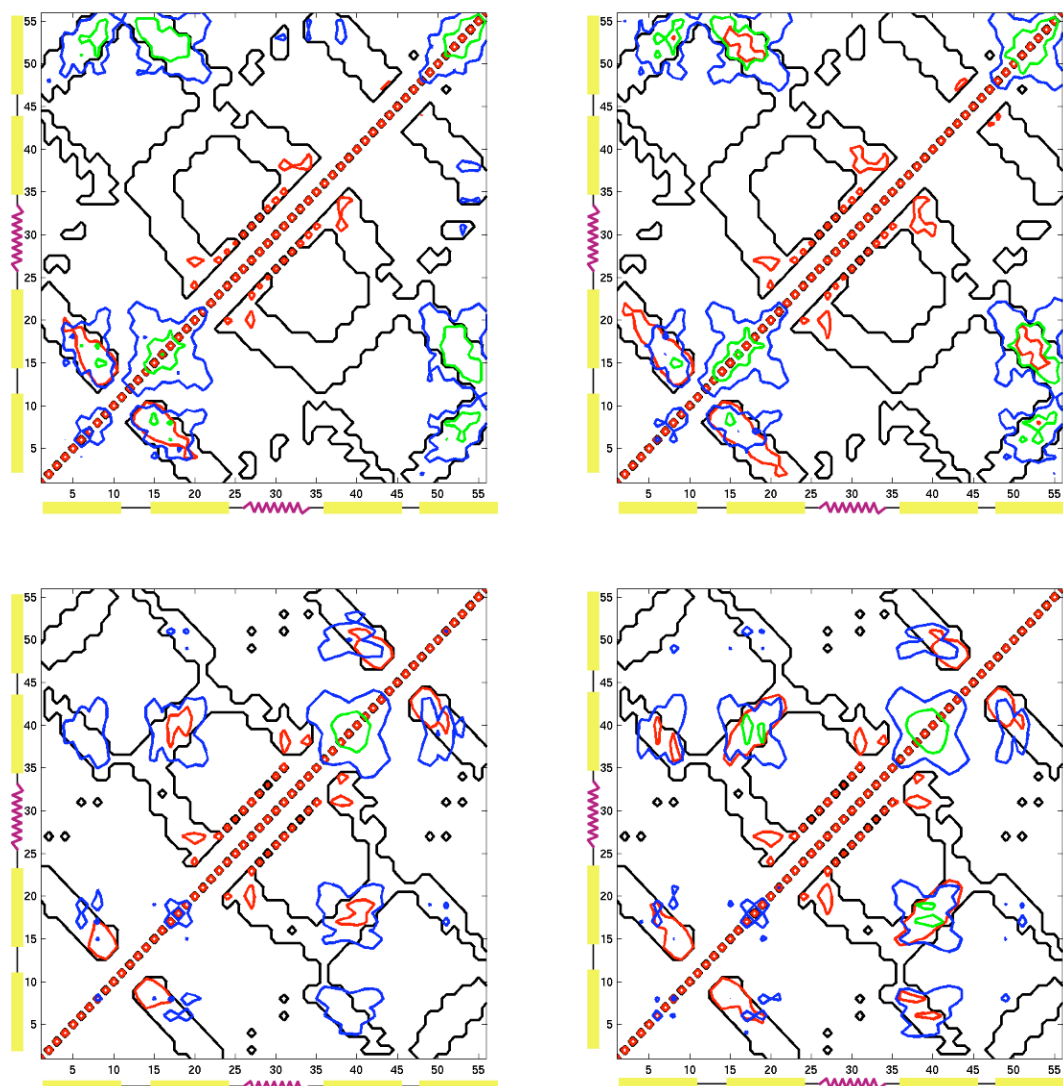


Figure 2.6: Contact map comparing native state (black) and *inter-chain* contacts made in at least 15 % (green) and 8% (blue) of the aggregated ensemble for (a) protein L's fast aggregation pathway, (b) protein L's slow aggregation pathway, (c) protein G's fast aggregating pathway, and (d) protein G's slow aggregation pathway, respectively. For reference, *intra-chain* contacts that are present in at least 60% of the aggregated ensemble of protein L are contoured in red in (a) and (b), and in at least 60% of the aggregated ensemble of protein G are contoured in red in (c) and (d).

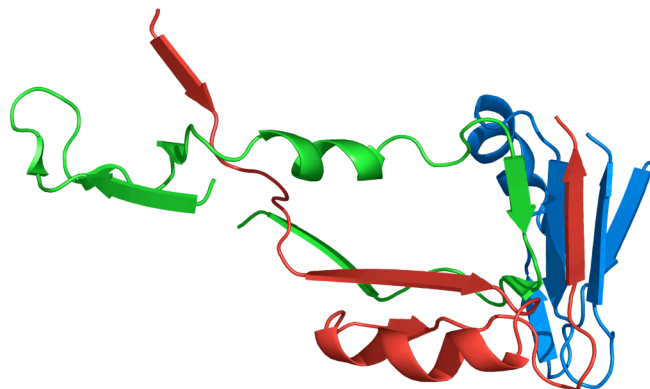


Figure 2.7: Ribbon diagram of a snapshot for the aggregation simulation of protein G that illustrates both native intra-chain and inter-chain contacts made. Figure created by PyMOL[42].

2.3 Discussion and Conclusion

The purpose of our study is to demonstrate that the different folding characteristics of protein L and G, both which form the same native topology, explain their different rates of aggregation. The aggregated ensembles for proteins L and G show intra-chain contact maps that strongly resemble the DSE, or intermediate ensemble if it exists, and therefore characterization of the DSE, early intermediates of the folding pathways, or even transition state ensembles of folding under non-aggregating conditions (low concentration) could provide information that will help explain the slower aggregation kinetics, and possibly the morphologies of aggregates, for different protein sequences.

The aggregation for protein G is slower than for protein L due to the presence of an intermediate in its folding pathway that quickly protects a number of regions of the sequence dispersed throughout the protein. While a number of studies have shown that intermediates can play a deleterious role by increasing protein aggregation [28-30, 48],

this work provides evidence that early on-pathway intermediates in folding could also play a protective role in abating unwanted aggregation. Correspondingly, the faster aggregation rate for L arises from the localization of stable structural elements in its DSE. The unstructured part of the chain leaves it vulnerable to domain swapping interactions with other chains that increase the aggregation rate and contributes to greater amount of inter-chain contacts. Therefore proteins which have localized structure in the DSE will be more aggregation prone than proteins with more diffuse elements of stable structure. In fact mutations that stabilize some elements of native structure in the second and/or fourth strand of protein L, could be made such that its folding pathway is not perturbed, although it should diminish its aggregation propensity relative to wild type.

It is a difficult experimental problem to determine the small populations of structure in the DSE at equilibrium, and the few experimental studies that exist have primarily focused on the DSE resemblance to the TSE [49-51]. In Figure 2.8a and 2.8b we show the resemblance between the DSE and TSE obtained from our model for protein L and protein G, respectively. It is interesting that the TSE population for protein G contains very similar or the same elements of structure populated in the DSE. Therefore information about the TSE does show correlation with aggregation propensity in this case. The DSE of protein L, however, exhibits association of β -strands 2 and 4 in both native and non-native configurations that are altogether absent in the TSE- an important difference since the that extra “pinning” site provides an alternative pathway that slows down the rate of protein L aggregation. In general, a more delocalized TSE should correlate with a reduction in the aggregation rate and amount of inter-chain contacts

formed in the aggregate, relative to sequences with a localized TSE, as we see when comparing protein L and G. Thus the “easier” characterization of the TSE through phi-value analysis may be a good guide about aggregation kinetics among different protein sequences, although the DSE would be more directly informative.

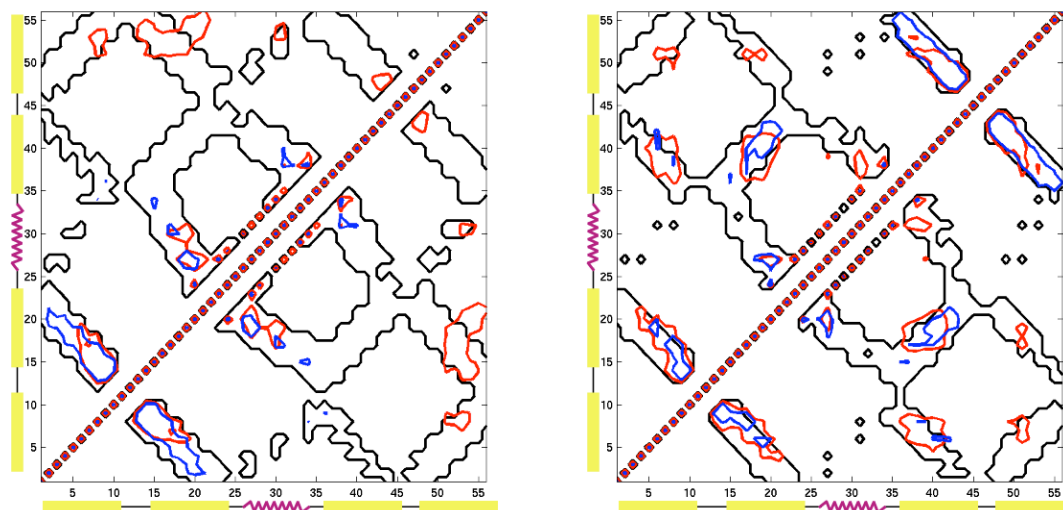


Figure 2.8: Contact map comparing native state (black), the denatured state ensemble (red), and the transition state ensemble (blue) for (a) protein L and (b) protein G.

The equilibrium experiments that report similarity in structure between the TSE and DSE are only suggestive of the role that it plays in the kinetics of folding [49, 50]. We note that the similarities seen between the TSE and DSE of our models of protein L and G do support the description of the DSE in the folding kinetics of proteins [52]. Protein L does exhibit a faster folding rate than protein G due to more localized contacts (helix and β -hairpin 1) formed in the DSE, and because it lacks a folding intermediate, all of which supports the connection between fast folding and minimal residual structure in the DSE. We might also suggest that residual *non-native* structure may also contribute

to a faster folding rate, since protein L's non-native association of β -strands 2 and 4 provide weak contacts (relative to native like interactions) and yet reduces the entropy of folding by drawing together disparate regions of the structure.

It is noteworthy that Capaldi and co-workers have determined that the folding mechanism of one member of a set of immunity proteins, Im7, involves an early intermediate, while the homologous members Im2, Im8, and Im9 are simple two-state folders [53, 54]. It would be interesting to see whether these and other structurally homologous protein families that show both two-state kinetics as well as on-pathway intermediates in folding correlate with differences in aggregation kinetics for the reasons that we have shown here for proteins L and G.

The primary conclusion of this computational study is that if experiments can characterize the structural signatures of the DSE, or possibly the TSE, then native contacts that are delocalized across the protein fold should correlate with slower aggregation rates. We suggest that there may be a functional advantage to a diffuse transition state ensemble or DSE to prevent not only misfolding [55-57], but also to aid aggregation resistance. This emphasizes that evolution has optimized protein sequences for functional robustness, not simply for folding rate; protein sequences that prefer slower pathways and/or folding intermediates may be evolutionary selected for, in part, due to their aggregation resistance [56, 57]. A further question is whether differences in folding pathways of homologous proteins might delineate differences in their functional role- i.e. that sequence differences in a given fold class have evolved to provide protection against aggregation depending on the specifics of their protein interaction partners in the cell. A

corollary of this work is that early intermediates in folding may be evolutionary selected for their protective role against unwanted aggregation, and thus could be useful to employ in reengineered sequences to slow aggregation and increase folding yield in industrial protein production. Furthermore, given the concentration conditions used in our study, and the fact that protein L and protein G form amyloid fibrils [43, 44], we would argue that these simulations are most directly relevant to the earliest aggregation events involving prefibrillar formation.

2.4 Methods

This work examines the question of aggregation kinetics and mechanism through computational coarse-grained models of two members of an α/β protein fold class, Ig-binding proteins L and G. We justify the use of coarse-grained models for the following reasons: They capture the correct spatial distribution of local and non-local contacts (Figure 2.1) of the most relevant native state features [7, 58, 59] which most influences the overall kinetics of protein folding [60, 61]. Coarse-graining in sequence should also be highly appropriate for aggregation studies, since it is clear that hydrophobic/hydrophilic amino acid sequence patterning plays a crucial role in determining a protein's ability to aggregate. Broome and Hecht [62] reported that alternating patterns of hydrophilic and hydrophobic amino acid residues occur significantly less often than other patterns, and Schwartz and coworkers [63] identify a more specific rule that blocks of three or more hydrophobic residues are disfavored

among wild-type proteins surveyed, indicating that there are sequence patterns that are particularly conducive to the formation of amyloid fibrils [64].

These minimalist models enable a sequence-driven connection to experimental protein folding mechanisms that is not reproducible by Go topology models. Since our models are based on physical potentials, we can engineer sequences that fold into α -helical, β -sheet, and mixed α/β protein topologies, and distinguish folding rates and mechanism between members within the same protein fold family. Therefore our computational model is also appropriate for protein engineering studies that have proven critical in understanding some basic aspects of aggregation phenomena.

The protein chain is modeled as a sequence of beads of three types, hydrophilic, hydrophobic, and neutral, designated by L, B and N, respectively [7, 59]. The pair-wise interaction between beads is attractive for hydrophobic-hydrophobic (B-B) interactions, and repulsive for all other bead pairs (although the strength of the repulsive interactions depend on the bead types involved). In addition to pair-wise non-bonded interactions, the other contributions to the potential energy function include bending and torsional degrees of freedom. The total potential energy function is given by

$$H = \sum_{\theta} \frac{I}{2} k_{\theta} (\theta - \theta_0)^2 + \sum_{\phi} \left[A(1 + \cos \phi) + B(1 - \cos \phi) + C(1 + \cos 3\phi) + D \left(1 + \cos \left[\phi + \frac{\pi}{4} \right] \right) \right] + \sum_{i, j \geq i+3} 4\epsilon_H S_1 \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - S_2 \left(\frac{\sigma}{r_{ij}} \right)^6 \right]. \quad (2.1)$$

where ϵ_H determines the energy scale and sets the strength of the hydrophobic interactions. The bond angle energy term is a stiff harmonic potential with force constant

$k_\theta = 20 \epsilon_H / \text{rad}^2$, and $\theta_0 = 105^\circ$. The second term in the potential energy designates the torsional, or dihedral, potential and is given by one of the following: helical (H), with $A = 0$, $B = C = D = 1.2 \epsilon_H$; extended (E), favoring β -strands, with $A = 0.9\epsilon_H$, $C = 1.2\epsilon_H$, $B = D = 0$; or turn potential (T), with $A = B = D = 0$, $C = 0.2\epsilon_H$. The non-bonded interactions are determined by: $S_1 = S_2 = 1$, a Lennard-Jones potential with a short range attractive minimum to represent the energetically favorable burial of hydrophobic groups for B-B interactions; $S_1 = 1/3$ and $S_2 = -1$, a repulsive interaction for L-L and L-B interactions; and $S_1 = 1$ and $S_2 = 0$, a softer repulsive interaction to mimic smaller amino acids for all N-L, N-B, and N-N interactions. For convenience all simulations are performed in reduced units, with mass m , length s , energy ϵ_H , and k_B all set equal to unity. Note that while the non-bonded potential is symmetric with respect to inversion, this is not true for the dihedral interactions. Thus the total energy function is not symmetric with respect to indice permutations, and we do not find mirror image states. Full details of the model can be found in our previous work[7, 59].

Folding and Aggregation Simulations

We perform constant-temperature simulations using Langevin dynamics in the low friction limit for three protein chains when characterizing the thermodynamics and kinetics of aggregation. Low friction stochastic dynamics enables the sampling of long timescale events (on the order of milliseconds or longer) such as folding and aggregation, but makes quantitative comparison to experimentally measured absolute time difficult. Therefore, we restrict our analysis to comparing timescales between protein L and G. We

performed 600 aggregation trajectories for both protein L and protein G at their reduced folding temperatures of $T_f=0.42$ and $T_f=0.41$, respectively. A reduced temperature is used by us for numerical convenience since it eliminates the use of small constants which accumulate error in an MD simulation. It is defined as $T^*=k_B T/\epsilon_H$, where $\epsilon_H=k_B=1$. A Langevin timestep equivalent to 0.005 unit time was used for all simulations.

In order to determine the kinetics of aggregation, we sought a thermodynamic definition of an aggregate based on the number of contacts by constructing a free energy surface describing aggregation. To this end, we collect multi-dimensional histograms [65, 66] over a number of different order parameters, including energy V , radius of gyration R_g , and various native-state or aggregation state similarity parameters χ . We collected histograms at 13 different temperatures: 0.90, 0.62, 0.60, 0.55, 0.50, 0.48, 0.46, 0.44, 0.42, 0.41, 0.40, 0.39, and 0.38. We run 10 independent trajectories at each temperature, and collect 5000 data points per trajectory. For each trajectory, the three chains start off in an arbitrary conformation at maximum separation in the periodic box. The chains are initially propagated at high temperature of 1.6 for 750,000 steps to randomize the starting configuration. The simulation is quickly cooled to the target temperature (5000 steps for target temperatures of 0.70 and 0.90, 10,000 steps for $0.5 \geq T \geq 0.62$, 20,000 steps for $T \leq 0.48$), then equilibrated for long times (500,000 timesteps 0.70 and 0.90, 1,000,000 steps for $0.5 \geq T \geq 0.62$, and 4,000,000 for $T \leq 0.48$) to ensure the simulation represents the equilibrium ensemble at the target temperature. The free energy landscape is characterized using the multiple, multi-dimensional weighted histogram analysis technique. From the histogram analysis, we constructed a projection of the free energy

surface on the parameters χ_{inter} and χ_{intra} . χ_{inter} is the number of bead pairs from different chains that are in close contact (within 1.28 distance units, corresponding to about 5Å). χ_{intra} is the number of bead pairs from the same chain in close contact. We selected the short contact distance of 1.28 to limit interactions counted as contacts to those that are very likely to be energetically favorable BB interactions. At contact length of 1.28 distance units, BB interactions are $-0.70\epsilon_H$, 70% of the minimum potential energy ($-1.0\epsilon_H$ at a distance of 1.122) and LL interactions are unfavorable ($+0.37\epsilon_H$). With these parameters, we identified the center of the free energy basin for aggregation based on the projection of the free energy surface onto the χ_{inter} and χ_{intra} parameters.

The kinetics of the aggregation process can be characterized by calculating a large number of mean first-passage times, the time required for an aggregation trajectory to reach χ_{inter} . In kinetics simulations, 3 chains start off in an arbitrary conformation at maximum separation in the periodic box. The chains are initially propagated at high temperature of 1.2 for 750,000 steps to randomize the starting configuration and ensure that each chain is in the unfolded conformation. The chains are then cooled extremely quickly (in 200 steps) to the folding temperature, and an equilibration period of 30,000 timesteps follows at the folding temperature. We subtract off this initial correlation time in which the high-temperature chain is briefly equilibrated at the target temperature (this is the computational dead time during the kinetics run). The chains are then propagated at the folding temperature and χ_{inter} is measured every 1000 steps until χ_{inter} reaches a rolling block average of the designated inter-chain contact number, $\chi_{inter}=38$ for G and $\chi_{inter}=28$ for L. The rolling block average is a short block average of the number of

contacts for the current and 9 previous sample points (the last 10,000 steps of simulation), implemented to reduce the noise in the number of contacts. The time this χ_{inter} is reached is recorded as the first-passage time. Trajectories are truncated at 6 million steps if χ_{inter} is not reached, and all kinetic fits are generated on trajectory data out to 5 million steps.

In order to examine structural differences in aggregates that form quickly and slowly, the population of kinetics runs was split at a point t^* . At t^* , assuming a two independent pathway model, $P(\text{trajectory is a member of the fast pathway} \mid \text{aggregated at } t^*) = P(\text{trajectory is part of the slow pathway} \mid \text{aggregated at } t^*)$. Trajectories aggregating before t^* are more likely to be from the fast pathway and represent the fast exponential timescale; trajectories aggregating after t^* are more likely to represent the slow exponential timescale. Any contamination of the populations where trajectories near t^* are incorrectly assigned will be small due to the order of magnitude separation in timescales.

To connect the aggregation properties to chain characteristics, findings from previous work characterizing the L and G single chain stationary points (transition and intermediate states) were included. Briefly, the structures along the folding pathway were isolated and for each structure the P_{fold} , the probability that a particular structure will find the native state before unfolding, was determined [67]. Structures with P_{fold} near 0.5 ($0.4 \leq P_{fold} \leq 0.6$) were identified as transition states (see [7] for full details). An ISE was postulated for protein G by observing the trajectories starting from a $P_{fold} \approx 0.5$ that did not unfold. To characterize the denatured state ensemble for the folding of individual chains, we simulate 16 independent single chain trajectories at constant temperature to

collect states at the folding temperature that reside in the unfolded basin, defined to be $0.0 < \chi < 0.4$ and $2.5 < R_g < 4.5$ as estimated from the partition function at T_f . When we compare the resulting distribution to the histogram partition function to validate that we are sampling the proper distribution of states, we find good agreement.

2.5 Acknowledgments.

We would like to acknowledge financial support from UC Berkeley and a subcontract award under the National Sciences Foundation Grant No. CHE-0205170. NJF gratefully acknowledges the Whitaker foundation for a graduate research fellowship. We thank Troy Cellmer and Harvey Blanch for communicating their aggregation amyloid data on protein L.

Chapter 3: Determining the Critical Nucleus and Mechanism of Fibril Elongation of the Alzheimer's A β ₁₋₄₀ Peptide

Reproduced with permission from the *Journal of Molecular Biology*. 2007 Jan 12;365(2):535-50. Copyright 2006 Elsevier Ltd.

Nicolas Lux Fawzi¹, Yuka Okabe², Eng-Hui Yap¹, and Teresa Head-Gordon^{1,2,3}

¹UCSF/UCB Joint Graduate Group in Bioengineering, Berkeley, CA 94720

²Department of Bioengineering, University of California, Berkeley, CA 94720

³Department of Chemistry, Cambridge University, Cambridge CB2 1EW, UK

Abstract

We use a coarse-grained protein model to characterize the critical nucleus, structural stability, and fibril elongation propensity of A β ₁₋₄₀ oligomers for the C_{2x} and C_{2z} quaternary forms proposed by solid state NMR. By estimating equilibrium populations of structurally stable and unstable protofibrils, we determine the shift in the dominant population from free monomer to ordered fibril at a critical nucleus of 10 chains for C_{2x} and C_{2z} forms. We find that a minimum assembly of 16 monomer chains is necessary to mimic a mature fibril, and show that its structural stability correlates with a plateauing in the hydrophobic residue density and a decrease in the likelihood of losing hydrophobic interactions by rotating the fibril sub-units. While A β ₁₋₄₀ protofibrils show similar structural stability for both C_{2x} and C_{2z} quaternary structures, we find that the fibril elongation propensity is greater for the C_{2z} form relative to the C_{2x} form. We attribute the increased propensity for elongation of the C_{2z} form as being due to a stagger in the interdigitation of the N- and C-terminal β -strands, resulting in structural asymmetry in the

presented fibril ends that decreases the amount of incorrect addition to the N-terminus on one end. We show that because different combinations of stagger and quaternary structure affects the structural symmetry of the fibril end, we propose that differences in quaternary structures will affect directional growth patterns and possibly different morphologies in the mature fiber.

3.1 Introduction

The aggregation of peptides or proteins into ordered amyloid fibril morphologies is associated with over 20 human diseases, including Alzheimer's disease, dialysis-related amyloidosis, and bovine spongiform encephalopathy[68, 69]. The fibrils have a characteristic "cross- β " structure where intermolecular β -sheets run along the long axis of the fibril, stabilizing the assemblies which can extend to microns in length[69]. Although early attention focused on the toxicity of the amyloid fibrils as the cause of disease, it is now hypothesized that oligomers formed during early aggregation are actually the major toxic species[11, 12]. This shift underscores the need to develop an understanding of the entire aggregation process that ultimately leads to the specific structure of the final amyloid fibril.

Alzheimer's is a neurodegenerative disease linked to the aggregation and amyloid fibril formation of a set of short ~40 residue peptides, amyloid β ($A\beta_{1-39,1-40,1-42}$), created by proteolytic cleavage of the amyloid precursor protein (APP)[70]. These fragments contain part of the C-terminal region of the APP protein, and are known to be highly prone to fibrilization in vitro and in vivo[21, 71-74]. The structure of the monomeric peptide has no well defined folded state although tertiary structures that are dependent on

solution conditions have been proposed from experimental and simulation work[75-78]. The backbone conformation can vary from α -helical structure in non-polar solutions as determined by solution NMR[75, 76] to disordered N- and C-terminal tails with a consistent turn region as determined from electrospray mass spectrometry and implicit solvent molecular dynamics[77, 78]. The structure of the $A\beta_{21-30}$ sub-peptide, encompassing a proteolysis resistance region of the full length sequence, has also been determined by NMR[79].

At the other extreme, the complete $A\beta_{1-40}$ amyloid fibril state has been extensively studied by Tycko and coworkers who have published a series of model structures based on constraints from solid state NMR[13, 80-83]. The proposed structure is shown in Figure 3.1 and is described as “U-shaped” monomers with two in-register parallel intermolecular β -sheet regions (N- and C-terminal β -sheet); the cross section of the fibril is made up of two monomers with hydrophobic C-terminal regions in van der Waals contact. The original NMR data [82] supported two possible intra-fibril contact types (“unflipped” and “flipped”) for the C-terminal β -strand, and eventually the “unflipped” form was eliminated based on tertiary sidechain-sidechain contacts [13].

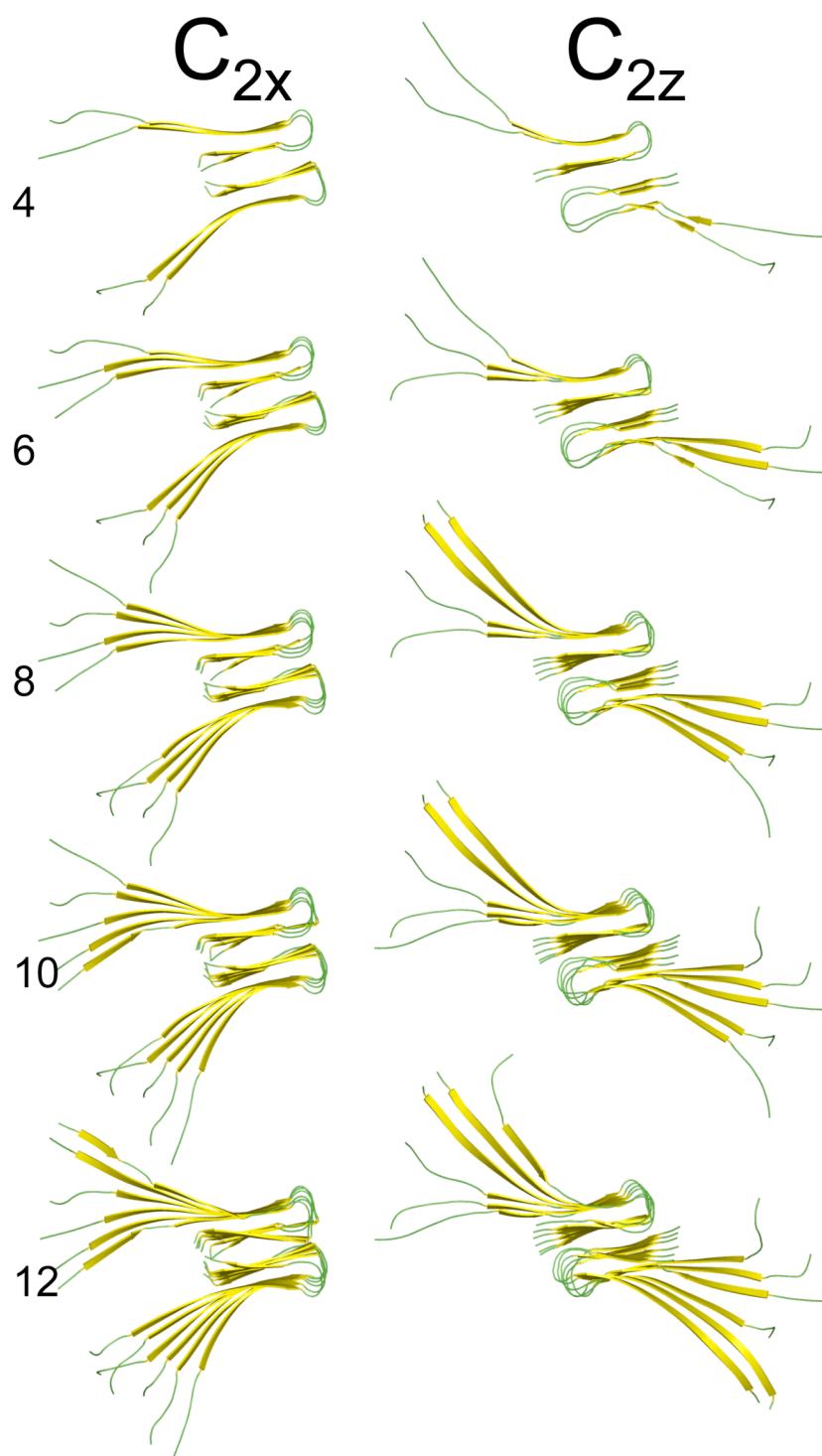


Figure 3.1: Examples of starting structures for C_{2x} and C_{2z} symmetry forms. Protofibril seeds composed of 4, 6, 8, 10, and 12 (14-20 not shown) monomers were simulated for protofibril stability.

Furthermore, two quaternary structures denoted as C_{2x} and C_{2z} were proposed[84], based on approximate C2 symmetry around the x axis (approximately orthogonal to the fibril axis and parallel to the β -strand directions) and C2 symmetry around the z axis (parallel to the fibril axis), respectively, and shown in Figure 3.1. Note that these are only pseudo-symmetry designators since there is imperfect matching of sidechain inter-digitation in the C-terminal region on opposite subunits of the relevant protofibril symmetry axis in both cases. More complete NMR data revealed that only the C_{2z} quaternary structure was likely to be formed *in vitro* based on specific 2D NMR cross-peaks that give tertiary contacts that are inconsistent with the C_{2x} quaternary form [13]. Most recently, however, a fibril made from shortened, mutated A β monomers covalently linked at the N-termini created fibrils with a likely C_{2x} symmetry, indicating that the C_{2x} form may be found under certain conditions[85].

Finally, the NMR data also supports inter-digitation of the N- and C- terminal β -strands to form side-chain contacts with a particular “stagger” of N- and C-terminal hydrophobic contacts[13], shown schematically in Figure 3.2. Based on isotopic dilution studies, side chain contacts are proposed between the C-termini of monomer i with the N-termini of monomer $i+1$ and $i+2$ (STAG(-2)) or between the N-termini of monomer i with the C-termini of monomer $i+1$ and $i+2$ (STAG(+2))[13]. In totality, the solid state NMR work is a truly seminal contribution to the amyloid field since these experimental models have provided well-defined structural constraints on the “folded state” of the A β ₁₋₄₀ monomer in the context of the formed fibril.

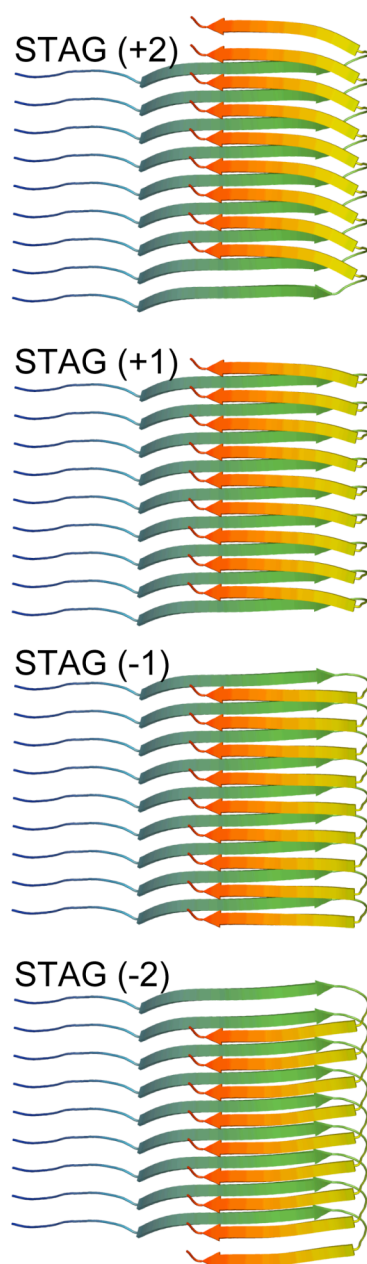


Figure 3.2: Interdigitation of the N- and C- terminal β -strands to form side-chain contacts between different monomer chains introduces a stagger in the strand alignments. Side chain contacts between the C-termini of monomer i with the N-termini of monomer $i+1$ and $i+2$ (STAG(-2)) or between the N-termini of monomer i with the C-termini of monomer $i+1$ and $i+2$ (STAG(+2))[13]. Our model naturally relaxes to the STAG (-1) definition, although we show the STAG (+1) for completeness.

Given the possible toxicity of the earlier protofibril states, the focus is now to understand how the A β monomers assemble into the highly ordered mesoscopic fibril as proposed by the NMR experimental models. The mechanism of fibrilization of full length A β peptides (A $\beta_{1-39,1-40,1-42}$) has been shown to follow an apparent nucleation-dependent polymerization[21, 74, 80, 85], whereby a small number of monomers associate through a free energy barrier corresponding to a critical nucleus size, beyond which initiates a gradient of favorable free energy or “down-hill” polymerization into a macroscopic fibril[86] (Figure 3.3). However, the structural characteristics and oligomer size of this ensemble of fibril nucleating species have yet to be determined, and the mechanism of monomer addition is unclear. This is in part due to the limited access of experimental characterization to this earliest aggregation stage, thus providing an opportunity for theoretical studies to bridge the experimental gap between the monomer and fibril endpoints and to develop testable hypotheses.

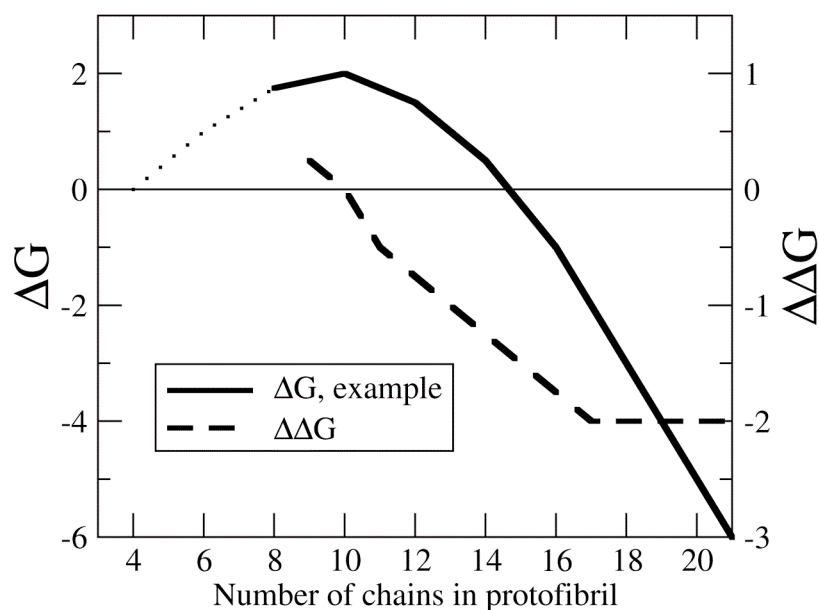


Figure 3.3: Free energy profile for the nucleation-polymerization reactions. Typical free energy (ΔG) profile and slope of ΔG ($\Delta\Delta G$) vs. number of chains in protofibril for fibril formation by a nucleation-dependent polymerization mechanism. At high number of chains, the protofibril is stable and free-energetically favorable, and the free-energy benefit to adding chains is constant, as seen in a constant slope of ΔG . Since the slope of ΔG is constant in this regime, the free-energy benefit to adding a chain or free-energy cost for removing a chain is the same as in an infinite fibril. As the number of chains decreases, the free-energy change for removing chains decreases, indicating that the fibril is approaching the number of chains in the critical nucleus. At the critical nucleus, the least free-energetically favorable species, the slope of ΔG is zero. (Typical ΔG data adapted from Ferrone[86])

Many computational studies using coarse-grained as well as all-atom models have focused on the formation of the anti-parallel β -sheet structure by sub-peptides of $A\beta$, particularly $A\beta_{16-22}$ [87-89]. The anti-parallel structure of these peptides, however, suggests that studies of the steps in fibril formation of this system will not lead to information regarding the nucleation and fibril forming properties of the in-register

parallel structures formed by the full-length $A\beta_{1-40}$ and $A\beta_{1-42}$ peptides. More recent simulation work has therefore focused on the full length $A\beta$ peptides. Coarse-grained simulations by Stanley and coworkers of the $A\beta_{1-40}$ and $A\beta_{1-42}$ monomers and dimers have reproduced some of the properties of the disordered peptides in solution[90, 91], but underscore the computational and modeling difficulty of forming structures resembling fibrils. All atom simulations conducted by Shea and coworkers give detailed insight into the monomer structure in dilute solution and *in vacuo*[77, 78]. In a set of all-atom molecular dynamics simulations with explicit water representation, Hummer and coworkers demonstrate that with incomplete NMR data from Tycko and coworkers, a set of four related but distinct minimum fibril models consistent with the NMR restraints are all structurally stable for at least a few nanoseconds[84]. All atom simulation of the full-length $A\beta$ peptide aggregation is likely difficult due to the extremely long experimental timescales (hours to days depending on conditions) and large system sizes (~ 10 peptides of 40 amino acids) necessary for fibril formation.

We have recently developed a new coarse-grained protein model which is a greatly enhanced version of coarse-grained models we have used in studying protein folding and non-disease protein aggregation[7, 92-96]. The new model, described in Methods, has been validated on folding thermodynamics and kinetics for proteins L and G, and provides higher structural resolutions ($\sim 3.0\text{\AA}$ C_α RMSD) of the folded state relative to our old model, especially for descriptions of β -sheets (Yap, Fawzi and Head-Gordon, manuscript in preparation). We use this new model for the first time to simulate

$A\beta_{1-40}$ oligomerization in order to address three primary questions regarding the association of $A\beta_{1-40}$ peptides into fibrils.

First, what is the number of peptides involved in the critical nucleus for subsequent fibril elongation and does it differ among quaternary forms? Starting from a mature fibril structure composed of 40 chains, we systematically shorten the protofibril and measure structural stability over an equilibrium ensemble for each n -chain oligomers for each quaternary form. By calculating the equilibrium populations of structurally stable and unstable protofibrils, we determine the shift in population dominated by free monomer to the ordered protofibril to quantify free energy profiles for our model as per Figure 3.3. Based on this thermodynamic analysis, we determine that the barrier in free energy occurs at a critical nucleus value of 10 chains for both quaternary C_{2x} and C_{2z} forms.

Second, given the hypothesis of a nucleation-dependent polymerization mechanism, and NMR guidance as to the structure of the monomer in the mature fibril[13, 80, 82], what is the minimum number of chains in an ordered oligomer necessary for assembly of a structurally stable protofibril? The commonly assumed nucleation dependent polymerization (Figure 3.3), suggests that beyond the critical nucleus size there is a minimum stable protofibril that reaches a constant $\Delta\Delta G$ for subsequent monomer addition, and thus initiates the behavior of a long fibril. We find that this constant addition free energy regime is evident from the thermodynamic analysis of our model, and determine that mature fibril behavior is reached at ~ 16 chains. We also show that the constant free energy for monomer addition correlates with a plateauing in

the hydrophobic residue density that in turn correlates with structural order and stability by decreasing the likelihood of losing hydrophobic interactions due to fibril sub-unit rotation.

Finally, what is the fibril elongation mechanism of $A\beta_{1-40}$ and can we distinguish it between the two quaternary structures? We find that the C_{2z} form shows a greater ratio of correct parallel N-termini addition to incorrect anti-parallel addition relative to the C_{2x} quaternary structure. We attribute this difference in elongation between the two quaternary forms as arising from differences in structure at the fibril ends due to the consequences of stagger in the inter-digitation of the N- and C-terminal β -strands. We find that the C_{2z} form exhibits a structural asymmetry in the fibril seed ends, with one side exposing the N-terminus region while the other side exposes the C-terminus region, while the fibril ends for the C_{2x} quaternary structure are not structurally distinguishable. This inter-digitation and interplay with quaternary structure suggest unidirectional growth of the protofibril for the C_{2z} quaternary form, while we expect bi-directional growth for C_{2x} based on our model. We also show that mixed stagger forms (+N on one half the fibril and -N on the other half) reverse the end symmetries of the C_{2x} and C_{2z} quaternary forms, leading to potentially different mature fibril morphologies.

3.2 Results

Symmetry of fibril ends for different quaternary forms

To examine the effect of stagger on the fibril quaternary structures, we built model structures for both C_{2x} and C_{2z} symmetries for internal stagger values of -2, -1, +1 and +2 (Figure 3.2). When the C_{2x} models are constructed with any pure stagger and

examined as a two monomer cross-section (down the fibril axis), one monomer has a protruding C-terminal strand, while the paired monomer has a protruding N-terminal strand (Figure 3.4). Although the resulting C_{2x} protofibril has only approximate C_2 symmetry around the fibril axis, due to imperfect interdigitation of the residues involved in the C-terminal hydrophobic interface[13], the resulting C_{2x} fibril ends are nearly indistinguishable (Figure 3.4). When fibrils with C_{2z} symmetry are constructed with any pure stagger, there is more perfect C_2 symmetry around the fibril axis relative to C_{2x} , however when examined in cross-section, the ends are distinguishable. For the case of C_{2z} fibrils, one end has both monomers presenting a protruding N-terminal strand, while the alternate end has both monomers exhibiting a protruding C-terminal strand (Figure 3.4). If instead we construct a C_{2z} fibril with a mixed stagger: i.e. a +1 stagger for one of the fibril halves and a -1 stagger for the other, the resulting C_{2z} fibril structure shows symmetrized ends while the C_{2z} fibril shows asymmetric ends. We note that the known solid state NMR constraints does not preclude the possibility of a mixed stagger. In what follows in later sections, we label the two ends as A or B to examine the consequences of symmetry or asymmetry of ends on structural stability and mechanism and rates of fibril growth in our models.

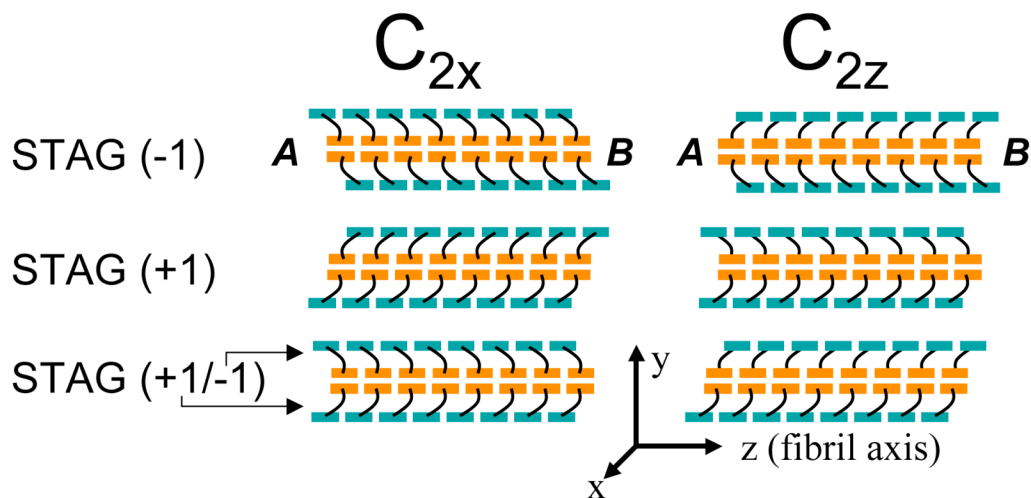


Figure 3.4: Effect of internal stagger on terminating ends of fibril. A schematic of 16 chain C_{2x} and C_{2z} fibrils are shown for internal staggers STAG (-1), STAG (+1) and mixed STAG(+1/-1).with N-terminal region colored in teal and C-terminal region colored in orange. (a) STAG (-1) C_{2x} has superimposable, symmetric ends. End A can be approximately superimposed on End B by a simple rotation of 180° about the x-axis (hence C_{2x}). STAG (-1) C_{2z} has distinct, asymmetric ends. End A exposes the C-terminal β -strands, and End B exposes the N-terminal β -strands. Ends A and B of the C_{2z} fibril cannot be superimposed on End A by any rotation. (b) C_{2x} STAG (+1), like C_{2x} STAG(-1), has superimposable, symmetric ends. C_{2z} STAG (+1), like C_{2z} STAG (-1) above it, has distinct, asymmetric ends. (c) C_{2x} STAG (-1/+1) has the top peptide STAG (+1) and bottom peptide STAG (-1). Mixing staggers in C_{2x} de-symmetrizes the C_{2x} ends. Mixing staggers in C_{2z} symmetrizes the C_{2z} ends so that each end has one subunit with an exposed N-terminal β -strand, and the other with an exposed C-terminal β -strand, unlike the two asymmetric ends in “pure” C_{2z} STAG (-1) or C_{2z} STAG (+1) models.

Structural Stability and Identification of Critical Nucleus

We next investigate the structural stability of fibril seed models for different seed sizes for both the C_{2x} and C_{2z} forms by simulating their dynamics at a constant

temperature of $T^*=0.45$ ($T \approx 337\text{K}$), and monitoring the amount of fibril order as a function of time. As a measure of fibril order, we define a structural similarity parameter, χ_f , which measures the fraction of residue pair-distances retained in the β -sheet regions, and restricted to the two exterior chains on each end of the protofibril structure and their two neighboring chains (see Methods for χ_f definition). The χ_f metrics allow a direct comparison between structures of different numbers of chains since only the exposed and subsequent layer are included, so that changes in the metric vs. number of chains is not simply due to the slower dynamics of a larger fibril seed. Since this metric includes contacts at the C-terminal interface between the subunits, it is also sensitive to translation and rotation of one subunit with respect to the other (perpendicular to the fibril axis), and thus measures disorder of the quaternary structure. Due to asymmetries in structure of the exposed ends depending on quaternary symmetry, we measure the structural integrity of both the A and B exposed ends of the fibril separately (Figure 3.4).

We simulate protofibrils with cross-sections composed of two “U-shaped” monomers ranging in number between 4 and 20 chains, as shown in Figure 3.1, and observe changes to the structural integrity of the fibril seeds by monitoring χ_f . In Figure 3.5 we show the time course of χ_f for different structured oligomer sizes, while Figure 3.6 plots the average of this metric for the last 500 τ of simulation time, $\langle \chi_f \rangle$, vs number of chains, for the two quaternary forms. The χ_f and $\langle \chi_f \rangle$ trends with increasing oligomer size shows increasing quaternary order due to a decrease in motion of one side of the fibril relative to the other, and thus providing greater stabilization of fibril ends. We note that at ~ 16 chains for both the C_{2x} and C_{2z} forms, χ_f and $\langle \chi_f \rangle$ saturates.

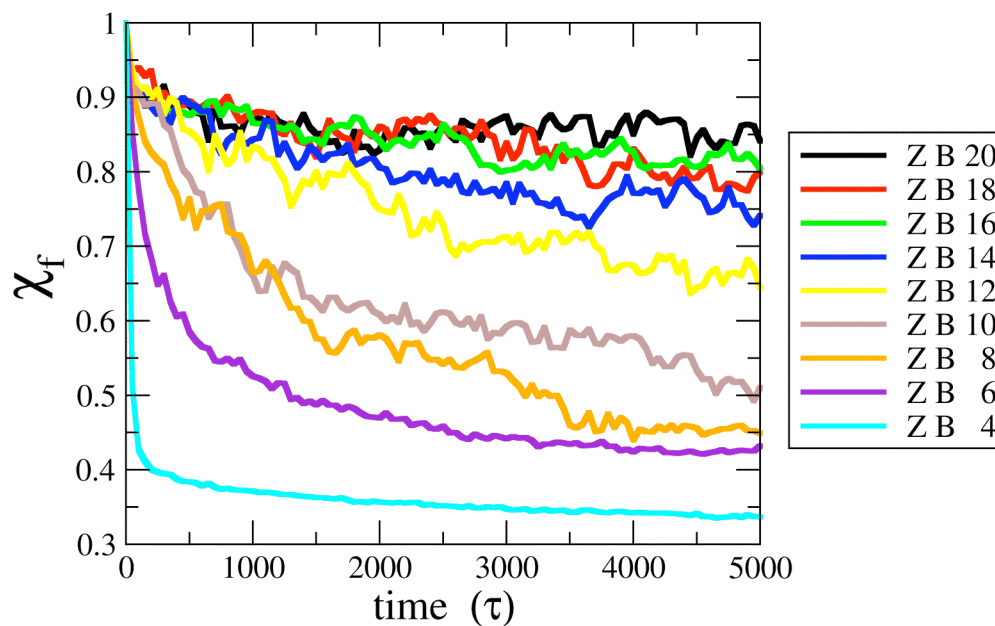


Figure 3.5: Time course for protofibril stability measured by χ_f . The metric χ_f measures the pair distances between the residues on both sides of the fibril, and thus is more sensitive to rotation of one subunit with respect to the other, and thus measures fibril disorder of the quaternary structure. The time course data averaged over all trajectories of C_{2z} fibrils for lengths 4 to 16 chains for fibril end B.

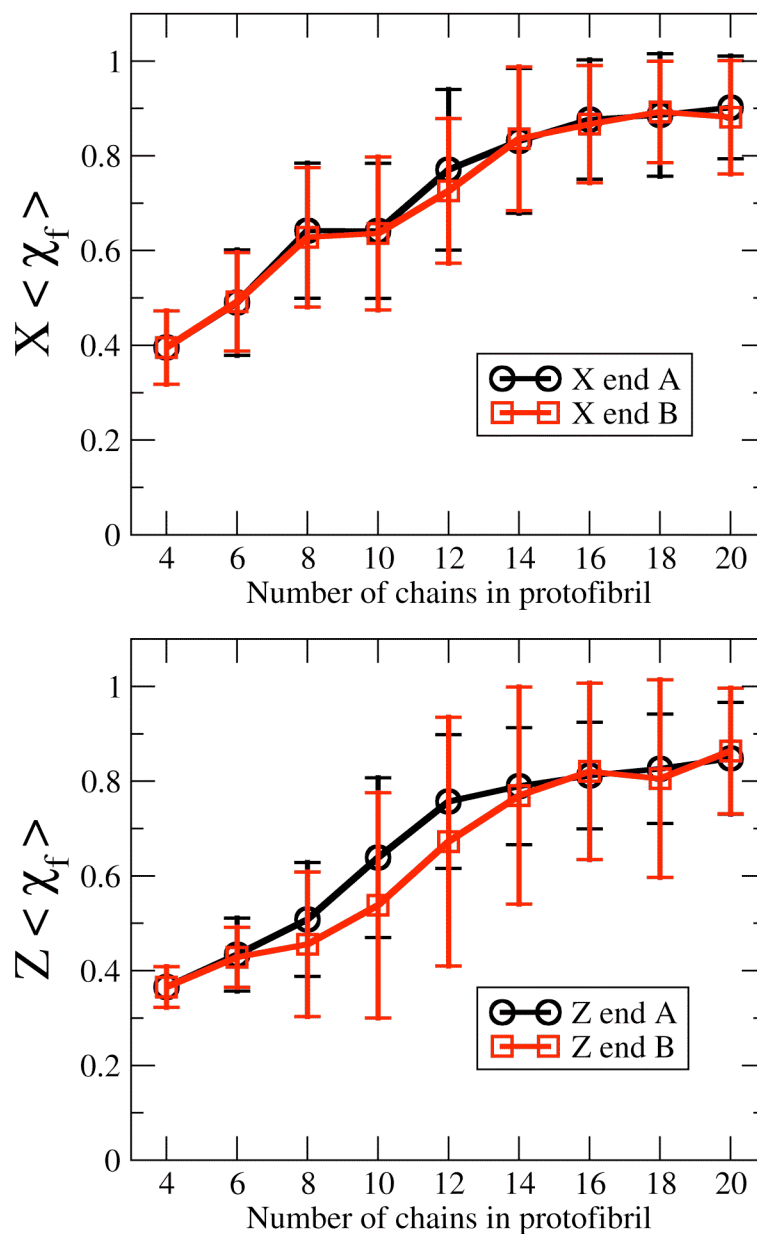


Figure 3.6: Protofibril stability measured by $\langle \chi_f \rangle$ vs. number of chains. $\langle \chi_f \rangle$ is an average measure of the fibril order of the edge chains for stable quaternary structure for (a) C_{2x} form and (b) C_{2z} for the two ends of the protofibril: end A (black) and end B (red). Error bars are standard deviation.

Based on the ensemble composed of the final structures of each of the trajectories for a given oligomer size, n , we calculate equilibrium populations of structurally stable

and unstable protofibrils based on a χ_f cutoff value of 0.7. The fraction of trajectories which correspond to $\chi_f > 0.7$ measures a population, C_n , of n -ordered monomers in a protofibril or seed with intact end monomers. This population is in equilibrium with the remaining fraction of trajectories corresponding to a protofibril with loss of structural order of one monomer end, and thus measures the population C_{n-1} . Based on thermodynamic arguments advanced by Ferrone[86] for nucleation-polymerization reactions relevant for aggregation kinetics, at equilibrium we can estimate the change in free energy, ΔG , per unit monomer as

$$\frac{d\Delta G}{dn} = -kT \ln \left(\frac{[C_{n-1}]}{[C_n]} \right) \quad (3.1)$$

where n is half the number of monomers and integration over all oligomer sizes allows us to generate a free energy curve like Figure 3.3 based on C_n and C_{n-1} populations measured in our model.

Table 3.1 gives the populations of C_n and C_{n-1} and Figure 3.7 plots calculated free energies as a function of oligomer size, and as a function of quaternary symmetry. It is evident that the critical nucleus size is 10 chains for both the C_{2z} and to C_{2x} quaternary structures. Below that number of chains there is a free energy barrier to association into ordered oligomer chains, and thus the equilibrium shifts in favor of the free monomer. At ~ 16 chains and above, consistent with the averaged time course data in Figure 3.6, the oligomer does not lose overall fibril structure, and now reaches a constant free energy gain for addition of new monomers to the ordered protofibril (Figure 3.7).

Number of chains, n	C_{2x} Symmetry			C_{2z} Symmetry		
	C_n	C_{n-1}	$\frac{d\Delta G}{dn}$	C_n	C_{n-1}	$\frac{d\Delta G}{dn}$
4	0.0000	1.0000	----	0.0000	1.0000	----
6	0.0312	0.9688	2.060	0.0208	0.9792	2.310
8	0.2604	0.7396	0.626	0.0729	0.9271	1.526
10	0.2708	0.7292	0.594	0.3028	0.6972	0.500
12	0.5625	0.4375	-0.151	0.5729	0.4271	-0.176
14	0.7569	0.2431	-0.682	0.7083	0.2917	-0.532
16	0.8333	0.1667	-0.966	0.8333	0.1667	-0.966
18	0.8750	0.1250	-1.168	0.8021	0.1979	-0.840
20	0.8854	0.1149	-1.227	0.8021	0.1979	-0.840

Table 3.1: Equilibrium populations of ordered fibrils, C_n and populations with free monomer, C_{n-1} , and calculated changes in free energy, ΔG , per unit monomer based on Equation 3.1.

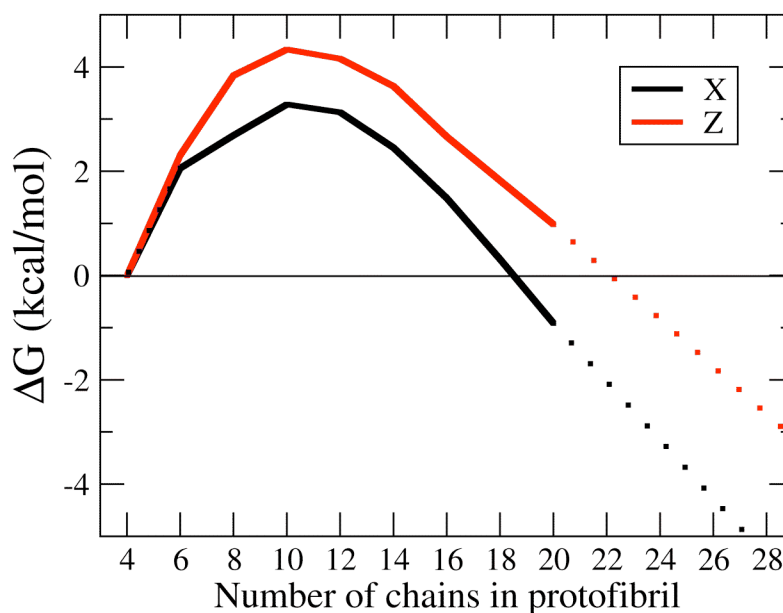


Figure 3.7: Free energy profile for free monomer and protofibril equilibrium. The free energy vs number of ordered chains in an oligomer is plotted for C_{2x} (X, black) and C_{2z} forms. The free energy shows a clear maximum at 10 chains for C_{2z} and 10-12 chains for C_{2x} , indicating the region of the critical nucleus. A constant, negative slope at ~16 chains and above is indicative of reaching a stable fibril regime.

The underlying molecular explanation for the increasing stability of a quaternary assembly up to ~16 monomer chains, and the constant free energy gain for subsequently larger protofibrils, is evident by evaluating the hydrophobic residue density of the starting structures at each seed size and each symmetry. Because hydrophobic interactions are thought to stabilize amyloid fibril structures, once a fibril reaches a certain length the average hydrophobic residue density should be a constant. To test this hypothesis, the hydrophobic residue density of the core of the equilibrated starting structures was measured by calculating the number of hydrophobic (B) residues within 2.0 units (7.6Å) of the tagged residue (excluding 1st and 2nd neighbors on the same peptide) divided by the volume, averaged over residues 29-40 and over all the peptides in the structure.

Figure 3.8 plots the average hydrophobic residue density vs. the number of chains in the oligomer for both symmetries. The average hydrophobic density correlates with the stability of the oligomers in Figures 3.5 and 3.6 and the linear regime of free energy once past the critical nucleus in Figure 3.7; as the oligomers get larger, the stability and hydrophobic density both grow, up until ~16 chains where both the structural stability and hydrophobic density level-off. Of note is that C_{2x} and C_{2z} form do not show strikingly different stabilities in this analysis, meaning that they are both reasonable fibril quaternary structures, similar to what was found by short all atom-simulation of 8 chain structures[84].

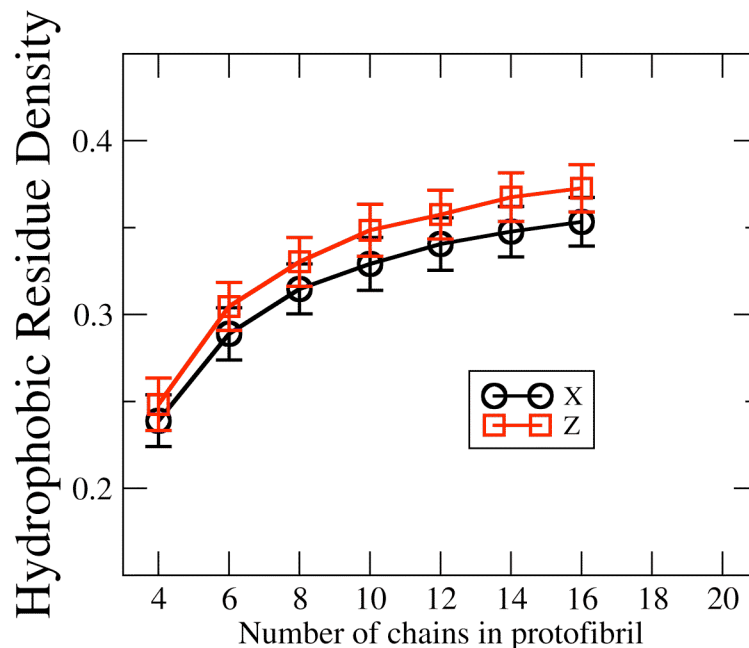


Figure 3.8: Hydrophobic residue density vs number of chains. Hydrophobic density (number of hydrophobic residues per unit volume) vs number of chains for the C_{2x} and C_{2z} forms after initial equilibration. Error bars are standard deviation for the 24 structures created from the 40 chain equilibration runs. The hydrophobic density for C_{2z} is higher than C_{2x} for all oligomer sizes.

Fibril Elongation Studies

Since the lag-time for forming amyloid fibril for $A\beta_{1-40}$ takes as much as a few days in the laboratory, even coarse-grained simulations of fibril formation from entirely disordered peptides may be intractable. Seeding a solution of $A\beta_{1-40}$ with fragments of pre-formed $A\beta_{1-40}$ fibrils, however, skips the lag phase and experiment shows that fibril formation proceeds rapidly relative to the unseeded experiments[80]. Although orders of magnitude faster than the lag time in fibril formation, fibril elongation is still a slow process relative to simulation timescales. Goto and coworkers recently measured a

sustained rate of amyloid fibril elongation to be 200 nm/minute, which corresponds approximately to 70 ms per monomer incorporation into the fibril[97].

Nonetheless, simulations which incorporate unstructured monomers into protofibril seeds should be more tractable than forming the fibrils from disordered monomers. The ability to propagate the elongation of the fibril through the addition of free monomers is a minimum necessary condition to show that the physics of the model represent the relatively fast fibril formation of the $A\beta_{1-40}$ system in seeding experiments. This simulation also enables another comparison of the C_{2x} and C_{2z} quaternary structures because the ability of the structure to propagate by elongation could be a criterion to determine which of the two structures is the most likely formed *in vitro* and perhaps *in vivo*; a structure that does not elongate will not be the structure that forms the amyloid fibrils measurable by solid state NMR.

The equilibrated 16 chain fibrils from the stability runs were used as seeds for fibril elongation simulations, since based on results in the previous section this oligomer size should be acting as a proper protofibril. Typical seeded fibril kinetics experiments for $A\beta_{1-40}$ use a peptide concentration on the order of $100\mu\text{M}$ [80, 97], equivalent to one peptide for a simulation box 270\AA on a side. Simulating a system at that dilution would require significant amount of computational time devoted entirely to diffusion of peptides towards the seed. To focus our study on the elongation of fibrils, our simulation conditions comprised of two equilibrated $A\beta_{1-40}$ peptides which are placed randomly and uniformly on the surface of a sphere with origin at the center of the fibril end, defined as the midpoint on the line connecting the 33rd bead on the two exterior fibril peptides. The

two peptides, one at each end of the fibril, were placed so that amino acid 20 was 5 units (19Å) from the center of the fibril end, and configurations where the peptide overlapped with the seed were excluded. Given the large seed size used here, the two peptides placed at opposite ends of the fibril rarely interact with each other. This procedure was the same for C_{2x} and C_{2z} symmetry forms. From ~2000 of these prepared starting structures, each were simulated for 1000τ (200,000 time steps) at $T^*=0.45$.

The number of trajectories resulting in the formation of partial (3 or more amino acids with formed hydrogen bonds) parallel additions in the N- and C- terminal β -sheet regions of the fibril seed, and “incorrect” anti-parallel additions to the fibril seed was then calculated. In this analysis, all parallel additions within the β -sheet regions are summed, even if they are not fully in register, though in-register parallel additions made up on average 75% of all parallel additions. (The data was also analyzed with out-of-register parallel additions either ignored or added to “incorrect” additions; the conclusions remained the same.) Simulations resulting in both N-and C-terminal in-register parallel addition did occur, but made up <0.5% of the population, making comparison between symmetries difficult. An example of an addition demonstrating both N- and C-terminal addition is shown below in Figure 3.10.

The fractions of simulations resulting in an addition of the C-terminal β -strand region of a random peptide to the protofibril seed is summarized in Figure 3.9A. Both C_{2x} and C_{2z} forms of the seed are capable of propagating monomer additions in correct parallel arrangements to both the N- and C-terminal sheets on both ends of the fibril, though the amount of parallel N-terminal additions is greater than C-terminal additions

for all symmetries and end combinations. Similarly, both forms have some percentage of trajectories (~0.5%-1% for C-terminal and 0.5%-1.5% for N-terminal) that result in anti-parallel or incorrect additions that will result in lengthening the timescale for extending a stable fibril structure since the incorrect addition will have to be “annealed out” before the fibril can continue to grow. The greatest distinction between the fibrils can be seen in Figure 3.9B which plots the ratio of the parallel to anti-parallel additions depending on which end of the fibril for which additions occur. Though the addition to the C-terminal β -sheet for each symmetry and fibril end combination is similar, the B end of the C_{2z} form shows almost 4x the amount of N-terminal parallel addition vs anti-parallel addition, approximately twice as much as any other symmetry and end combination. Unlike the stability of the fibril ends, this addition metric distinguishes the C_{2x} and C_{2z} forms and demonstrates that monomer addition to amyloid fibrils may result in unidirectional or uneven growth from the fibril ends for C_{2z} , while we expect bidirectional growth for C_{2x} based on the results of our model shown in Figure 3.9.

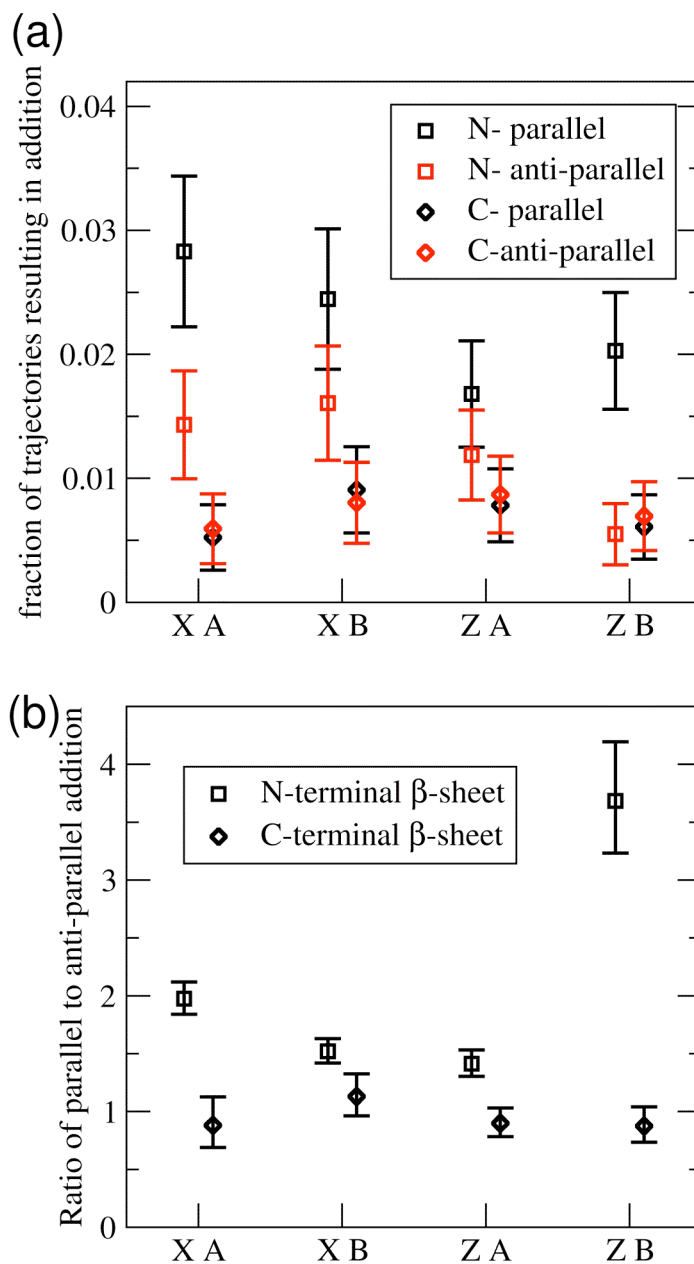


Figure 3.9: Monomer additions to protofibrils for C_{2x} and C_{2z} fibrils. (a) Fraction of trajectories resulting in partial parallel (black) and anti-parallel (red) additions to the N-terminal (\square) and C-terminal (\diamond) β -sheets. Error bars are standard deviation approximated from distributions with binary outcomes. (b) Ratio of partial parallel to anti-parallel additions to the N-terminal (\square) and C-terminal (\diamond) β -sheets. Error bars are 95% confidence interval for “relative risk” measure comparing binary outcomes.

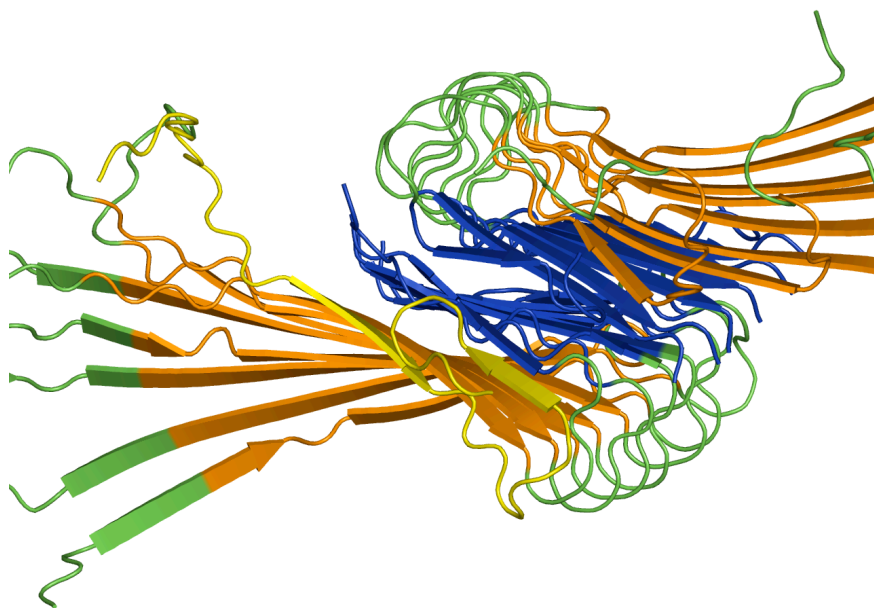


Figure 3.10: Example addition to fibril seed by free peptide. A peptide (yellow) with a random initial configuration without contacts with the seed is shown with partial in-register parallel addition to both N- and C- terminal β -sheets of the fibril seed.

The differences among quaternary symmetries and the possibility of unidirectional growth most likely arises from the effects of internal stagger that Tycko and coworkers have suggested for $A\beta_{1-40}$ based on the isotopic dilution experiments[13] and shown in Figures 3.2 and 3.4. We find that not only is there a different type of local symmetry splitting between the C_{2x} and C_{2z} forms like the experimental models involving +N or -N staggers, but we have shown that there are important local symmetry splitting in the exposed protofibril ends of the two proposed quaternary structures (Figure 3.4). We suggest that the end of the C_{2z} fibril with exposed N-terminal regions is better able to nucleate in-register parallel additions than the C-terminal exposed region because the hydrophobic, hydrophilic, and aromatic residue patterning in N-terminal residues 17-21 is accessible to the free peptide without non-specific hydrophobic interactions with the C-

terminal hydrophobic cluster. A free peptide approaching the exposed N-terminal region have more favorable interactions encouraging an in-register parallel addition – an anti-parallel arrangement does not result in as much favorable enthalpy (interchain hydrophobic interactions) for the same entropic cost (peptide backbone entropy), and therefore occurs relatively less frequently. When the N-terminal hydrophobic cluster of a free peptide approaches the end of a fibril with a buried N-terminal region and exposed C-terminal residues, the patterning of amino acids on the C-terminus is more generic so that both parallel and anti-parallel arrangements are equally likely.

We confirm this hypothesis by examining the parallel and anti-parallel additions to the C_{2x} fibril ends, which have one subunit with the N-terminal peptide exposed, and the other with the C-terminal peptide exposed. On both ends of the C_{2x} fibril, the N-terminal exposed subunit has only ~10% of the anti-parallel additions to that end, while it has ~25% of the parallel additions. We therefore suggest that the C_{2z} end B, where both subunits have N-terminal exposed regions, has greater in-register parallel addition due to the internal stagger. We also note that > 80% of the anti-parallel additions have the anti-parallel register of $A\beta_{16-22}$ fibrils, indicating anti-parallel β -sheet formation in this region, and exposing the limitations of the $A\beta_{16-22}$ fragment for understanding fibril elongation mechanisms.

3.3 Discussion and Conclusions

We have used a coarse-grained protein model to simulate $A\beta_{1-40}$ oligomers to determine both the critical nucleus and a minimum assembly of N-monomer cross-

sections of a mature fibril necessary for a structurally stable protofibril seed, and used that seed to measure fibril elongation propensities for different quaternary forms. Determining the critical nucleus as well as the minimum number of peptides necessary for a stable protofibril is an essential piece of information for experimentalists searching for signatures of these kinetic steps and the modeling of rate equations for aggregation kinetics, as well as theoreticians seeking to simulate a minimum size system capable of describing the structural properties of the mature fibril.

The first important conclusion of our study is to question the underlying assumption that the $A\beta_{1-40}$ nucleation event is a sudden transition from isolated disordered monomers to some minimal organization in either the monomer (i.e. a nucleated turn and/or β -strands) or intermolecular monomer-monomer interactions. Our approach instead works backward from an unambiguously ordered and stable protofibril with quaternary structure to see at what oligomer size do different levels of order breakdown. We find that the critical nucleus in our model corresponds to loss of quaternary order, i.e. a loss of registry in orientation of the two fibril halves, that partially destabilizes edge chains through loss of hydrophobic contacts from the other fibril sub-unit (Figure 3.11a). Below the critical nucleus the instability of edge chains is due to insufficient stabilization of inner monomer-monomer interactions within the same fibril half sub-unit as well (Figure 3.11b). In the linear regime of free energy represented by 16 chains (Figure 3.11c) quaternary order is well established and hence edge chain monomers are stabilized by ordered monomers in both fibril halves. This result implies a much greater level of structural order than usually assumed for the smallest oligomer

sizes, and seems consistent with the goals of a reductionist approach that seeks to determine whether structural order exists at the level of the monomer[78, 79, 98-100].

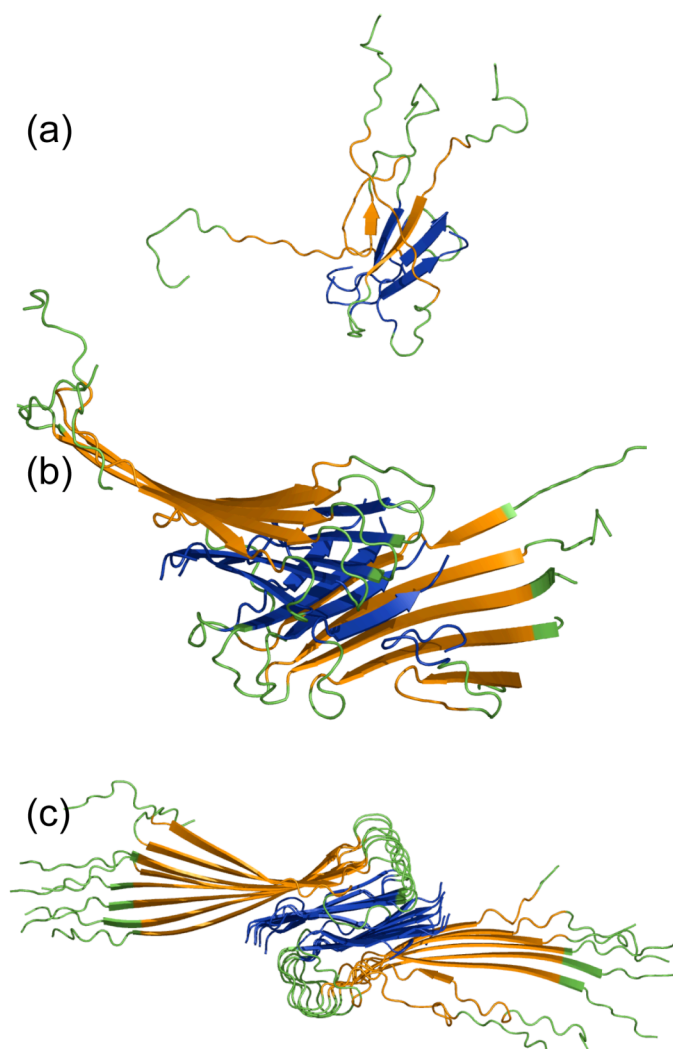


Figure 3.11: Comparing structural stability of example structures of varying oligomer length. Representative oligomer structures after 5000τ constant temperature simulations depicting greater structural stability as number of chains increases. (a) 4 chain simulation shows complete loss of fibril structure (b) 10 chain simulation shows that although a significant fraction of intermolecular β -sheet is retained, the fibril subunits rotate with respect to one another, leading to disorder and loss of contacts in the edge chains (c) 16 chain simulations shows retention of fibril order, and a clear fibril axis.

This alternate view of working backward from known order to disorder leads naturally to a means for determining free energy trends in nucleation-polymerization mechanisms to back out a critical nucleus size within our model. Based on equilibrium ensemble populations for a given proposed protofibril size, and analyzed with a fibril order metric χ_f , we can determine free energy barriers for nucleation of the thermodynamically scarce species that shifts the equilibrium from free monomer to stable, and therefore polymerizable, protofibril. We find a critical nucleus size of 10 chains, which is in pleasing agreement with the results by Teplow and coworkers which showed that kinetic models of amyloid formation fit time course data when the number of chains involved in the aggregation nucleus for A β aggregation is set to 10 chains [21]. Below this critical nucleus the edge chains of the 4-, 6-, and 8-chain protofibril structure are unstable, shifting the population from ordered fibril, C_n to populations with increased free monomer C_{n-1} . The true protofibril state, i.e. the minimum size protofibril capable of elongation, is ~ 16 chains, and subsequent monomer additions involve a constant gain in free energy that are insensitive to the number of interior chains.

Beyond this minimum stable seed size, adding chains to the protofibril does not increase the stability of the outermost chains, which we have shown correlates with a leveling off of the hydrophobic residue density and helps compensate for the unfavorable entropy of ordering the monomers in the sub-unit halves as well as the two subunit halves with respect to each other. This is an intuitive result because the hydrophobic interactions are thought to stabilize amyloid fibril structures and once a fibril reaches a certain length, the average hydrophobic residue density should be a constant. Hecht and coworkers have

also demonstrated that the two additional hydrophobic residues (isoleucine and alanine) at the C-terminus of $A\beta_{1-42}$ are responsible for increased fibril forming propensity of $A\beta_{1-42}$ compared with $A\beta_{1-40}$ [73]. These additional hydrophobic residues would shift the plateauing of hydrophobic density to smaller oligomer sizes for $A\beta_{1-42}$. In turn, we hypothesize that there is a corresponding shift in the size of a viable seed for fibril formation *in vitro* to fewer numbers of chains relative to $A\beta_{1-40}$ [71-73, 101] and which may be correlated with greater disease virulence *in vivo* [101-103] of the $A\beta_{1-42}$ versus $A\beta_{1-40}$ sequences. The calculation of hydrophobic density therefore may be predictive for the size of critical nucleus and/or protofibril regime for any new $A\beta$ fragment or mutation, or other systems which assemble into fibrils in their aggregated state.

For proposed fibril models with no quaternary structure, such as that suggested by Lührs et al [104] for methionine sulfoxide 35 (Met35ox) mutants of $A\beta_{1-42}$, the χ_f metric will overestimate the critical nucleus size since it is sensitive to quaternary disorder due to rotations of the fibril halves. However, for a protofibril with a cross section of one peptide, only half the number of chains would be necessary to reach the plateau in hydrophobic density compared to a protofibril with a two peptide cross section since the hydrophobic density of all the peptides in the protofibril are now averaged over a single subunit. Thus for any $A\beta$ system that does not form quaternary structure, we would predict a reduction in the size of the critical nucleus and protofibril capable of elongation relative to $A\beta_{1-40}$, though the height of the free energy barrier may be higher. However whether the Met35ox mutant is a good model for $A\beta_{1-42}$ is open to question, since the oxidized methionine residue is a disruptive mutation for stabilizing the hydrophobic

interface of the two halves of the quaternary structure. Thus the best model for WT A β ₁₋₄₂ likely remains one in which some type of quaternary structure is present as it is for A β ₁₋₄₀.

The number of chains where the free-energy for addition of another monomer becomes constant corresponds to the point of minimum protofibril size where the edges and quaternary structure are stable and behave like a long fibril. In our model this point corresponds to ~16 chains which we use for characterizing the propensity to add additional monomers to understand polymerization in more molecular detail and to compare different quaternary forms. Based on our simulations, which quantify parallel additions and “incorrect” anti-parallel additions for both fibril seed ends, both C_{2x} and C_{2z} are capable of elongation. However we suggest that C_{2z} may be the dominant amyloid fibril form because it can more readily propagate in-register parallel addition with a much lower “error” through anti-parallel additions. The most recent NMR studies have suggested that the quaternary structure of one particular fibril of A β ₁₋₄₀ involves the C_{2z} form [13].

One possible reason for this preference for C_{2z} over C_{2x} most likely arises from a topological “frustration” for addition to C_{2x}. In the C_{2x} form, the C-terminal β -strands of the peptides in both subunits are parallel; the same is true for the C_{2x} N-terminal β -sheets in both subunits. In contrast, the N-terminal (and C-terminal) β -strands are anti-parallel in the C_{2z} form. If the C-terminus of a free peptide approaches the C_{2x} form fibril in an orientation anti-parallel to the C-terminus of one of the edge peptides (i.e. in a direction not suitable to extending the in-register parallel structure of the fibril), the free peptide is

similarly anti-parallel to the C-terminus of the other exposed fibril peptide. In this case it must break any favorable interactions (i.e. hydrophobic clustering) and flip orientation in order to continue the correct fibril elongation. For the C_{2z} , conversely, if a free peptide approaches in a configuration anti-parallel to one of the fibril peptides, it is parallel to the other exposed fibril peptide terminus and can form in-register parallel β -sheet contacts with the appropriately oriented exposed fibril peptides thus incorporating the free peptide into the fibril. However, our simulations are much too short to observe the multiple binding and unbinding events would be necessary to demonstrate that this mechanism contributes to the difference in additions between C_{2x} and C_{2z} .

We have concluded from our simulation data that it is the structural consequences of the internal stagger[13] that results in higher rates of in-register parallel addition, and more importantly on fewer growth-halting anti-parallel additions for the C_{2z} fibril with distinct ends. This opens up the possibility for unidirectional growth of the protofibril for this quaternary form, while we expect bi-directional growth for C_{2x} based on our model. Although NMR experiments support a +2 or -2 stagger, our model naturally equilibrates to a different STAG value of -1, and analysis of the NMR data for $A\beta_{1-40}$ does not rule out the possibility of a mixed stagger, i.e. +N inter-digitation for one fibril sub-unit half and -N inter-digitation for the other fibril sub-unit. Model building shows that a mixed stagger quaternary structure for the C_{2z} form symmetrizes the fibril ends, while it results in end asymmetry of the C_{2x} form (Figure 3.4), thereby reversing the structural end symmetries of the two quaternary forms and potentially their elongation mechanism.

Although our model and experiment show both C_{2z} and C_{2x} are viable quaternary structures, and that mixed staggers are theoretically possible, we advance the much more speculative conclusion that macroscopic morphology differences in the mature fiber may be due to different quaternary and stagger configurations that affect directionality in fibril growth. We note that $A\beta_{1-40}$ is known to form fibrils of at least two distinct architectures[81] which may be accounted for by distinct staggers and quaternary forms. Finally we end the discussion by noting that the finite length of our simulation times makes the absolute percentages of any type of monomer addition rather low ($\sim 3\%$), although it may increase with longer simulation runs. However it opens up the question as to whether the $A\beta$ monomer is the dominant unit for fibril elongation. One direction we will pursue is whether small oligomers are more viable addition units for fibril elongation which has been suggested by Kaye and co-workers[105].

3.4 Methods

Coarse-Grained Protein Model

The three-flavor coarse-grained model we developed has been used to study the folding and aggregation propensities of members of the ubiquitin α/β fold class[7, 92-96]. We have recently updated the minimalist model to improve the faithfulness to real proteins while retaining its simplicity, which we describe here.

To better discriminate between hydrophobic residues of different sizes, we have updated the model to allow four flavors, consisting of large hydrophobic, small hydrophobic, neutral/small hydrophilic, and large hydrophilic, designated B, V, N, and L

respectively. The amino acid sequence of the A β peptide was mapped to its four-flavor sequence using the mapping shown in Table 3.2.

20	4	20	4	20	4	20	4
Trp	B	Met	B	Gly	N	Glu	L
Cys	B	Tyr	B	Ser	N	Asp	L
Leu	B	Val	V	Thr	N	Gln	L
Ile	B	Ala	V	Lys	N	Asn	L
Phe	B	Pro	N	His	N	Arg	L

Table 3.2: Mapping 20-letter (20) amino acid code to coarse-grained four-letter (4) code

The Hamiltonian governing the interactions in the system is given by

$$\begin{aligned}
 H = & \sum_{\text{angles}} \frac{1}{2} k_{\theta} (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \left[A[1 + \cos(\phi + \phi_0)] + B[1 - \cos(\phi + \phi_0)] + \right. \\
 & \left. C[1 + \cos 3(\phi + \phi_0)] + D[1 + \cos(k\phi + \phi_0 + \frac{\pi}{4})] \right] \\
 & + \sum_{\substack{i, j \geq i+3 \text{ for intrachain} \\ i, j \text{ for interchain}}} 4\epsilon_H S_1 \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - S_2 \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \sum_{\substack{i, j \geq i+3 \text{ for intrachain} \\ i, j \text{ for interchain}}} U_{HB}
 \end{aligned} \tag{3.2}$$

where θ is the bond angle, ϕ is the dihedral angle formed by 4 consecutive C_{α} positions, and r_{ij} is the distance between beads i and j . ϵ_H sets the energy scale and gives the strength of the large (B) hydrophobic contact. The bond angle term is a stiff harmonic potential with a force constant of $k_{\theta} = 20 \epsilon_H / \text{rad}^2$, and the optimal bond angle θ_0 is set to 105° . Each dihedral angle in the chain is designated to be either structured (S), a weighted sum of helical and extended potentials, or turn (T) primarily placed in regions of the peptide containing glycine to account for the greater backbone mobility. The

parameters A, B, C, and D are chosen to produce the desired minima (see Table 3.3), and ϕ_0 is set to 0.17 for the helical portion of the helical (H) portion of the S potential and to -0.35 for the extended (E) portion (see Table 3.3). The k parameter is set to 1 for all dihedral potentials in this study.

Dihedral Type		A (ϵ_H)	B (ϵ_H)	C (ϵ_H)	D (ϵ_H)	ϕ_0 (rad)	Local minima (global minima in bold)
S = Σ {	0.3 * H (Helical)	0	1.2	1.2	1.2	+0.17	-70°, + 50 °, 170°
	0.75 * E (Extended)	0.9	0	1.2	0	-0.35	-165 °, -45°, +75°
T	Turn	0.2	0.2	0.2	0.2	0	-60 °, 0 °, + 60 °

Table 3.3: Parameters for Dihedral Types

The third term in Equation 3.2 represents non-bonded interactions, and is determined according to the bead flavors B, V, N and L: $S_1 = S_2 = 1$ for B–B interactions; $S_1 = S_2 = 0.5$ for all V–V and V–B interactions; $S_1 = 1/3$ and $S_2 = -1$ for L–L, L–V and L–B interactions; and $S_1 = 1$ and $S_2 = 0$ for all N–X interactions. Of the possible interaction combinations, attractive potentials result for interactions between hydrophobic beads (B–B, V–B and V–V interactions). The interactions among all other combinations of beads are repulsive, although the form of repulsion depends on the bead types involved. The sum of van der Waals radii σ is set at 1.095 to mimic the large excluded volume of amino acid side chains. This term is evaluated for all bead pairs within a distance cutoff of 5.5 units (21Å).

The last term in Equation 3.2 describes a direction-dependent hydrogen bond interaction to better represent the interactions that form and stabilize β -sheets and α -helices. The functional form is inspired by the Mercedes Benz (MB) model of water,

introduced by Ben-Naim[106] and further developed by Dill and coworkers[107]. The hydrogen bond potential between two residues i and j is given by:

$$U_{HB} = \varepsilon_{HB} F(r_{ij} - r_{HB}) G(|\mathbf{t}_{HB,i} \cdot \mathbf{u}_{ij}| - 1) H(|\mathbf{t}_{HB,j} \cdot \mathbf{u}_{ij}| - 1) \quad (3.3)$$

where

$$F(r_{ij} - r_{HB}) = \exp\left[-\frac{(r_{ij} - r_{HB})^2}{\sigma_{HBdist}^2}\right], \quad (3.3a)$$

$$G(|\mathbf{t}_{HB,i} \cdot \mathbf{u}_{ij}| - 1) = \exp\left[\frac{|\mathbf{t}_{HB,i} \cdot \mathbf{u}_{ij}| - 1}{\sigma_{HB}^2}\right], \quad (3.3b)$$

$$H(|\mathbf{t}_{HB,j} \cdot \mathbf{u}_{ij}| - 1) = \exp\left[\frac{|\mathbf{t}_{HB,j} \cdot \mathbf{u}_{ij}| - 1}{\sigma_{HB}^2}\right] \quad (3.3c)$$

The hydrogen bond strength is modulated by the value ε_{HB} , and is set at $1.6\varepsilon_H$. The distance-dependent term F is a Gaussian function centered at the ideal hydrogen bond distance r_{HB} , set to 1.125 in accordance with our survey of PDB structures. For the direction-dependent terms G and H , we use a modified exponential instead of a Gaussian function to smoothen the potential energy surface. The vectors $\mathbf{t}_{HB,i}$ and $\mathbf{t}_{HB,j}$ are unit normal vectors for the planes described by $(i-1, i, i+1)$ and $(j-1, j, j+1)$ respectively, and \mathbf{u}_{ij} is the unit vector between residues i and j . The width of functions F , G and H are set

by $\sigma_{hBdist} = 0.5$ and $\sigma_{HB} = 0.45$. The hydrogen bond potential is evaluated for all i-j bead-pairs within a cutoff distance of 3.0 units.

Model Building

A model of an amyloid fibril building block was constructed in single-bead representation according to the constraints specified by Petkova et al[82]. Since this model is a single-bead representation of a protein, the (ϕ, ψ) angle constraints were converted into local secondary structure assignments and then applied to the model. The 20-letter sequence of the $A\beta_{1-40}$ peptide and the corresponding coarse-grained (CG) primary and secondary structure is:

1°sequence	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
1°sequence (CG)	LVLBLNLNNBLVNNLNBVBBVLLVNNLNNVBBNBBVNNV
2°structure (CG)	SSSSSSSSSSSSSSSSSSSSSSSTTSSTTSSSSSSSTT

To construct the amyloid fibril building block, in-register parallel intermolecular β -sheet models were made with 40 starting chains, one for the C_{2x} and one for C_{2z} form. Each strand in the models contains a disordered N-terminal region (residues 1-9), an N-terminal β -sheet region (residues 10 to 24), a turn region (residues 25-29), and a C-terminal β -sheet region (residues 30-40). In comparison to the model of a fibril presented in Petkova et al., we have the C-terminal β -strand “flipped” in orientation, where the residues packed against the N-terminal β -strand are even numbered, as determined by the most recent NMR data[13]. Models were built with N- and C- terminal strands without

stagger, but interdigitation of structures into staggered structures can be seen in equilibrated structures at finite temperature.

Two possible conformers of the amyloid fibril were built. Given that z is defined as the fibril axis and x as the direction of the β -strands, fibril models with approximate C_2 symmetry around each of these axes, named C_{2z} and C_{2x} respectively, were constructed. Models for each different seed size (4, 6, 8, 10, 12, 14, 16, 18, 20) were created by retaining the inner-most chains from the equilibrated 40 chain seed starting structures of the two symmetries. The outer-most chains were discarded to ensure that edge effects (loss of perfect fibrillar order of the exterior chains) were not incorporated into the seeds. Once equilibrated, the beads representing the N- and C- terminal β -sheets inter-digitate to form contacts internal to each subunit of the fibril with a particular value of “stagger” as shown in Figure 3.2. The most recent solid state NMR work has suggested that the stagger is either STAG(+2) or STAG(-2), although our models under thermal equilibration give STAG(-1).

Simulation Protocol

We use constant-temperature Langevin dynamics with friction parameter $\zeta = 0.05$. Bond lengths are held rigid by using the RATTLE algorithm[108]. All simulations are performed in reduced units, with mass m , energy ϵ_H , and k_B all set equal to unity. The forty chain C_{2x} and C_{2z} fibril models were equilibrated with Langevin dynamics at a temperature of 0.45 for 1500τ (300,000 steps). This procedure was repeated 24 times so that the stochastic dynamics generated 24 equilibrated starting structures of a 40 chain fibril seed for C_{2x} and C_{2z} . Three to five simulations of each of 24 models of each

symmetry were run for 5000τ (1,000,000 steps) at $T^*=0.45$ ($T \approx 337\text{K}$). The reported protofibril stability data are based on statistics collected approximately 120 independent simulations per chain number and symmetry. Statistics on the chain conformation were gathered every 50τ (10,000 steps). Structural stability for each time point was quantified by the χ_f parameter:

$$\chi_f = \frac{1}{M} \sum_{\alpha=1}^4 \sum_{\beta>c}^4 \sum_i^N \sum_j^N h\left(\varepsilon - |r_{\alpha,i;\beta,j} - r_{\alpha,i;\beta,j}^0|\right) \quad (3.4)$$

The metric is the sum over bead i on chain α and bead j on chain β (where i and j range from ([17-21] [30-34]) and α and β range over the 4 chains making up the exterior and neighboring chains on each end), h is the Heaviside step function, ε is the tolerance set to 0.5 distance units ($\sim 1.9\text{\AA}$), $r_{\alpha,i;\beta,j}$ is the distance between bead i on chain α and bead j on chain β , and $r_{\alpha,i;\beta,j}^0$ is the pair distance in the initial structure, and M is a normalizing constant counting the total number of pairs (in this case it is equal to 600).

To investigate the addition of monomers to the protofibril seeds of different symmetries, monomer simulations decorrelated at high temperature ($T^*=1.0$) and equilibrated at $T^*=0.45$ as single chains (infinite dilution) were placed as described in the Results at the ends of equilibrated protofibrils of length 16 chains.

3.5 Acknowledgements

We thank Troy Cellmer and Harvey Blanch for useful discussion throughout this study, and Elizabeth Verschell for assistance with the coarse-grained force field. We

thank Robert Tycko for useful discussion and for the atomic coordinates for his models of A β ₁₋₄₀. We also thank Kevin Kohlstedt, a DOE Computational Science Graduate Fellow, for careful reading of the manuscript during his internship at Berkeley Lab. N.L.F. thanks the Whitaker Foundation for a graduate research fellowship. Y.O. thanks the Guidant Foundation for a summer research fellowship. THG gratefully acknowledges a Schlumberger Fellowship while on sabbatical at Cambridge University. Molecular graphics for this paper created in PyMOL [109]. This work was supported in part by a grant from the NIH.

Chapter 4: Protofibril Assemblies of the Arctic, Dutch and Flemish Mutants of the Alzheimer's A β ₁₋₄₀ Peptide

Reproduced with permission from *Biophysical Journal*. 2007 Nov 21. Electronic publication ahead of print. doi:10.1529/biophysj.107.121467 Copyright 2007 The Biophysical Society.

Nicolas Lux Fawzi¹, Kevin L. Kohlstedt³, Yuka Okabe², and Teresa Head-Gordon^{1,2}

¹UCSF/UCB Joint Graduate Group in Bioengineering, Berkeley, CA 94720

²Department of Bioengineering, University of California, Berkeley, CA 94720

³Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208

Abstract

Using a coarse-grained model of the A β peptide, we analyze the Arctic (E22G), Dutch (E22Q), and Flemish (A21G) familial Alzheimer's disease (FAD) mutants for any changes in the stability of amyloid assemblies with respect to the wild-type (WT) sequence. Based on a structural reference state of two protofilaments aligned to create the "agitated" protofibril as determined by solid-state NMR, we determine free energy trends for A β assemblies for the WT and FAD familial sequences. We find that the structural characteristics and oligomer size of the critical nucleus vary dramatically among the hereditary mutants. The Arctic mutant's disorder in the turn region introduces new stabilizing interactions that better align the two protofilaments, yielding a well-defined protofibril axis at relatively small oligomer sizes with respect to WT. By contrast, the critical nucleus for the Flemish mutant is beyond the 20 chains characterized in this study, thereby showing a strong shift in the equilibrium toward monomers with respect to larger protofibril assemblies. The Dutch mutant forms more ordered protofilaments than

WT, but exhibits greater disorder in protofibril structure that includes an alternative polymorph of the WT fibril. An important conclusion of this work is that the Dutch mutant does not support the agitated protofibril assembly. We discuss the implications of the structural ensembles and free energy profiles for the FAD mutants in regards to interpretation of the kinetics of fibril assembly using chromatography and dye-binding experiments.

4.1 Introduction

Alzheimer's disease (AD) is characterized by the appearance of neuritic plaque deposits comprised primarily of amyloid β peptide[2], whose chemophysical properties are central to understanding the disease state. Amyloid β is created by proteolytic cleavage of the amyloid precursor protein (APP), as a 40 or more virulent 42 residue sequence ($A\beta_{1-40}$ or $A\beta_{1-42}$) with unknown function[2, 69]. While many familial Alzheimer's disease (FAD) mutants of the APP protein are external to the $A\beta$ peptide sequence and typically influence $A\beta$ processing, a set of mutants that cluster near amino acid positions 21 through 23 in the amyloid β peptide itself have drawn special attention due to possible changes of peptide biochemistry[2]. Some of the most well-studied FAD mutants of amyloid β include the Dutch (E22Q)[15, 18], Flemish (A21G)[14, 17], Italian (E22K)[110, 111], Arctic (E22G)[16], Iowa (D23N)[112], and double Dutch/Iowa mutants (E22Q, D23N)[112], all of which have been characterized for both $A\beta_{1-40}$ and $A\beta_{1-42}$ both *in vitro* and *in vivo*. In spite of the locality of the mutation, the FAD mutants show dramatic diversity in presence or absence of AD dementia symptoms and intracerebral hemorrhaging[112], exhibit variations in $A\beta_{1-42}$ levels in media from cells

transfected with a given mutant[16], and show strong differences in the regions of the brain tissue or vasculature in which amyloid plaques are deposited[16, 113].

More relevant to this study are the strong differences in the kinetics of the formation of fibril assemblies of WT and mutant A β that make up the amyloid plaque[19-22]. *In vitro* studies have found that the Dutch mutant nucleates and fibrillizes more readily than WT, that the Arctic mutation has a higher propensity to form protofibrils (either distinct from, or precursors to, the fibril state) although fibrillization rates are comparable to WT, while the rate of fibril formation is greatly reduced for the Flemish mutant relative to WT[16]. Morelli et al. showed that proteolytic enzymes more easily degrade monomeric WT A β , Italian, and Flemish mutants, while proteolysis of the Arctic and Dutch mutant protein is not as efficient- likely due to their rapid sequestration into protofibril or fibril morphologies that inhibit degradation by the enzyme[114]. While *in vitro* experiments have shown that different polymorphs of the mature A β fibril can contribute to variation in cell viability[81], and synaptic activity is greatly impaired in the presence of the insoluble plaque[115], biochemical evidence is accumulating that immature and/or soluble oligomer states may be the more prevalent cytotoxic species[116-120]. Again the FAD mutants show distinct differences; cognitive deficits arising from the Arctic mutant were traced to a non-fibrillar form of the A β peptide, whereas the severity of memory loss symptoms for carriers of the Dutch mutation were consistent with interference from the mature fibrillar A β species[121].

A convenient separation of the soluble oligomers and mature fibril regimes may be gleaned from the mechanism of fibrillization of full length WT A β peptides which has

been shown to follow a nucleation-dependent polymerization mechanism[21, 74, 80, 122]. The kinetic model developed by Ferrone[123] assumes that the observed lag phase is due to the formation of a critical nucleus- the assembly of monomers into a certain oligomer size corresponding to the largest free energy barrier- beyond which a gradient of favorable free energy or “down-hill” polymerization progresses into a mature fibril. However, the structural characteristics and oligomer size of the soluble nucleating species have yet to be determined experimentally for either the WT or familial mutants, and the mechanism of polymerization that eventually delineates a mature fibril is unclear.

A number of important computational studies have addressed the monomer conformation and oligomers assemblies of the WT and FAD mutants, both on the full length sequence as well as A β fragments[87-90, 124-134]. We have chosen in this study to focus on the A β_{1-40} peptide since the best quality experimental structural data is available for this system[13, 81-83]. Given the ability of A β_{1-42} and A β_{1-40} to cross seed fibril growth, we believe the A β_{1-40} structure is relevant to the fibril form of A β_{1-42} .

We have recently developed[135] and used a coarse-grained protein model to characterize the critical nucleus, structural stability, and fibril elongation propensity of WT A β_{1-40} protofibrils[136]. We pursue a coarse-grained C $_{\alpha}$ model unlike previous all-atom studies because these models enable us to retain physico-chemical interactions through model physics faithful to the true system while enabling a full statistical characterization of the ensemble properties for each mutant, something not attainable for much more computationally expensive all atom models. These coarse-grained models capture both sequence specific interactions and geometrically accurate α -helical and β -

sheet secondary structure geometries while retaining the simplicity of a C_α protein model (see Methods section). By careful parameterization of the interaction potentials between coarse grained amino acid positions, these models capture well the excluded volume and hydrophobic interactions of the true system. The inclusion of a direction dependent backbone hydrogen bond potential enables the model to capture cooperative assembly of secondary structures with faithful β -sheet geometry resulting in native state RMSD of $\sim 3\text{\AA}$ for globular proteins relative to experimental NMR structures[135]. Although favorable opposite charge interactions can be modeled as attractive interactions, this model is limited by the lack of explicit electrostatic interactions and we leave the analysis of specific charged interactions to future studies with an enhanced model. We have previously characterized this model for the WT sequence and perform all analysis of mutants as comparisons to this sequence. Within our model we have represented two different quaternary symmetry forms proposed by solid state NMR for so-called “agitated” fibrils[13, 81-83]. As shown in Figure 4.1, the cross section of the fibril is made up of two “U-shaped” monomers with hydrophobic C-terminal regions in van der Waals contact in a pseudo-symmetry C_{2z} form, and larger (proto)fibril peptide assemblies propagate this dimer motif down the fibril axis. Our nomenclature is to define a (proto)fibril as being composed of two (proto)filaments of in-register parallel intermolecular N-terminal and C-terminal β -sheet regions that can be organized by the C_2 symmetry operation about the fibril z-axis (see Figure 4.1). *Protofibril* refers to a fibril that is well below micron size lengths.

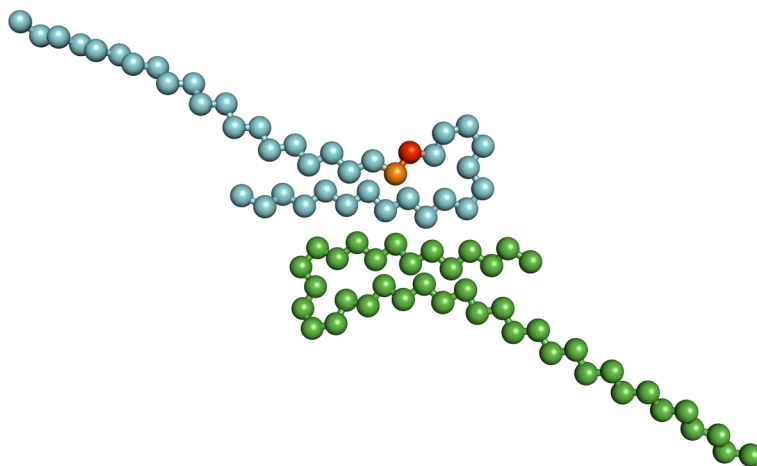


Figure 4.1: Ideal cross-section of agitated fibril morphology. The two monomer cross-section of the bead model with C_{2z} symmetry based on SS-NMR data [17, 31-33] after equilibration within a large WT fibril (40 chain). Ala 21, site of the Flemish mutation, is pictured in orange on one monomer. Glu 22, site of Arctic and Dutch mutations, is pictured in red.

Given this model, we calculated equilibrium populations of structurally stable and unstable protofibrils for WT $A\beta_{1-40}$ as a function of the number of dimer cross-sections, and evaluated a free energy profile for monomer-protofibril equilibrium[136]. We determined a critical nucleus of 10 chains for WT $A\beta_{1-40}$, characterized as having well formed intermolecular β -sheets, but lacking structural integrity at the C-terminal interface so that the protofilaments do not align along the fibril axis[136]. Beyond the critical nucleus we found that 16 monomer chains showed the fibril extension propensities of a mature fibril, for the reason that a sufficient hydrophobic density is reached to stabilize the C-terminal interface and therefore align the protofilaments along the fibril axis. At this length the $\Delta\Delta G$ for cross-section addition is a constant and defines polymerization equilibrium as shifted strongly in favor of the fibril form. Thus the oligomer size below or above 10 chains for WT $A\beta$ delineate the concentration conditions for further study as

to whether the cytotoxic species correspond to early soluble aggregates or mature insoluble (proto)fibrils[136].

In this work we extend our coarse-grained model study of amyloid assemblies of WT A β to the Dutch, Flemish, and Arctic FAD mutants. Our work starts with the assumption that the best experimental model of the complete WT A β_{1-40} amyloid fibril, derived from fibrils prepared under “agitated” conditions[13, 81-83], is also an appropriate structural model of these FAD mutants. An important component of this work is to test whether the agitated fibril morphology is an appropriate model for mutations at positions 21 and 22. Given the importance of the salt bridge defined by D23 and K28 in stabilizing the agitated fibril assemblies, the Italian and Iowa FAD mutations at position 23 that lose the ability to neutralize charge will be unlikely to conform to the reported SS-NMR agitated structure[13], and are not considered in this study but have been recently examined by Zheng et al [134].

We find that the free energy trends for A β assemblies among the familial and WT sequences show that the structural characteristics and size of the critical nucleus shifts dramatically among the mutants, even though the single point mutations are localized in the same region of the A β peptide. The Arctic mutant’s disorder in the turn region introduces new stabilizing interactions that better align the two protofilaments to yield a well-defined fibril axis. By contrast, we find that the critical nucleus for the Flemish mutant is beyond the 20 chains characterized in this study, thereby indicating a strong shift in the equilibrium toward monomers with respect to larger ordered protofibril assemblies. We find that the Dutch mutant forms more ordered protofilaments than WT,

but more disorder in protofibril structure that includes an alternative polymorph of the WT fibril. We discuss the implications of the structural ensembles and free energy profiles for the FAD mutants in regards to interpretation of the kinetics of fibril assembly using chromatography and dye-binding experiments[16, 137].

4.2 Methods

Coarse-Grained Protein Model

The coarse-grained model we developed has been used to study the folding and aggregation properties of members of the ubiquitin α/β fold class[7, 92-96], and we have recently updated it to improve its faithfulness to real proteins while retaining its simplicity[135]. The coarse-grained model consists of an unbranched chain of beads, each representing a single amino acid. Beads are assigned interaction type and strength using a Lennard-Jones functional form based on a mapping from the 20 amino acids to our 4 bead types: B, strong attraction; V, weak attraction; N, weak repulsion; L, strong repulsion. Interactions between beads 3 or more positions apart are represented by potentials of mean force corresponding to bead flavor, and solvation water is treated implicitly by incorporating favorable interactions between hydrophobic groups. Bonds between beads are kept rigid at 1 reduced distance unit (approximately 3.8Å) representing the distance between C_α positions in a peptide chain. Angles formed by three consecutive beads are represented by a harmonic potential with mean 105° , the average of the C_α pseudo bond-angle in extended and helical secondary structures. A single torsional potential, “S,” which has competing minima for helical ($\sim 60^\circ$) and beta-sheet ($\sim 180^\circ$) arrangements is applied for every dihedral angles formed by four

consecutive beads with the exception of some of the dihedral angles where one of the central beads is a glycine in the 20 amino acid sequence. For these dihedrals, we replace the helical/extended torsional potential with a “floppy” potential, “T,” where the barriers to transition between helical and extended type angles are reduced, in this way capturing the greater conformational flexibility of the peptide chain near glycine residues. We have used this model to differentiate sequence driven folding mechanisms of proteins L and G with $\sim 3\text{\AA}$ RMSD models to the native PDB structures[135], as well as determine the critical nucleus and fibril elongation propensity of the wildtype $A\beta_{1-40}$ fibrils. We refer the reader to our recent work [135, 136] for full specification of the model applicable to this study.

A model of an amyloid dimer cross-section was constructed in the single-bead representation of our model according to the constraints specified by Petkova and co-workers[82]. Since this model is a single-bead representation of a protein, the (ϕ, ψ) angle constraints were converted into local secondary structure assignments and then applied to the model. The resulting 20-letter sequence of the WT $A\beta_{1-40}$ peptide and the corresponding coarse-grained (CG) primary and secondary structure are:

1°sequence	DAEFRHDSGYEVHHQKLVFF AE DVGSNKGAIIGLMVGGVV
1°sequence (CG)	LVLBLNLNBLVNNLNBVBB VL LVNNLNNVBBNBBVNNVV
2°structure (CG)	SSSSSSSSSSSSSSSSSS SSS STTSSTTSSSSSSSTT

We highlight in bold what aspects of the model change under the Arctic (E22G) or Flemish (A21G) mutations. The amino acid sequence and secondary structure assignment for our model of the Flemish $A\beta$ peptide (A21G) is:

1°sequence DAEFRHDSGYEVHHQKLVFF**G**EDVGSNKGAI IGLMVGGVV
 1°sequence (CG) LVLBLNLNBLVNNLNBVBB**N**LLVNNLNNVBBNBBVNNVV
 2°structure (CG) SSSSSSSSSSSSSSSSSSS**TT**SSSTTSSTTSSSSSSSTT

which changes a bead with small attraction to one of small repulsion, while making the dihedral angles in that vicinity of the chain floppier given the greater conformational flexibility of the glycine backbone. Correspondingly, the amino acid sequence and secondary structure assignment of the Arctic A β peptide (E22G) is:

1°sequence DAEFRHDSGYEVHHQKLVFFA**G**DVGSNKGAI IGLMVGGVV
 1°sequence (CG) LVLBLNLNBLVNNLNBVBBV**N**LVNNLNNVBBNBBVNNVV
 2°structure (CG) SSSSSSSSSSSSSSSSSSS**TT**SSSTTSSTTSSSSSSSTT

This also makes the dihedral angles floppier, but in a region of the chain shifted by one amino acid, while at the same time changing a more strongly repulsive bead interaction to a weaker one. Finally, since the Dutch A β peptide mutation (E22Q) does not involve a mutation to glycine, the glutamine mutation is represented only at the level of a primary sequence:

1°sequence DAEFRHDSGYEVHHQKLVFFA**Q**DVGSNKGAI IGLMVGGVV
 1°sequence (CG) LVLBLNLNBLVNNLNBVBBV**V**LVNNLNNVBBNBBVNNVV
 2°structure (CG) SSSSSSSSSSSSSSSSSSSSSSSSTTSSTTSSSSSSSTT

in which an L bead that describes repulsion due to the alignment of negative charge down each of the protofilaments is changed to a V bead that qualitatively makes the interactions attractive. We justify this change from L bead to attractive V to represent E22Q by noting that glutamine-glutamine interactions of this geometry are favorable in polyglutamine aggregates[138]. Globular proteins that contain a sequence run of glutamines are known to form β -sheets that are stabilized by hydrogen bonds between

carbonyl and amide moieties of the glutamine side chain chemistry as well as hydrophobic interactions between aligned non-polar regions of the glutamine side-chains, and polyglutamine fibrils show similar hydrogen bonding patterns that stabilize the intermolecular assemblies[138].

These three mutations, Flemish, Arctic and Dutch have clear coarse-grained bead and dihedral mutations which we have described above. Representing change of charge mutations from acidic to basic amino acids such as the Italian mutation (E22K) are difficult to represent in our current formulation of the coarse-grained model which does not include explicit treatment of electrostatic effects. We believe that the mutations we have pursued here capture the type of change that would be seen in the full atomistic system since the perturbation of the WT system, to which we always compare in its coarse-grained form, is well represented by the mutations. We note that the Arctic and Dutch mutations do result in a change in the charge of the peptide which undoubtedly influences monomer and disordered oligomer thermodynamics and dimerization kinetics in a pH and salt dependent manner. Our present study, however, aims to examine how these mutations, local changes to the sequence represented well by bead and backbone dihedral angle differences, affect the global structure and thermodynamics of the protofibril and fibril assemblies in standard physiological buffer conditions for which our original model has been parameterized.

Model Building

To construct the amyloid fibril, in-register parallel intermolecular β -sheet models were made with 40 starting chains for the C_{2z} form. Each strand in the models contains a

disordered N-terminal region (residues 1-9), an N-terminal β -sheet region (residues 10 to 24), a turn region (residues 25-29), and a C-terminal β -sheet region (residues 30-40). In comparison to the original model of a fibril presented by Tycko and coworkers [82], we have the C-terminal β -strand “flipped” in orientation, where the residues packed against the N-terminal β -strand are even numbered, as determined by the most recent NMR data [13]. Models were built with N- and C- terminal strands without stagger, but interdigitation of structures into staggered structures can be seen in equilibrated structures at finite temperature. Once equilibrated, the beads representing the N- and C- terminal β -sheets inter-digitate to form contacts internal to each subunit of the fibril with a particular value of “stagger” [13]. The most recent solid state NMR work has suggested that the stagger is either STAG(+2) or STAG(-2) [13], although our models under thermal equilibration give STAG(-1) [136]. Models for different seed sizes (4, 6, 8, 10, 12, 14, 16, 18, 20) were created by retaining the inner-most chains from the equilibrated 40 chain starting structures to ensure that edge effects (loss of perfect fibrillar order of the exterior chains) were not incorporated into the seeds.

Simulation Protocol

We use constant-temperature Langevin dynamics with friction parameter $\zeta = 0.05$. Bond lengths are held rigid by using the RATTLE algorithm[108]. All simulations are performed in reduced units, with mass m , energy ϵ_H , and k_B all set equal to unity. The 40 chain C_{2z} fibril models were equilibrated with Langevin dynamics at a temperature of 0.45 for 1500τ (300,000 steps). This procedure was repeated between 50 (Arctic and

Flemish) and 100 (WT and Dutch) times so that the stochastic dynamics generated 50 or 100 equilibrated starting structures of a 40 chain fibril seed for C_{2z} . One to three simulations of each of the 50-100 models were run for 5000τ (1,000,000 steps) at $T^*=0.45$ ($T \approx 337K$).

The reported protofibril stability data are based on statistics collected approximately 50-150 independent simulations per chain number. Statistics on the chain conformation were gathered every 50τ (10,000 steps). Structural stability for each time point was quantified by two different variants of the χ parameter:

$$\chi = \frac{1}{M} \sum_{\alpha=1}^{N_c} \sum_{\beta > \alpha}^{N_c} \sum_i^N \sum_j^N h(\varepsilon - |r_{\alpha,i;\beta,j} - r_{\alpha,i;\beta,j}^0|) \quad (4.1)$$

The generic χ parameter evaluates the sum over bead i on chain α and bead j on chain β and α and β range over the N_c chains making up the exterior and neighboring chains on each end, h is the Heaviside step function, ε is the tolerance set to 0.5 distance units ($\sim 1.9\text{\AA}$), $r_{\alpha,i;\beta,j}$ is the distance between bead i on chain α and bead j on chain β , and $r_{\alpha,i;\beta,j}^0$ is the pair distance in the initial structure, and M is a normalizing constant counting the total number of pairs.

The two variants of the χ parameter involve different ranges of the restricted sum over chains α and β , and beads i and j in Equation 4.1. χ_f measures β -strand order on an individual protofilament and alignment of the protofilament with the fibril axis, by evaluating i and j over the range from [17-21; 31-35], and over 4 monomer chains on each end (2 independent contributions from each end involving a total of 8 chains). P_f measures the nativeness of an individual protofilament, by evaluating i and j over the

range from [17-35], including both β -strand regions as well as the turn connecting these regions. Since P_f is isolated to a single protofilament, each protofibril end has two values of P_f that are binned independently (4 independent contributions involving 8 chains total).

Free energy profiles

Based on the ensemble composed of the final structures of each of the 50 independent trajectories for each sequence and for each oligomer size, n , we can calculate equilibrium populations of structurally stable and unstable protofibrils based on population differences measured by either P_f or χ_f . For chain lengths and mutant combinations for which the population of either stable or unstable protofibrils is very small, we run an additional 100 trajectories for a total of 150 trajectories in order to reduce the error of our population estimates. The fraction of trajectories corresponding to $P_f > 0.7$ or $\chi_f > 0.7$ measures a population, C_n , of n -ordered monomers in a protofibril with intact end monomers and a well-defined fibril axis. This population is in equilibrium with the remaining fraction of trajectories corresponding to a protofibril with loss of structural order corresponding to $P_f < 0.7$ or $\chi_f < 0.7$, and thus measures the population C_{n-1} . We have chosen P_f and χ_f dividing surfaces of 0.7 based on the best single value of the parameters that divides the high and low chain number populations. We confirmed the choice of value by visual examination of structures with a range of P_f and χ_f values and found the values to accurately divide ordered and disordered structures.

Based on thermodynamic arguments advanced by Ferrone[123] for nucleation-polymerization reactions relevant for aggregation kinetics, at equilibrium we can estimate the change in free energy, ΔG , per unit monomer as

$$\frac{d\Delta G}{dn} = -kT \ln\left(\frac{[C_{n-1}]}{[C_n]}\right) \quad (4.2)$$

where n is half the number of monomers, and kT , is the Boltzmann constant multiplied by the temperature. Integration over all oligomer sizes allows us to generate a free energy curve based on C_n and C_{n-1} populations measured in our model for the different sequences.

4.3 Results

We investigate the structural stability of fibril seed models for the Arctic, Dutch, and Flemish mutants ranging from 2 to 10 dimer cross-sections (i.e. 4 to 20 monomer chains) under the C_{2z} symmetry form of the agitated fibril morphology. Each of these protofibril sizes are simulated using Langevin dynamics at a constant temperature of $T^*=0.45$ ($T \approx 337K$), and we monitor the amount of fibril order as a function of time. As a measure of fibril order, we define two different structural similarity parameters (see Methods). The first order parameter, P_f , measures the structural similarity of the ends of the protofilament subunits with respect to perfect fibril order. The second order parameter, χ_f , measures β -strand order over the ends of the whole protofibril, and thus is sensitive to disorder at both the level of the protofilament and the quaternary structure of the protofibril.

Figure 4.2 show the histograms of populations of P_f order for the final structures for 4, 8 and 20 peptide assemblies of the three FAD mutants with respect to the WT A β sequence. We see that fibril structural similarity at the level of a protofilament increases in order of FAD mutants: Flemish < Arctic < WT < Dutch at any size protofilament,

regardless of the number of peptides. Since a higher concentration of peptides in solution should drive the equilibrium toward larger assemblies, the larger sized protofilaments would be more likely to be found in solutions at higher concentrations. The greater disorder by this metric for the Flemish and Arctic mutant is clearly a consequence of the glycine mutation that permits greater flexibility of the backbone dihedral angles in the N-terminal β -sheet region. The population distributions are largely no different for 8 chains vs. 20 chains for the Arctic sequence, while there is a systematic gain of some structure for the Flemish mutant as concentration increases, although both are less fibril-like than WT. Figure 4.3 shows that while the Arctic mutant exhibits disorder in the turn regions, it still retains its β -strand pairings, unlike the Flemish mutant that loses the attachment of the edge monomer to the protofibril. By contrast, the Dutch sequence shows structural enhancements over WT by a primary sequence mutation that eliminates charge repulsion between peptides on the same protofilament, so that its populations are more ordered than WT at any chain assembly size. The enhancement of protofilament order for Dutch exaggerates the twist down the protofilament axis with respect to WT, as shown in Figure 4.4.

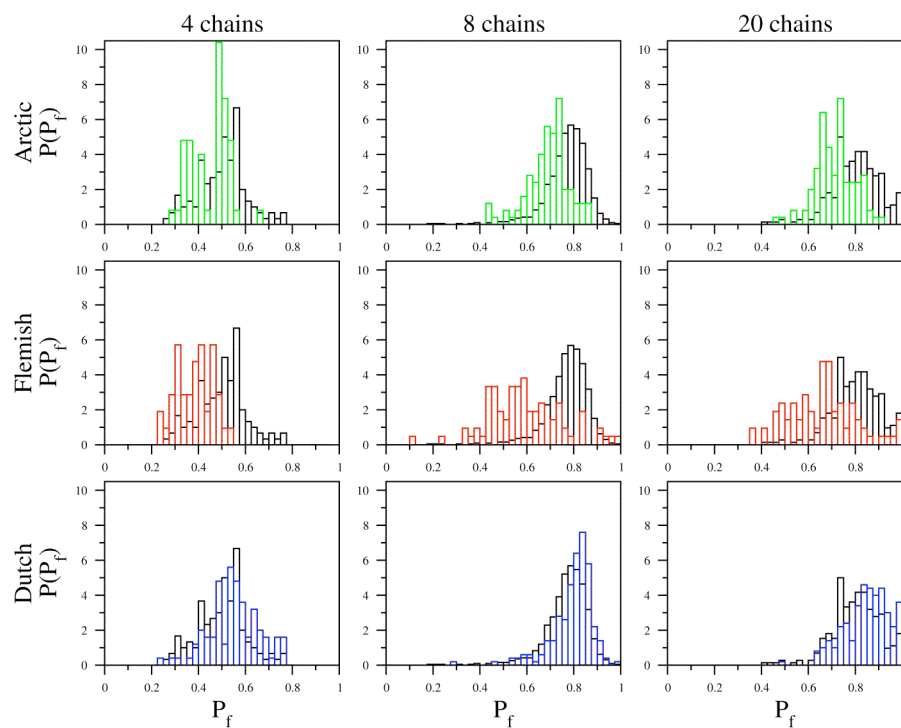


Figure 4.2: Population histograms with respect to protofilament order (P_f) for 4, 8, and 20 chains. The histograms emphasize that protofilament order increases for the FAD mutants as: Flemish (red) < Arctic (green) < WT (black) < Dutch (blue), at any oligomer size.

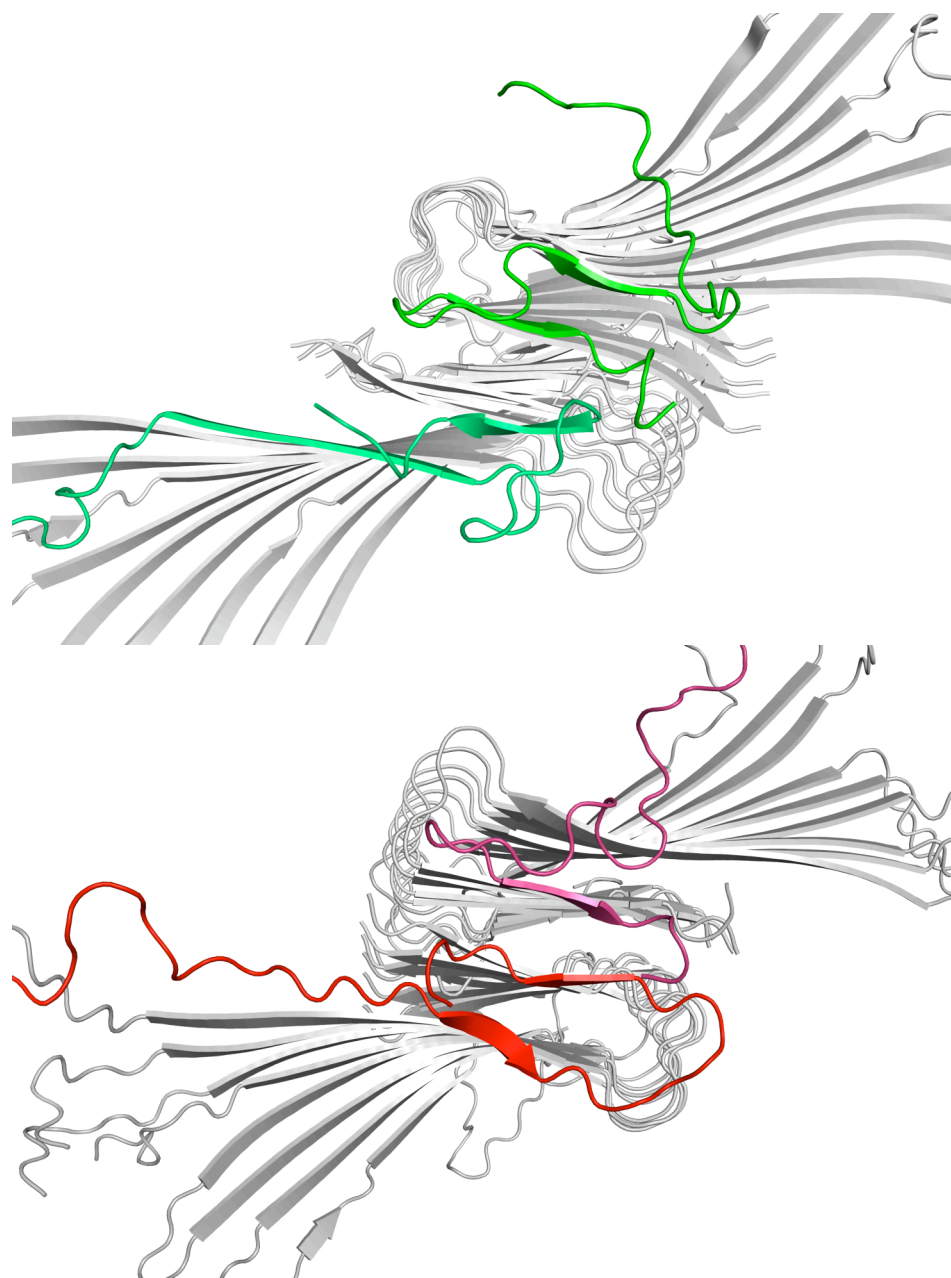


Figure 4.3: Representative protofibril structure of the Arctic and Flemish mutants. While both FAD mutants show disorder in the turn region, the Arctic mutant (green) retains much better β -strand structure over the whole cross-section at the end of the 5000τ trajectories, while the Flemish mutant (red) has almost lost a monomer after the same amount of time.

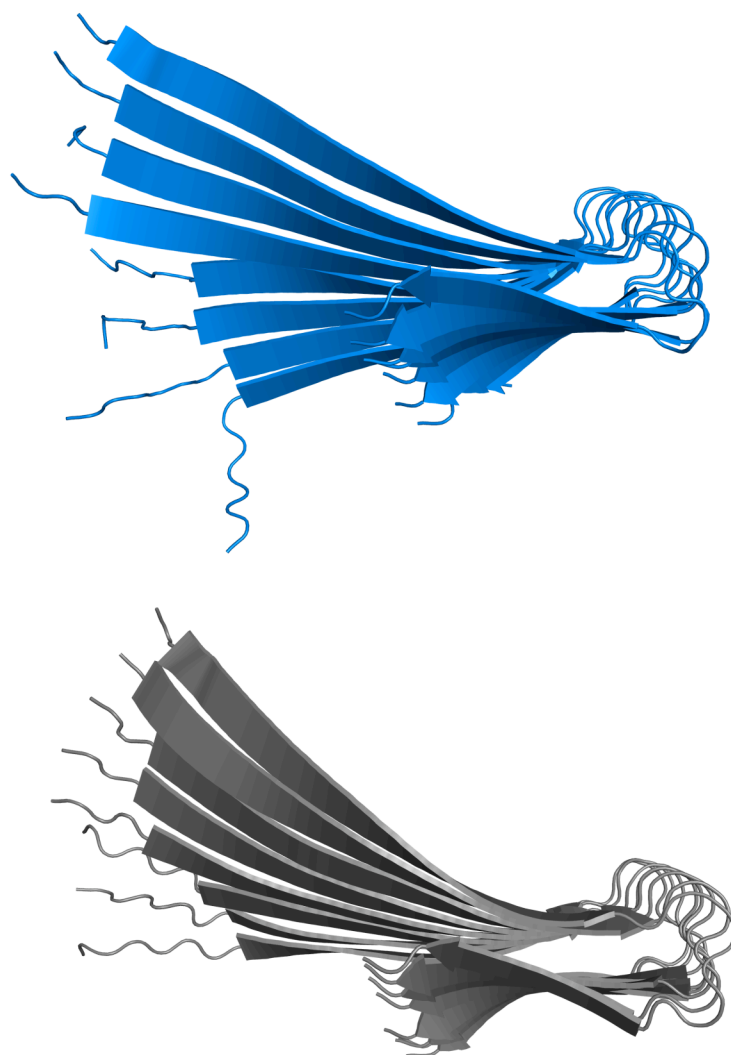


Figure 4.4: Representative protofilament structure of the Dutch mutant compared to the WT sequence. Already after only initial equilibration from the model build, the Dutch mutant (blue) shows a greater twist of the intermolecular β -sheet down the protofilaments with respect to WT (black).

The protofilament order trends for the mutant and WT sequence do not predict the trends in our χ_f metric that measures retention of order across the protofibril ends. In Figure 4.5 we show the histograms of populations of χ_f for 4, 8 and 20 peptide assemblies of the three FAD mutants with respect to the WT A β sequence. The 4 chain assemblies are equivalent among the sequences: no protofibrils are present at such low

concentrations. However fibril structural similarity at the level of a protofibril is different among the sequences at 8 chains to yield a different order for FAD mutants: Flemish < Dutch < WT < Arctic. Note that in Figure 4.5 top right, the χ_f at 20 chains for Flemish mutant at 20 chains has a far lower population at χ_f greater than 0.7, never adopting the level of protofibril order that is reached by the Arctic, Dutch and WT mutants at 8 chains.

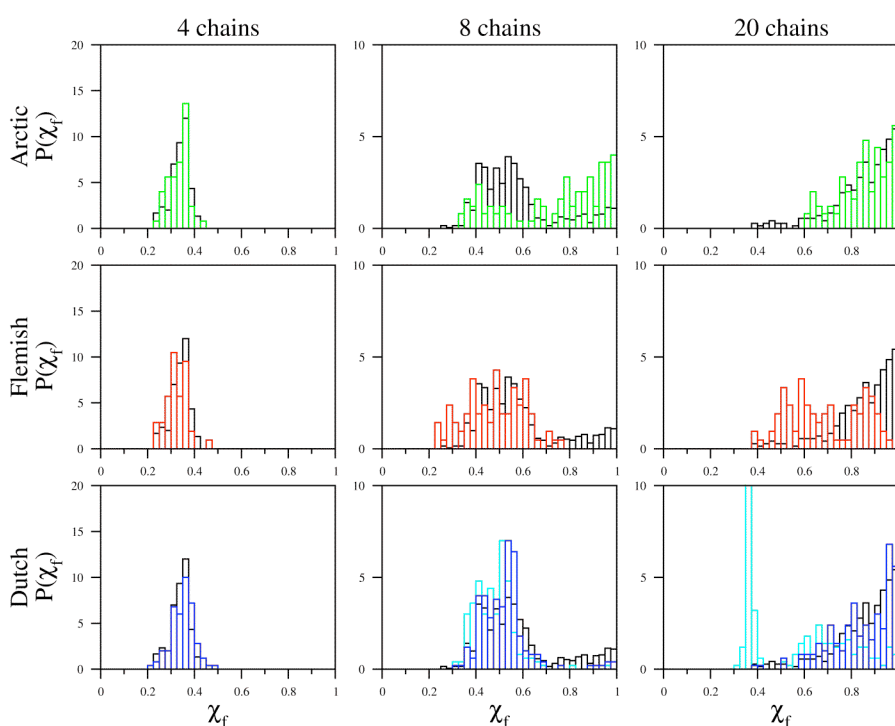


Figure 4.5: Population histograms with respect to protofibril order (χ_f) for 4, 8, and 20 chains. While no protofibrils are present for the 4 chain assemblies for any sequences, the level of protofibril structure is different among the sequences at 8 chains to yield the following trend for FAD mutants: Flemish (red) < Dutch (aqua) < WT (black) < Arctic (green). Even with the addition of the Dutch polymorph as a reference state (blue), there is slightly more disorder for the Dutch with respect to WT.

The position of the glycine mutant results in qualitatively different behavior in the structural integrity of the protofibril. The glycine mutation at position 21 is far enough into the N-terminal β -strand to diminish fibril integrity across the whole end cross-

section of β -strands. By contrast, the glycine mutation at position 22 pushes the disorder nearer to the turn region, thereby retaining β -strand order over the whole cross-section. In both cases, new but non-specific stabilizing interactions between the turn region and the β -strands prevent the protofilaments from rotating with respect to each other so that both retain a well-defined protofibril axis (Figure 4.3).

While the protofilament assemblies are better formed for the Dutch mutant, the agitated fibril morphology is not a viable reference state for ordered protofibril structure (Figure 4.5, aqua). In fact a new polymorph (comprising 50% of the population of the 16, 18, and 20 chain protofibril, and about 40% for 14 chain, 35% for 12 chain, and 20% for 10 chain) is seen in which the protofilaments show a shift in register of β -strand alignment at the interface (Figure 4.6). Even when this new polymorph serves as an additional reference state for fibril order (Figure 4.5, blue), there is still some disorder for the Dutch mutant when it is compared to WT at the same number of chains, as seen by the shallower negative slope for Dutch, which is due to more rotational freedom of one protofilament with respect to another.

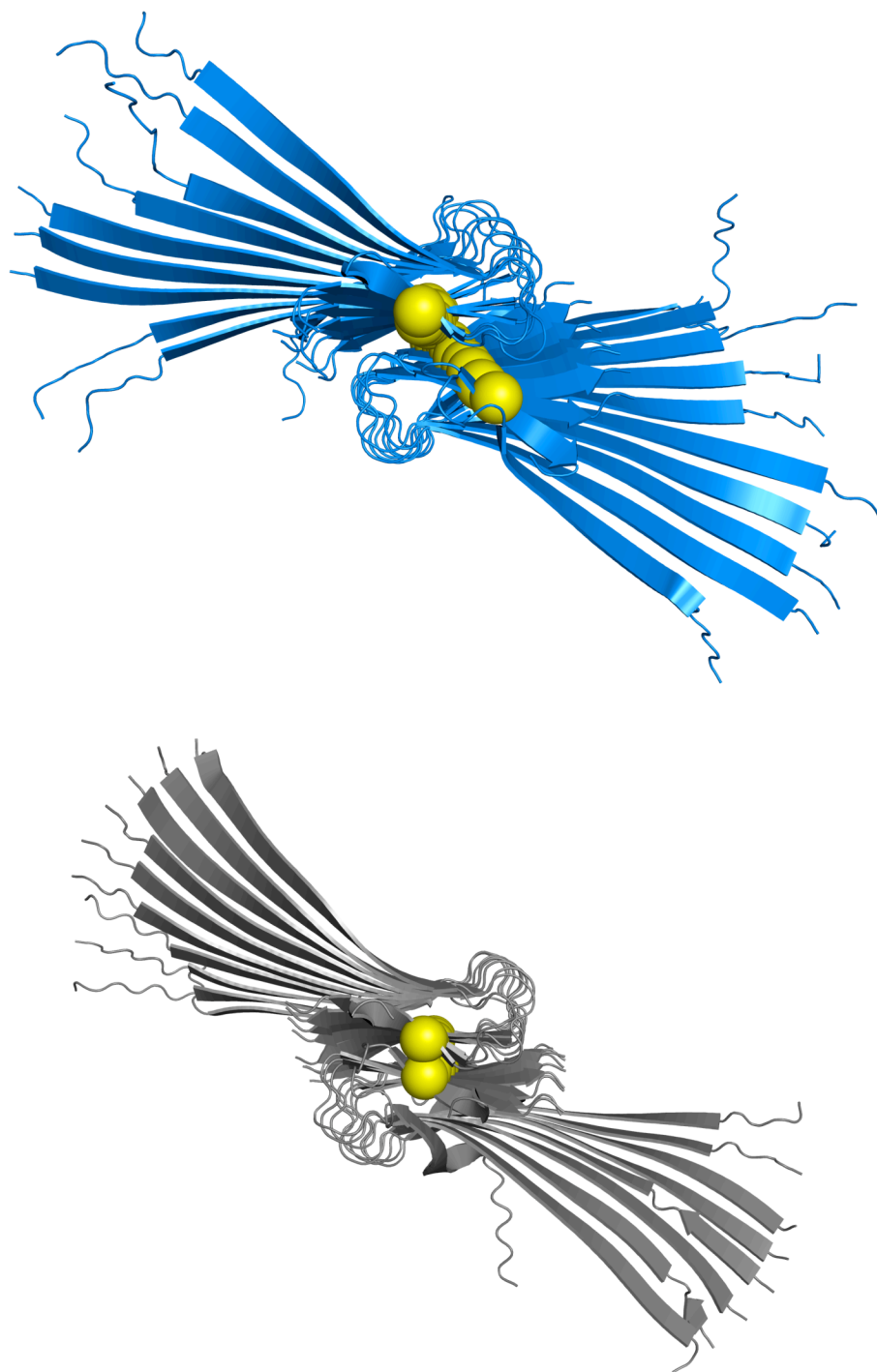


Figure 4.6: Representative protofibril structure of the Dutch mutant with respect to WT. A comparison of the Dutch polymorph (blue) with respect to the agitated fibril morphology [17, 31-33] favored by the WT sequence (black) at the end of the 5000 τ trajectories. The yellow spheres represent amino acid 33 on each monomer chain, which

shows how it is displaced due to a register shift of the C-terminal β -strands at the interface under the Dutch polymorph.

4.4 Discussion

The kinetics of pre-fibrillar Arctic and WT A β peptides have been quantified by chromatographic methods that measure rates of appearance and disappearance of monomer and/or A β oligomer assemblies based on their mass[16], with no information as to their structural characteristics. A more structurally informative kinetic assay is based on Congo Red or Thioflavin T dye-binding fluorescence[137] which measures the disappearance of monomer into growing fibril assemblies that have cross β -strand order, whose long-time saturation indicates the formation of mature fibrils. However even this kinetic measurement is not particularly sensitive to the structural details of the oligomeric assemblies that are accumulating in the measured kinetic profiles.

Our examination has revealed that substantial differences in structural ensembles exist between the four different A β sequences based on P_f and χ_f . Both the P_f and χ_f metrics are consistent in the formation of good cross β -strand order, so that kinetic assays based on Congo Red or Thioflavin T dye-binding fluorescence are equally relevant to both of these reaction coordinates. The only difference between P_f and χ_f is that the latter assumes a higher level of structural organization so that pairs of protofilaments are well-defined with respect to the fibril axis to adopt the agitated fibril morphology. How do these structural ensembles for the different mutants connect to the observed differences in their kinetic rates of fibrillization? We make this connection under the assumption of a dynamic equilibrium between monomer and protofibril states,

with the equilibrium constant allowing us to define a free energy profile as a function of protofibril assembly size (see Methods).

Figure 4.7a plots the free energies as a function of oligomer size for the WT and familial mutant sequences based on protofibril order, with $\chi_f > 0.7$. We find that the size of the critical nucleus shifts dramatically to a smaller number of monomer cross-sections for the Arctic mutant corresponding to 6-8 chains, and exhibits a greater drive to form protofibrils with respect to WT given the smaller free energy barrier. For the Flemish mutant we find that the critical nucleus is shifted to beyond 20 chains analyzed in this study, thereby always favoring the monomer. The free energy profile for the Dutch mutant using the agitated fibril morphology relevant for WT is a poor measure of order into higher order protofilament assemblies. Even when we add the additional polymorph as a reference structure, we find that the Dutch mutant has the same critical nucleus size and slightly larger barrier to protofibril order with respect to WT. Given that the fibrillization kinetics are faster for Dutch relative to WT, this result suggests that the Dutch mutant does not favor the higher order assemblies of protofilament-protofilament organization that arise from variations of the agitated protofibril morphology.

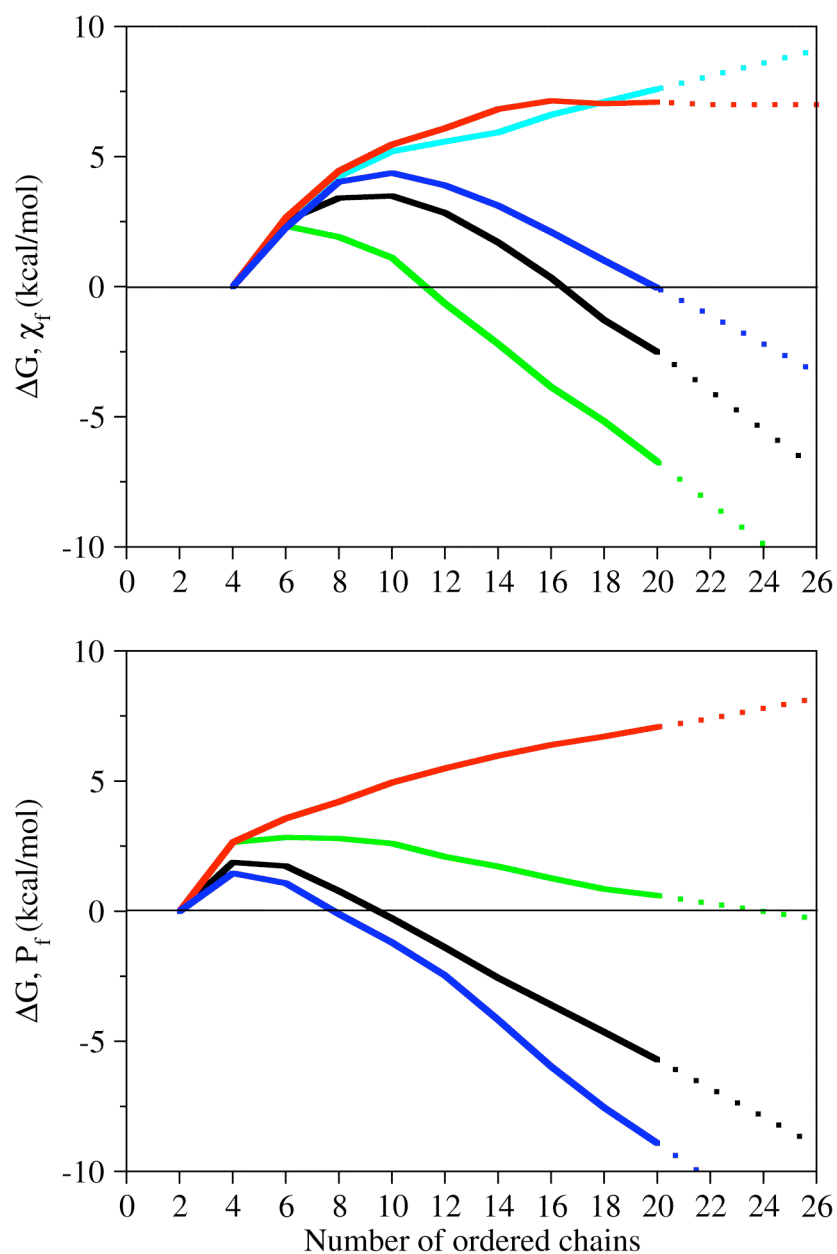


Figure 4.7: (a) Free energy profile for free monomer and protofibril (χ_f) equilibrium for the WT and FAD mutants. The free energy shows a maximum corresponding to the critical nucleus size of 6 chains for Arctic (green), 10 chains for WT (black), and no preferred order for either Dutch (aqua) or Flemish (red) mutants. The constant, negative slope beyond the critical nucleus is indicative of reaching a stable fibril regime in which the equilibrium shifts decidedly away from the monomer form. When the alternative polymorph for the Dutch mutant (Figure 4.6) is added as a reference for fibril order (blue), there is now a critical nucleus of 10 chains for the E22Q mutant but with a larger

free energy barrier and shallow slope indicating a slower approach to protofibril order for the Dutch mutant. (b) Free energy profile for free monomer and protofilament (P_f) equilibrium for the WT and FAD mutants. The free energy shows a maximum corresponding to the critical nucleus size of 6 chains for Dutch (blue), 6-8 chains for WT (black), and no preferred order for either Arctic (green) or Flemish (red) mutants. The constant, negative slope beyond the critical nucleus is indicative of reaching a stable regime in which the equilibrium favors the protofilament form, which is more strongly evident for Dutch over WT.

As a measure of lower order assemblies, Figure 4.7b exhibits the free energies as a function of oligomer size for the WT and familial mutant sequences based on protofilament order, with $P_f > 0.7$. We see that the Dutch mutant shows the smallest critical nucleus and free energy barrier relative to all other sequences. The Dutch mutant preference for lower order assemblies involving only protofilaments which are not subject to the free energy barrier for ordering and aligning a two-filament cross section fibril [136] may explain its significantly enhanced fibrillization kinetics using dye-binding assays of cross β -sheet structure. Alternatively, a higher order assembly of a substantially different polymorph other than the agitated fibril morphology may be relevant for its fibrillization mechanism. By this measure the Arctic mutant shows a flat free energy curve indicating that its structural stability arises primarily from protofilament-protofilament alignments to define a fibril axis, and that the agitated fibril assembly is a good model for this mutation. Again the Flemish mutant is disordered and never exhibits a stable protofilament regime. This result is similar to that found by all atom simulations of a Flemish mutation dimmer where the A21G mutation destabilized the dimeric assembly[129] and here we demonstrate this behavior continues to ordered oligomers larger than dimers and connect this local disorder to protofibrillar instability. Given the nature of the A21G mutation that so strongly favors the monomer over ordered

cross β -sheet structure, we believe no specific alternative ordered protofibril reference state exists for the Flemish mutant.

4.5 Conclusions

We have used a coarse-grained protein model to measure structural stability trends of A β protofibril assemblies for WT as well as for Arctic (E22G), Dutch (E22Q), and Flemish (A21G) mutant sequences. We find that although the single point mutations are localized in the same region of the A β peptide, their structural ensembles are quite distinct, and the mutations can disrupt organization at the level of protofilament up through protofibril order. By measuring the equilibrium populations of monomer \Leftrightarrow protofilament or monomer \Leftrightarrow protofibril as a function of protofibril size, we determine free energy profiles that are consistent with the attainment of cross β -sheet structure measured by dye-binding assays, while providing better structural information upon which to develop new hypotheses for experimental investigation.

We find that while both the Arctic and Flemish sequences promote greater disorder of the β -turn region of the A β peptide, the difference in sequence position of the glycine mutation radically alters fibril order stability. The glycine mutation at position 21 in the Flemish mutant disrupts the exterior N-terminal strand regions, thereby degrading order throughout each protofilament and at the interface between protofilaments. Regardless of the detection method (P_f vs. χ_f) for cross β -sheet structure, the dynamic equilibrium strongly favors the monomeric peptide for the Flemish mutant. The greater resistance of the Flemish mutant to order into fibril assemblies of any size suggests that it is capable of both greater fragmentation into smaller oligomers that can readily diffuse,

while at the same time possibly promoting amorphous aggregation to yield large plaques by recruiting other proteins and extracellular constituents into its more unstructured A β aggregates. Our results suggest it would have no definitive single fibril morphology reference state.

By contrast, the glycine mutation at position 22 is enough removed from the N-terminal strands that they retain their protofilament order, although it does increase the flexibility in the turn region of the A β monomer. The more flexible loop can form new contacts that stabilize the fibril at the interface so that little rotation between the protofilaments is exhibited beyond 6 chains. It seems likely however that while the critical barrier is rapidly reached at lower concentrations than WT, the new stabilizing contacts could slow the addition of monomer beyond that point, i.e. that there is a separation between rapid attainment of small oligomers that do not translate into more rapid rates of fibrillization into large assemblies. This would be consistent with chromatography methods that measure more rapid disappearance of monomer into oligomer formations for Arctic relative to WT, but find little difference in rates of forming fibrils from these oligomer states[16].

The Dutch mutant shows an increase in protofilament order, i.e. better alignment of β -strands on the N-terminal (amino acids 17-21) and C-terminal β -strand regions (amino acids 31-35) and little disorder in the turn region. However structural rearrangements in the monomer creates a new twist angle in the protofilament that does not allow the protofilaments to align along the fibril axis consistent with the agitated fibril structure found for the WT sequence. In fact, rearrangement between the

protofilaments results in a new polymorph of the Dutch protofibril that is substantially populated. The enhanced fibrillization kinetics measured by dye-binding assays of cross β -sheet structure for the Dutch mutant may stem in part from its possible preference for lower order assemblies involving only protofilaments. Alternatively, a higher order assembly of protofilaments into a different polymorph other than the agitated fibril morphology may be relevant for its fibrillization mechanism.

4.6 Acknowledgements

We thank Robert Tycko for the atomic coordinates for his models of $A\beta_{1-40}$. K. L. K. thanks the DOE Computational Science Graduate Fellowship for support and for his internship at Berkeley Lab. N.L.F. thanks the Whitaker Foundation for a graduate research fellowship. Y.O. thanks the Guidant Foundation for a summer research fellowship. Molecular graphics for this paper created in PyMOL (DeLano Scientific, San Carlos, CA). We thank Dr. Jonathan Kohn for careful reading of the manuscript. This work was supported in part by a grant from the NIH.

Chapter 5: Structure and dynamics of the A β ₂₁₋₃₀ peptide from the interplay of NMR experiments and molecular simulations

Reproduced with permission from Journal of the American Chemical Society, submitted for publication. Unpublished work copyright 2007 American Chemical Society.

Nicolas Lux Fawzi¹, Aaron Phillips², Jory Z. Ruscio³, Michaeleen Doucleff², David E Wemmer^{2,4}, Teresa Head-Gordon^{1,3,4}

¹UCSF/UCB Joint Graduate Group in Bioengineering, Berkeley, CA 94720

²Department of Chemistry, University of California, Berkeley, CA 94720

³Department of Bioengineering, University of California, Berkeley, CA 94720

⁴Physical Biosciences Division, Lawrence Berkeley National Laboratory
Berkeley CA 94720

Abstract

In this study, we use molecular dynamics to calculate a structural ensemble that describes the solution structure of the A β ₂₁₋₃₀ peptide. From it we directly predict NMR observables such as three bond scalar coupling constants (³J_{H_{NH} α}), chemical shift values, ¹³C relaxation parameters, and ROESY cross peaks. New generation TIP4P-Ew and Amber ff99SB force fields and high quality sampling of the equilibrium ensembles provide excellent predictions of the NMR experimental observables, allowing us to describe the structural diversity of the A β ₂₁₋₃₀ conformations more completely than previous studies on this peptide fragment. We find that the structural ensemble of the A β ₂₁₋₃₀ peptide involves a majority population (~60%) of unstructured conformers, lacking any secondary structure or persistent hydrogen-bonding networks. However the remaining minority population contains about ~14% of conformers with a β -turn centered at Val24

and G25, and a 7% population with evidence of the Asp23 to Lys28 salt bridge, important to the fibril structure. The accurate prediction of experimentally measured parameters is a strong validation of the molecular simulation, requiring both structurally and dynamically correct ensembles over the entire peptide. This serves as a validation of the approach for characterization of disordered peptides and proteins in general, and specifically for A β ₁₋₄₀ and A β ₁₋₄₂, for which collection of detailed NMR data will be more challenging due to aggregation and fibril formation on experimental timescales at physiological conditions.

5.1 Introduction

The amyloid β (A β) peptide, comprised of a family of 39 to 42 residue long fragments cleaved by proteolysis of the APP protein, is the major species in amyloid fibril plaque found in the brains of patients with Alzheimer's disease (AD)[2, 69]. It is not yet clear whether the actual fibril plaque may be disease causing[81], or whether it is the pre-fibrillar monomer and oligomeric states that have received attention recently, since they also appear to be cytotoxic and related to the Alzheimer's disease state[116-120, 139]. A knowledge of the structures of the A β peptide on the aggregation pathways from monomer to fibril is therefore critical to a molecular approach for the prevention or control of aggregation outcomes *in vitro*, with the hope of potentially changing the course of the disease *in vivo*.

While significant progress has been made in the characterization of highly ordered amyloid fibril structures of A β peptides under physiological conditions[6, 13, 81-

83, 140], the aqueous monomeric form is difficult to study by standard structural biology techniques due to its propensity to sequester into these ordered fibril assemblies. Recent studies by NMR spectroscopy have found some regular structure in the $A\beta_{1-40}$ and $A\beta_{1-42}$ peptides in non-aqueous solvents[75, 141-143], but the high concentration of apolar solvent (hexafluoroisopropanol, trifluoroethanol, SDS micelles) are far from the relevant physiological state. Other studies have calculated averaged quantities such as scalar coupling constants[144] and spin relaxation constants[145, 146] for aqueous solution ensembles of the peptide, but these data measure only local rather than tertiary structure and hence provide an incomplete description of the structural ensemble.

In an attempt to probe for any significant structure in the monomeric state by an alternative technique, Lazo *et al.* subjected the $A\beta_{1-40}$ and $A\beta_{1-42}$ peptides to digestion by multiple proteases[24]. Despite containing a large number of potential proteolytic sites throughout the sequence, it was determined that the fragment spanning residues 21 to 30 of the peptide was a significant product for each enzyme, indicating relative resistance to cleavage. Little digestion was seen when synthetic $A\beta_{21-30}$ was subjected to the same protease conditions. These results may imply that some structure in the 21 to 30 region protects $A\beta_{1-40}$ and $A\beta_{1-42}$ from protease degradation, and that this structure is retained in the $A\beta_{21-30}$ fragment.

The $A\beta_{21-30}$ peptide consists mainly of a hydrophilic region that in the fibril state of full-length $A\beta_{1-40}$ has been shown to form a turn connecting the two flanking hydrophobic β -sheet regions[82] that are characteristic of amyloid states of peptides. Structure and chemical detail in this peptide may play a significant role in the physico-

chemical properties of the structures that lead to AD. It is also noteworthy that many of the familial associated disease (FAD) mutants of the APP protein are located between residues 21 to 23, each of which leads to dramatically different *in vitro* fibril formation properties and *in vivo* clinical outcomes[15-18, 110-113, 147, 148] Additionally, the buried salt bridge between Asp23 and Lys28 in the fibrillar form determined by solid state NMR[13, 80, 82] is also encompassed by this peptide; it has been shown that mutating either residue dramatically affects fibril formation[80, 112]. Finally this small zwitterionic peptide fragment has a net charge of -1 at pH 7 due to the presence of two acidic and one basic residue in addition to the N- and C-termini, and contains no aromatic and only a single aliphatic residue (Val24), making the peptide extremely soluble in water.

A number of studies have experimentally and computationally studied the A β_{21-30} peptide to determine what stable structure in the A β_{21-30} monomer accounts for the protease resistance. Teplow and coworkers studied the wild-type A β_{21-30} peptide[24] as well as five FAD mutants[23] using rotating frame Overhauser effect spectroscopy (ROESY) NMR experiments. Based on a single cross peak interpreted as a long-range ROE, Teplow and coworkers have proposed that A β_{21-30} folds into a single population corresponding to a unique bend structure, identified through long range (i,i+8) Glu22 H $_{\alpha}$ to Ala30 HN and (i,i+6) Glu22 sidechain to Lys28 sidechain ROESY crosspeaks. Replica exchange simulations using the OPLS all-atom model for the peptide and explicit TIP3P water by Shea and coworkers show 40% of the peptide ensemble is folded into two distinct bend structures stabilized primarily by the Asp23 sidechain interactions with

Ser26 sidechain and backbone, although they found no long-range interactions that were observed in the NMR experiment[149]. Stanley and coworkers studied the peptide by a coarse-grained model in which they find folded structure, stabilized by hydrophobic interactions between Val24 and Lys28, as well as Asp23 to Lys28 electrostatic interactions[150]. They also examined the peptide in five ~ 100 ns CHARMM-27/TIP3P explicit water molecular dynamics trajectories (without using any accelerated sampling technique), each with a different combination of density, starting structure and salt concentration[151]. Though the authors acknowledge that their simulations are far too short to sample the equilibrium ensemble, they report contacts between hydrophobic regions of Val24 and Lys28 that are more stable on the nanosecond timescale than charged interactions between Lys28 and Glu22 or Asp23. Finally Mousseau, Derreumaux and coworkers used an activation-relaxation sampling technique combined with the OPEP coarse-grained model of A β_{21-30} and found several clusters of structures, all sharing a turn formed between Val24 and Lys28 but stabilized by different sidechain contacts[152].

Although significant progress has been made recently in the interpretation of NMR observables for disordered peptide and protein systems[153-156] combining multiple and independent structural constraints for a system with significant disorder often leads to an inadequate description of the ensemble diversity[154]. By contrast molecular dynamics simulations of disordered systems has the opposite challenge where the simulated ensemble is directly observable with good statistical confidence but the accuracy is difficult to assess due to uncertainties of the underlying empirical force

fields[157, 158]. While simulations of folded proteins in their native state have been shown to quantitatively reproduce NMR observables (model-free order parameters, relaxation times T_1 and T_2) [25, 159, 160], partially structured peptides and natively unfolded proteins present a new challenge for simulations in which small energetic biases or inaccuracies can dramatically affect the populations of structures in the equilibrium ensemble. Stock, Schwalbe and coworkers have recently shown that many earlier generation peptide-water empirical force field combinations simply do not reproduce the average structural ensemble for a simple disordered system such as polyalanine[161].

In this study, we calculate a structural ensemble for the $A\beta_{21-30}$ peptide using molecular dynamics simulations with more recently parameterized empirical force fields for protein and water that quantitatively agrees well with multiple NMR determined parameters. We find that the quality of the peptide force field and the dynamical properties of the water model dramatically affect the results. A combination of Amber ff99SB[25] and TIP4P-Ew[26] demonstrate the ability of the simulation to directly predict the NMR observables such as three bond scalar coupling constants ($^3J_{\text{H}^{\text{NH}\alpha}}$), chemical shift values, relaxation parameters, and ROESY cross peaks. An accurate prediction of all of these parameters is a strong validation of the molecular simulation, requiring both structurally and dynamically correct ensembles over the entire peptide. We use the simulated ensemble as a tool to interpret the ensemble averaged NMR information for the $A\beta_{21-30}$ peptide.

From this interplay of experiment and simulation, we determine that the observed medium range ROEs from our experiment on $A\beta_{21-30}$ arise from independent populations

of local turn structure in regions 23 to 27 and 27 to 30, different from what was reported by Teplow and coworkers[23, 24]. Though it is possible to construct an artificial ensemble of structures that simultaneously satisfy all ROE interactions interpreted as distance restraints, this minimized NMR ensemble is not stabilized by any significant electrostatic or hydrophobic interactions. The validated simulated ensembles make clear that the ROE interactions are not satisfied simultaneously and in fact arise from separate structural populations that together comprise only ~40% of the total equilibrium ensemble. The very good quality of results of this validation study on A β ₂₁₋₃₀ paves the way for simulating the structural ensemble of the A β ₁₋₄₀ and A β ₁₋₄₂ systems with high quality of prediction. This is a particularly important consideration since solution NMR experiments on these peptides are inherently more difficult due to peptide aggregation.

5.2 Methods

NMR Experiments

The A β ₂₁₋₃₀ peptide (AEDVGSNKGA) was synthesized (Anaspec, San Jose CA) and purified to 98% purity by reverse-phase HPLC. NMR samples contained 10 mM A β ₂₁₋₃₀ and 25 mM ammonium d₄-acetate in 90% H₂O, 10% ²H₂O or 100% ²H₂O. The solution pH was adjusted to 6.0 with 20 μ l 1M NaOH. NMR data were collected at 283 K on a Bruker Avance 500 MHz and Bruker Avance II 800 or 900 MHz spectrometers. All data were processed with NMRPipe[162] and analyzed with NMRView[163] and CARRA[164]. All spectra were recorded at 10°C to facilitate comparison with previous NMR studies on this peptide.

Chemical shift assignments were obtained with a 2D ^1H - ^1H TOCSY. Distance restraints were obtained from a 2D ^1H - ^1H ROESY experiment with a spinlock using $180^\circ(x)$ - $180^\circ(-x)$ pulses[165] and 300 ms mixing time. A total of 4096 and 1400 points (States-TPPI) were collected in t_2 and t_1 , respectively. Sweepwidths in both dimensions were 7184 Hz and 8503 Hz on the 800 MHz and 900 MHz spectrometers, respectively.

To build a single structural model fitting all of the data simultaneously using a standard NMR structure determination approach, ROESY crosspeaks were classified as strong, medium, weak, and very weak based on peak intensity and converted to 2.9, 3.3, 5.0, and 6.0 Å distant restraints, respectively. The set of 155 manually assigned distant restraints were used to calculate 1000 structures with the program CYANA[166]. The 50 lowest energy structures of the minimized ensemble was analyzed with the program Pymol[167]. Structural statistics and hydrogen bonds present in the structure ensemble, detected with the Amber suite [168]program *ptraj* are presented in the Results.

Spin-lattice (T_1) and spin-spin (T_2) relaxation times for natural abundance ^{13}C at the C_α position were measured for all non-glycine amino acids using the same 100% $^2\text{H}_2\text{O}$ sample described above. T_1 was measured at 500 MHz by inverse-detected inversion recovery with delay times 5, 25, 50, 150, 400, 600, 1000, 2000, and 2500 ms[169]. T_2 was measured at 600MHz by an inverse detected CPMG experiment with delay times 0, 20.48, 40.96, 61.44, 82.92, 122.88, 143.36, 163.84, 204.8 ms[169]. Relaxation parameters were fit from the data as described previously[169].

Three bond scalar coupling constants $^3J_{\text{HNH}\alpha}$ were measured by lineshape fitting from a 2D double-quantum filtered COSY. To ensure that the experimental lineshape

was not adversely affected by limited digital resolution 4096 points were collected in both t_1 and t_2 . Quadrature detection in t_1 was obtained according to the States-TPPI method and the digital resolution was matched to t_2 with linear prediction.

Simulation Protocol

We simulate the zwitterionic A β_{21-30} peptide represented by the Amber ff99SB fixed charge empirical force field[25]. Amber ff99SB is a recent reparameterization of the backbone dihedral angles by Simmerling and coworkers to correct previous problems with secondary structure propensities of the original ff99 parameters. Amber ff99SB quantitatively captures the distribution of backbone ϕ/ψ angles compared with quantum mechanical calculations and validation on model peptide and protein systems. We have chosen to run two separate sets of simulations where we solvate the peptide with TIP3P[170] and TIP4P-Ew water models, respectively. TIP4P-Ew is a re-parameterized version of the standard TIP4P water model for use with Ewald summation techniques, and reproduces many salient thermodynamic and dynamic features of bulk water properties when compared with experiment[26]. Its excellent performance for temperature trends of these properties is especially relevant for this experimental study, which is conducted at 10°C.

The long-range electrostatic interactions are calculated using Particle Mesh Ewald method (PME)[171] and a cutoff of 9.0 Å is used for real space electrostatics and LJ interactions. The equations of motion are integrated with 1fs timesteps in all simulations. Replica exchange simulations[27, 172] are prepared by building the A β_{21-30} structure in *tleap* and solvated it with 1578 TIP3P or 1579 TIP4P-Ew water molecules, respectively.

A single sodium (Na^+) ion is included in the system to balance the peptide net charge. Each system is then briefly equilibrated using Andersen thermostats[173] to bring the system up to 300K temperature, then equilibrated for 125ps at constant pressure with Berendsen (weak) coupling at 1 bar (default parameters) and 300K temperature. The average density of the last 100ps of the constant pressure simulation is then calculated and a snapshot containing position and velocity information with that density (within 0.001g/cm^3) is selected as the starting structure for each temperature replicas. The *sander* module of Amber9 simulation package is used to propagate the 64 temperature replicas exponentially spaced between 270 K and 507 K with exchange attempts every 1 ps.

Each separate replica exchange simulation was run for a time between 45ns and 50ns per replica. In order to check convergence of the structural ensembles, a second replica exchange simulation is started with a coordinate file from the first replica exchange simulation after 20ns equilibration for all replicas, and convergence of equilibrium structural populations is compared to the first production run. We find that the equilibrium populations differ by no more than 7% in linear average of the pair distances, although we comment later on the challenge of converging these averages and their effect on the predictions in the Results.

We use the ptraj module of AMBER to analyze the DSSP defined secondary structure[174], hydrogen-bonds and electrostatic/saltbridge interactions. We specify all possible donors and acceptors for the hydrogen bond analysis.

Simulation of NMR Experimental Observables

We predict the chemical shifts for all protons and C_α and C_β carbon atoms by averaging the shift for each member of this Boltzmann weighted simulation ensemble calculated by the SHIFTS [175-177] and SHIFTX[178] programs. We similarly calculate the predicted scalar coupling constant ${}^3J_{\text{H}^{\text{NH}\alpha}}$ by evaluating the ϕ dihedral angle for every member of the correctly weighted ensemble and calculate the ensemble member's instantaneous scalar coupling constant. The values are then grouped and averaged in bins to obtain the overall average and standard deviation. We use the Karplus equation

$$J(\phi) = A \cos^2(\phi - 60) + B \cos(\phi - 60) + C \quad (5.1)$$

where $A = 6.51$, $B = -1.76$ and $C = 1.60$ [179]. The chemical shifts and predicted scalar coupling constant ${}^3J_{\text{H}^{\text{NH}\alpha}}$ were also calculated on a population on “unstructured” conformations. A conformation was considered unstructured if DSSP [174] did not identify any secondary structure category for any of the ten residues.

To calculate the spectral density functions for ${}^1\text{H}$ - ${}^1\text{H}$ spin pairs from the simulated data to predict ROESY data and for ${}^{13}\text{C}$ - ${}^1\text{H}$ pairs to predict the relaxation parameters (T_1 and T_2), we compute time correlation functions at 284K from constant energy (NVE) trajectories. Thirty NVE trajectories of 20ns for each peptide-water model combination were run. Starting structures for these trajectories were selected from the replica exchange simulation of each water model at 284K separated by 1ns to ensure decorrelation. Since only coordinate information (not velocity) was saved for the structural ensemble, structures were equilibrated at 284K for 100ps prior to the 20ns constant energy runs.

Following the method of Peter et al.[180], we evaluate the following time correlation function

$$C(\tau) = \left\langle \frac{1}{r^6(t)} \right\rangle^{-1} \left\langle \frac{P_2(\cos \chi_{t,t+\tau})}{r^3(t)r^3(t+\tau)} \right\rangle \quad (5.2)$$

from the simulation trajectories using ptraj, where P_2 is the 2nd order Legendre polynomial, χ is the angle between the interspin vector in the laboratory reference frame connecting each of the ~1800 pairs of protons of interest at time t and $t+\tau$, and $r(t)$ is the instantaneous pair distance, the angle brackets denote a thermal average and that average is normalized by the average spin-spin distance. Time correlation functions from the thirty trajectories are averaged. It is important to note that time correlations functions normalized by the inverse of the $1/r^6$ averaged spin-spin distance time must be reweighted by the the $1/r^6$ average spin-spin distance in each trajectory in order to compute the average over trajectories. This resulting average numerical correlation function is then fit to a triple exponential form, which is then Fourier transformed to define the spectral density functions

$$J(\omega) = \int_{-\infty}^{\infty} C(\tau) e^{i\omega\tau} d\tau \quad (5.3)$$

We note that the peptide tumbles and locally reorients rapidly enough for all relevant spin-spin vector time-correlation functions to decay to zero within the time of our 20ns dynamics simulations.

The quantities in Equations 5.2 and 5.3 allow us to define the T_1 and T_2 relaxation at the relevant ^1H and ^{13}C Larmor frequencies[169]. T_1 is defined as the inverse of the spin-lattice relaxation rate

$$R_1 = \frac{1}{T_1} = R_1^{\text{DD}} + R_1^{\text{CSA}} \quad (5.3a)$$

where R_1^{DD} and R_1^{CSA} are the dipolar and chemical shift anisotropy (CSA) components of the spin-lattice relaxation rate constants. T_2 is defined as the inverse of the spin-spin relaxation rate

$$R_2 = \frac{1}{T_2} = R_2^{\text{DD}} + R_2^{\text{CSA}} + R_a \quad (5.3b)$$

where R_2^{DD} and R_2^{CSA} are the dipolar and CSA components of the spin-spin relaxation rate constants and R_a is the sum of the relaxation rate constants for pseudo-first order processes such as chemical exchange and diffusion which we ignore for this analysis.

The dipolar relaxation rates

$$R_1^{\text{DD}} = (1/20)K^2[J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C)] \quad (5.3c)$$

$$R_2^{\text{DD}} = (1/40)K^2[4J(0) + J(\omega_H - \omega_C) + 3J(\omega_C) + 3J(\omega_H) + 6J(\omega_H + \omega_C)] \quad (5.3d)$$

are the main contributors to the overall relaxation. The constant factor K is defined as

$$K = \frac{\mu_0}{4\pi r_{\text{eff}}^3} \hbar \gamma_a \gamma_b \quad (5.3e)$$

where μ_0 is the permeability of free space, \hbar is Planck's constant, and γ_a and γ_b are the gyromagnetic ratios for the nuclei of interest which for the ^{13}C relaxation experiment are carbon and hydrogen, and r_{eff}

$$r_{eff} = \left\langle \frac{1}{r^6(t)} \right\rangle^{-1/6} \quad (5.3f)$$

is the appropriately averaged internuclear distance between atoms .

The chemical shift anisotropy contributions to the spin-lattice and spin-spin relaxation rates are incorporated in our T_1 and T_2 predictions by assuming an axially symmetric chemical shift tensor with chemical shift tensor parallel and perpendicular component difference, $\Delta\delta$, equal to 25ppm[169].

$$R_1^{CSA} = (2/5)\Delta\delta^2\omega_c^2 J(\omega_c) \quad (5.3g)$$

$$R_2^{CSA} = (2/5)\left(\Delta\delta^2\omega_c^2/6\right)[4J(0) + 3J(\omega_c)] \quad (5.3h)$$

We can therefore explicitly calculate the full time correlation functions and analytically FT these fits, and hence do not need to fit our spectral density functions through a Lipari-Szabo model-free analysis[181, 182], which has limited applicability when the system of interest lacks the separation of internal and external motion timescales as in the case of the A β_{21-30} peptide.

We also predict the ROESY spectra from our structural ensemble and dynamical trajectories by calculating the intensity

$$I(t_{mix}) = X e^{-\Lambda t_{mix}} X^{-1} I(0) \quad (5.4a)$$

where X and Λ are the eigenvectors and eigenvalues of the full relaxation matrix, R , composed of the diagonal elements

$$\rho_{ii} = \sum_{j=1, \neq i}^n \frac{1}{10} K^2 \left[\frac{3}{2} J_{ij}(2\omega_0) + \frac{9}{4} J_{ij}(\omega_0) + \frac{5}{4} J_{ij}(0) \right] \quad (5.4b)$$

and off-diagonal elements

$$\sigma_{ij} = \frac{1}{10} K^2 \left[\frac{3}{2} J_{ij}(\omega_0) + J_{ij}(0) \right] \quad (5.4c)$$

where ρ is the direct dipolar relaxation rate and σ is the cross-relaxation rate for all proton pairs as described by van Gunsteren and coworkers[180, 183] and K is defined as above, with $\gamma_a\gamma_b$ equal to γ_H^2 . Unlike the extended atom model (no aliphatic hydrogens) of these previous studies, we simulate all hydrogen atoms explicitly for each methyl group and hence calculate all pair correlation functions, including neighboring methylene and methyl group protons. We ignore water proton coordinates as is the standard assumption in the NMR experiment.

We solve this coupled system of differential equations for the magnetization matrix at the mixing time used by the NMR experiments for both H_2O and 2H_2O . We simulate the experimental conditions of heavy water solvation on the relaxation matrix by removing the exchangeable hydrogens including backbone amides (HN), hydrogens in basic NH_3^+ groups, and hydroxyl hydrogens (HO) from the spin-matrix which we accomplished by setting all pair distances with these protons to 30\AA . In order to generate peak predictions, we sum the peak volume contributions (including positive contributions from cross peaks dominated by spin diffusion) for degenerate spins (methyl groups) as well as those from spins within the same residue that are indistinguishable at the resolution of our NMR experiments. We note that this method explicitly accounts for peak intensity effects caused by methyl group rotation since spectral density functions are calculated for each proton in a methyl group and the individual intensities from indistinguishable peaks are summed to compare to experimental spectra. Amine peak

volumes in the H₂O spectra are scaled by a factor of 0.9 to approximately account for the presence of 10% deuterium exchanged amide protons from the 10% ²H₂O used for NMR lock. Predicted cross peaks to basic amine and hydroxyl groups are filtered from the predictions since these cross peaks would be significantly broadened by exchange with solvent protons on the NMR timescale.

5.3 Results

Chemical shifts

In Figure 5.1 we show the C_α, C_β, H_α and H_N experimental chemical shifts for Aβ₂₁₋₃₀ as compared to the calculated chemical shifts over our simulated ensemble for different force fields and different trajectories. For the experimental chemical shift values we subtracted the reference value of the chemical shift of a random coil at 25°C from the average shift for each amino acid[184]; while the carbon shifts show a very weak dependence on temperature and the 25°C random coil reference shift is used, the amide proton random coil shift reference for each amino acid is adjusted to a value appropriate for 10°C [185]. Together the chemical shift data emphasize that the peptide is largely unstructured.

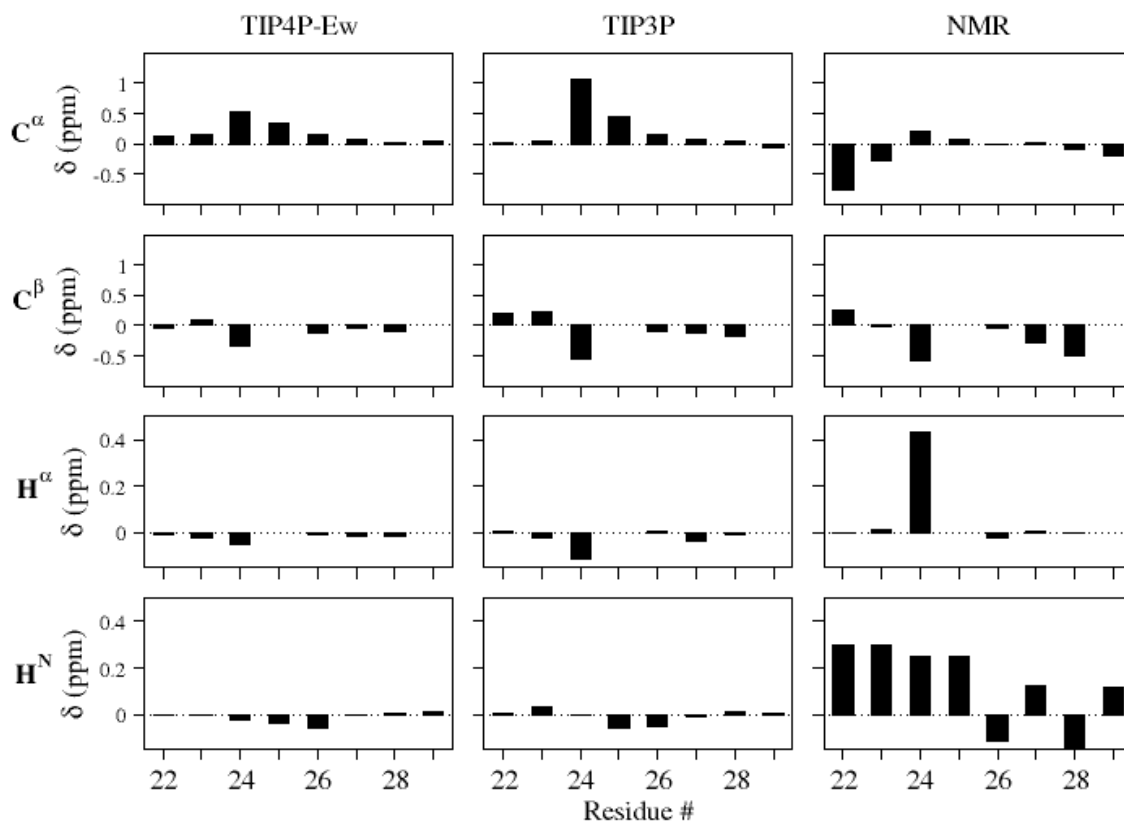


Figure 5.1: ^{13}C and ^1H chemical shifts from experiment and simulation for $\text{A}\beta_{21-30}$. Experimental NMR shifts are calculated as difference from the peptide measurements and tabulated random coil values. For ^{13}C we used the tabulated 25°C random coil shifts, while the reference amide proton shifts for each residue are adjusted to take into account the difference in temperature (25°C vs 10°C). The predicted chemical shifts are SHIFTS calculations averaged over the full ensemble after subtracting SHIFTS calculations averaged over the unstructured subpopulations (as defined by DSSP to represent our calculated reference random coil population).

In order to take into account any limitations of the calculations of the chemical shift theory, we calculate the average chemical shifts of our unstructured population, which comprises $\sim 60\%$ of the TIP4P-Ew and $\sim 40\%$ of the TIP3P ensemble. We then use these values as a random coil reference state that is subtracted from the total simulated

ensemble. The resulting ensemble averaged chemical shifts calculated with SHIFTS and SHIFTX are less than ~ 1 ppm for C_α and C_β shifts from our calculated random coil shifts, are similar between the TIP4P-Ew and TIP3P simulations, and show reasonable qualitative agreement with experiment. While the calculated amide proton shift values appear to deviate significantly from the NMR values, this may be due to the inability of the theory to capture all of the chemically relevant contributions to amide shifts[175], as opposed to inadequacy of the structural ensemble.

Scalar coupling measurements

The $^3J_{\text{H}^{\text{NH}}\alpha}$ scalar coupling constants, measured from a high resolution COSY spectrum, are compared to the prediction from the simulations by evaluating the ensemble average coupling constant calculated by the Karplus equation from the backbone ϕ dihedral angles along the peptide (Figure 5.2). The simulations show well converged values between trajectories and the two different water models show only small differences, less than 1 Hz. The overall agreement between simulation and experiment is quite good, making clear that both are consistent with an ensemble that is largely random coil, consistent with the chemical shift data.

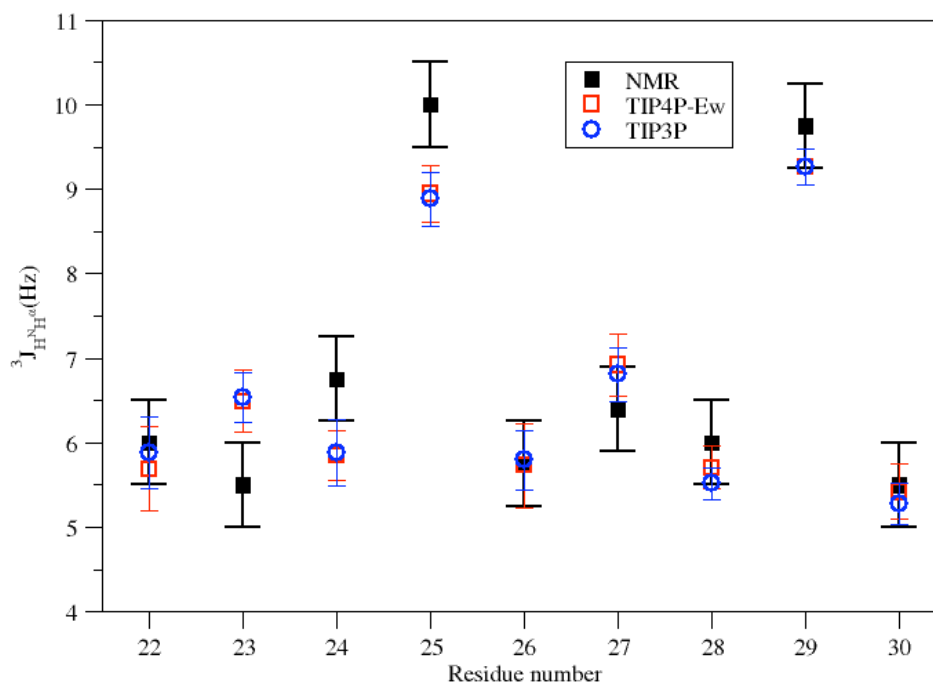


Figure 5.2: Scalar coupling constant ${}^3J_{\text{HNH}\alpha}$ calculated from the self consistently converged individual replica exchange ensembles for different empirical force fields and compared to experimentally determined coupling constants. Error bars are experimental uncertainty for NMR values as well as simulated standard deviations calculated for converged trajectories split into three sections. The scalar coupling constants for the glycines are added for the redundant C_α hydrogens in the simulation to connect to experiment, which is unable to distinguish them.

NMR Spin Relaxation

The C_α spin-lattice and spin-spin relaxation times measured for the non-glycine positions of $A\beta_{21-30}$ are presented in Table 5.1. Since the chemical shifts for ${}^{13}\text{C}$ - ${}^1\text{H}$ pairs for Glu22 and K28 overlap, T_1 and T_2 at these positions cannot be distinguished and the parameters are treated as an average of relaxation times at the two positions. Since the combined data for these positions fit well to a single exponential form, T_1 and T_2 relaxation times are similar for these positions. The experimentally determined relaxation

parameters vary by a maximum of 25% for non-terminal residues positions, indicating that the peptide does not contain significantly stiffer and/or slower moving regions on average.

T_1 and T_2 calculated from simulations of the peptide solvated with TIP4P-Ew model water show excellent agreement with the experimental values (Table 5.1). Non-terminal amino acid relaxation times are within $\pm 10\%$ for both T_1 and T_2 . Terminal amino acids show significantly longer relaxation times indicating faster motion, consistent with the experiments, although the simulated relaxation times are larger. The discrepancy is greatest for the C-terminal alanine, which may indicate that the simulations predict less structure in this region than in the experiment, although viscosity differences at lower temperature in $^2\text{H}_2\text{O}$ may influence the dynamics. To test this, we use a time scaling factor of 1.2 to elongate the time correlation functions as a simple approximation for the larger viscosity of $^2\text{H}_2\text{O}$ compared to $^1\text{H}_2\text{O}$ at 10°C , and we determine that the C-terminal T_1 and T_2 are then within 15% of the experimental values while the non-terminal relaxation parameters change by only a few percent (Table 5.1).

		A21	E22	D23	V24	S26	N27	K28	A30
T1	Experiment	415	298*	244	291	285	274	298*	475
At	TIP4P-Ew	492	292	268	272	274	273	287	530
500MHz	Scaled	457	276	253	260	262	261	275	485
	TIP3P	853	469	398	389	386	396	413	907
T2	Experiment	403	265*	269	230	235	241	265*	372
at	TIP4P-Ew	445	262	248	236	245	239	242	475
600MHz	Scaled	398	238	225	215	224	218	220	425
	TIP3P	860	474	403	386	386	393	402	858

Table 5.1: ^{13}C NMR Spin Relaxation times T_1 and T_2 for non-glycine C_α positions from experiment and TIP4P-Ew, TIP3P and time scaled TIP4P-Ew, in milliseconds. Glu22 and Lys28 resonances overlapped such that T_1 and T_2 could not be independently measured. The relaxation times calculated from the overlapped peaks are indicated with an asterisk.

T_1 and T_2 relaxation times calculated from the TIP3P simulations result are ~ 1.8 times larger than experimentally observed, for both terminal and non-terminal positions. This overestimate of relaxation time is a result of the faster dynamics of peptide motion in the TIP3P solvent. To demonstrate this difference the averaged vector autocorrelation function for the $^{13}\text{C}_\alpha$ and H_α pair at the Val24 position for both the TIP3P and TIP4P-Ew simulation is presented in Figure 5.3. The TIP3P simulations result in time correlation functions with fitted decay parameters more than twice that of the TIP4P-Ew simulations. The faster peptide motion in TIP3P is likely a result of the unrealistically low viscosity (faster self-diffusion) properties of the TIP3P model, speeding up the peptide dynamics.

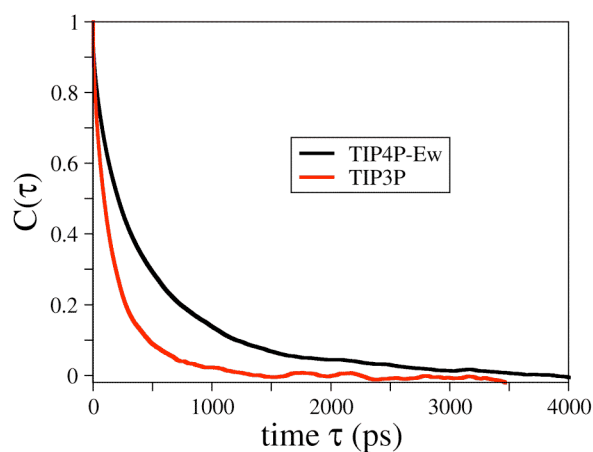


Figure 5.3: Normalized average vector time correlation function for Val24 C_α to H_α position for constant energy trajectories solvated with TIP4P-Ew and TIP3P. The TIP3P model (red) shows a dramatically faster decay for all vector time correlation functions relative to TIP4P-Ew, and we present this as an example.

Experimental and Simulated ROESY Crosspeaks

The 2D ROESY experiments in 90% H_2O :10% $^2\text{H}_2\text{O}$ and 100% $^2\text{H}_2\text{O}$ yielded a set of 155 assigned ROESY crosspeaks (see supplementary material for spectra). Although the majority of the crosspeaks were for intra-residue (83) and sequential (44) pairs, 28

weak medium range ROE interactions were also detected. These medium range ROE crosspeaks comprise several $i,i+2$ and $i,i+3$ interactions and two extremely weak $i,i+4$ interactions; no longer range ROE crosspeaks are observed and no strong patterns of α -helical or β -sheet contacts are evident. Lazo *et al.* report a long-range $i,i+8$ Glu22 H_α to Ala30 HN cross-peak[24], and more recent work from the same group assigns the peak to an overlap of both an $i,i+8$ Glu22 H_α and $i,i+2$ Lys28 H_α to Ala30 HN in a spectrum collected at 500MHz[23]. The observation of the long-range ROE is critical to their proposed NMR model, which is significantly collapsed. However, due to the higher resolution of the spectrum we have collected at 900 MHz, we interpret the cross peak to be solely attributed to the $i,i+2$ contact between Lys28 H_α and Ala30 HN, as shown in Figure 5.4.

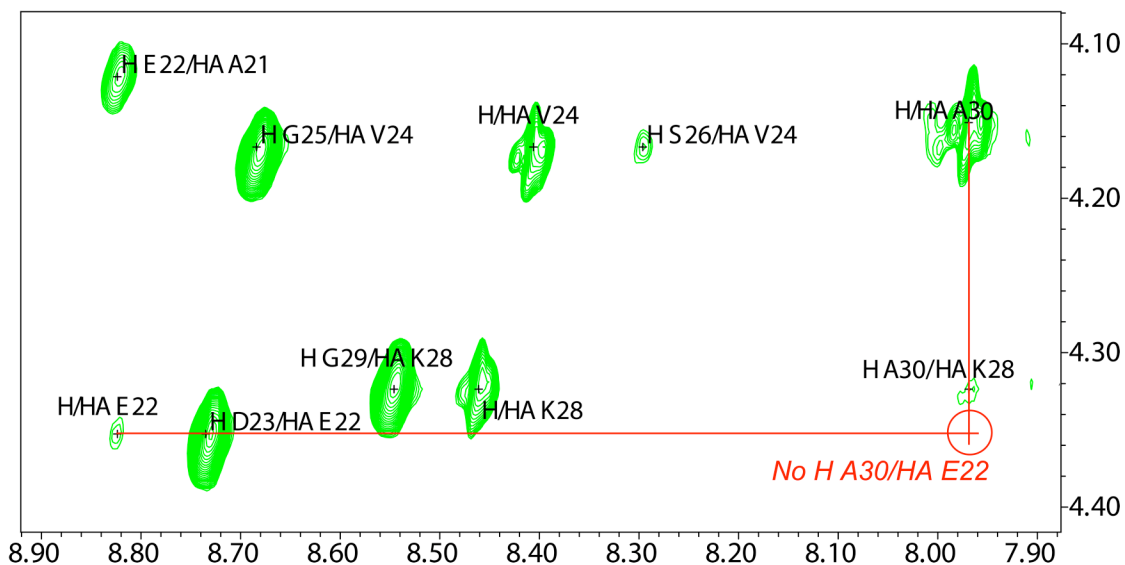


Figure 5.4: Fingerprint region of ROESY spectrum in H_2O of $A\beta_{21-30}$ demonstrating the cross peak interpreted by Lazo *et al.* and Grant *et al.* as H_α Glu22 with HN Ala30 in their 500MHz experiments is clearly resolved as only H_α Lys28 to HN Ala30 in our 900MHz experiment.

An additional set of long-range interactions between the Glu22 and Lys28 sidechains are reported by Lazo *et al.* In our higher field spectra, we are able to distinguish the small chemical shift difference between HB3 Lys28 and HB2 Glu22 (Figure 5.5). At lower field, these resonances would overlap very close to the limit of the resolution of the experiment and hence be difficult to distinguish. We see no evidence of true Glu22 to Lys28 ROE interactions and believe that the previously reported long range cross peaks between these residues are also due to misassignment of peaks too close to distinguish at lower field.

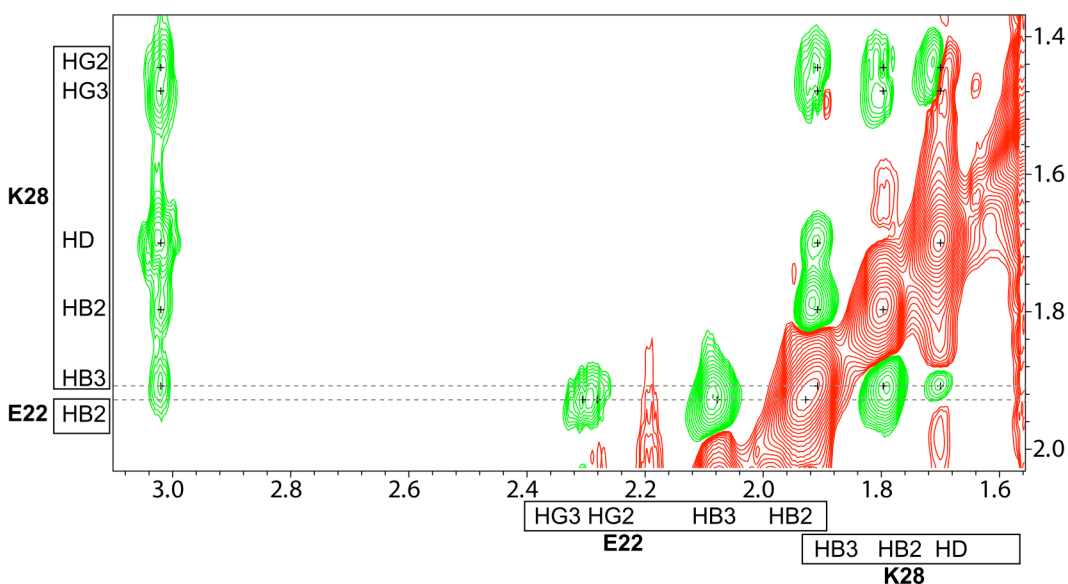


Figure 5.5: HB3 Lys28 has a nearly overlapping chemical shift with HB2 Glu22, potentially leading to cross peak misassignment in previous studies at lower field.

The majority of the medium range cross peaks suggest turn or partial collapse structure for residues 22 through 27, although medium-range interactions are also seen from Ala21 to Asp23 and Val24, indicating the peptide backbone is not simply extended in the N-terminal region. Two additional pairs of interactions are found in the C-terminal

region between the side chain and H_{α} of Lys28 with the backbone of Ala30, as well as the methyl group (HB) of Ala30 with both HB methylene atoms of Asn27.

Predicted ROESY cross peaks from simulation were calculated for a 300 ms mixing time and compared to the cross peaks observed in the H_2O and 2H_2O experimental spectra. By comparing the cross peaks predicted from the simulations to the experimentally observed cross peaks, the quality of the match between simulation and experiment can be evaluated. For each water model we have run two separate replica exchange simulations to explicitly evaluate convergence of the ROESY cross peaks. After 50ns, each simulation appears to be converged as commonly defined in the biomolecular simulation literature[144], where equilibrium averages of the quantities of interest, in this case the $\langle r^{-6} \rangle$ distances, change less than a few percent with additional simulation time. However these ensemble differences result in noticeable changes in the corresponding peak volumes that are derived from r^{-6} averages, and therefore peak ranking. We therefore specifically highlight below the few cross peaks where the combined ensemble cross peak predictions do not make evident the significantly different predictions from the independent simulations for both TIP4P-Ew and TIP3P.

In order to directly compare simulations to experiment, we must determine the constant relating the arbitrary experimental cross peak intensity scale to the simulated intensities for diagonal magnetization of unity at mixing time of 0 ms. Determining this value is especially important since the simulated cross peak intensity corresponding to the experimental noise level must be determined to evaluate the ability of the predictions to separate observed and unobserved peaks. The constant relating the simulation and

experimental scales is calculated by determining the slope of least-squares fit line (constrained to pass through the origin)[186] for the experimental intensities vs simulated cross peak volumes of all cross peaks for all distinguishable pairs separated by four or more bonds, an important condition since cross peaks between pairs separated by few bonds are subject to significant TOCSY intensity contributions, evident from lineshape distortion. The analysis includes all long range, medium range, and sequential peaks, as well as distant intra-residue pairs. This procedure was repeated for two water models, and two experimental solvation conditions (water or heavy water) where the ensembles from the two independent replica exchange simulations for each water model are combined. This constant factor was used to calculate the noise level by multiplying the intensity of the weakest peak assigned for the appropriate experiment (experimental intensity value of -700 for H₂O and -800 for ²H₂O).

Since medium and long ranged ROE interactions contain the most information about peptide structure, we focus only on these interactions and ignore intra residue and sequential interactions in our comparisons of simulations to experiment. The top-ranked intensities for the medium and long range ROE cross peaks predicted to be above the noise level are shown in Tables 5.2 and 5.3. The predictions are marked with a 1 if they are experimentally observed, 0 if they are not observed in the experiment, U for a potential experimental cross peak so weak that it cannot reliably be assigned as such. An important observation is that both TIP3P and TIP4P-Ew ensembles pick out many of the 28 experimentally observed medium-range ROE interactions from the 600 possible distinguishable medium and long-range interactions. These peaks that are both predicted

and observed are “true positives”, suggesting that both the TIP3P and TIP4P-Ew ensembles are similar to the true ensemble. The TIP4P-Ew simulations of the cross peak of strongest intensity between the HB methyl group of Ala21 and the HG methyl groups of Val24 led to the assignment of this peak in the experimental spectra, initially hidden beneath an experimental artifact.

The TIP4P-Ew and TIP3P ensembles have 12 and 14 experimentally assigned cross peaks, respectively, predicted above the noise from the 900 MHz H₂O, and 14 and 15, respectively, of the assigned cross peaks in the 800 MHz ²H₂O experiment (note that there is redundancy of cross peaks between the two experiments so that there are only 28 distinct cross-peaks in total). The true positive cross peaks encompass *i,i+2*, *i,i+3* and *i,i+4* medium range cross peaks from across the entire peptide.

The predictions also show a number of “false positive” cross peaks *i.e.* cross peak predicted above the estimated noise level in the simulations that are not experimentally observed. In the TIP4P-Ew predictions, all of the false positive cross peaks, excepting one from the H₂O and four additional from the ²H₂O spectra, are all *i,i+2* or *i,i+3* and involve residues and regions of the chain that have other observed ROE interactions, suggesting that the simulations are bringing together the correct regions of the peptide and that these false positives are due to small differences in the detailed structure, not wholesale failure of the empirical models. Due to the contribution of $1/r^6$ distance averaging to peak volumes, a difference of a factor of 2 in volume corresponds to only a factor of $2^{1/6}$ (~1.12) in average distance. In other words, false positive cross peaks predicted to be weak but still above the noise by our simulations may be just below the

background noise in the experiment if the simulated distances are closer than in the experiment by only a factor of 1.12. We therefore focus on false positives predicted more than a factor of 2 above the noise. In this case the quality of the water models are revealed in that the TIP3P simulations have far too many unobserved or unassignable cross peaks as the top predicted cross peaks. This suggests that the TIP3P ensemble has simulated distances for these pairs that are spatially closer than for most of the true positive medium range pairs in the simulation and closer than in the true ensemble measured in the experiment. By contrast, the TIP4P-Ew simulations have far fewer false positives topping the list, and hence are more likely to be close to the true ensemble. In fact, the TIP3P ensemble is clearly more collapsed, corresponding to more peak observations above the estimated noise as compared to TIP4P-Ew. Only a single long-range false positive is greater than a factor of two above the noise in the TIP4P-Ew predictions, and this high intensity is supported only by one of the two replica exchange simulations, suggesting that fully converged simulations may decrease the predicted intensity. Therefore we focus on the TIP4P-Ew results in the remainder of the Results.

The more significant discrepancies between the experiments and simulation involve “false negatives”- experimentally observed cross peaks that are not predicted to be observable by simulation. Since the noise in the experiment is not uniform over the spectra, if we dip just below the estimate for noise by just a factor of 2 reveals that the TIP4P-Ew simulations predict an additional set of 2 to 6 cross peaks in the two experiments. We would thus classify these as true positives as well.

However, the additional missing crosspeaks are a full factor of 10 below the noise level, and hence are genuine false negatives. One set of false negatives are the interactions between β methylene protons of Asp 23 and Ser 26, which involve four cross peaks seen experimentally. Since one of the possible four cross peaks of methylene pair interactions for these two residues is predicted among the top 10 $^2\text{H}_2\text{O}$ cross peaks by both TIP4P-Ew simulations, it is apparent that the simulation is bringing together the correct areas of the side-chains, but apparently not the correct detailed geometry in this region. We hypothesize that the structure of the turn stabilizing the Asp23-Ser26 sidechain contacts slightly favors the proximity of a single β -methylene pair from each residue, instead of a more even arrangement predicted by the experiments, with the further possibility that spin diffusion effects are suppressing the peaks further. One of the four cross peaks in the experiment is much smaller, indicating that even from the experimental observations, the distance between the protons is not expected to be entirely equal. Furthermore, the experimental cross peaks may imply more equivalent distances than in the actual underlying ensemble. Magnetization selectively ROE transferred through space to a single hydrogen in a methylene pair would subsequently be more evenly redistributed between the pair due to TOCSY type (through bond) transfer. TOCSY transfer is created by the rotating frame pulse during ROESY mixing, and is difficult to remove completely[165], and is an effect that is not accounted for by the predicted ROESY. These same arguments apply to the β methylene protons of Asn 27 and Ala 30, which involve four cross peaks seen experimentally, while the simulations find 1 of the crosspeaks above the noise, and 2 crosspeaks just below the noise, while the

Table 5.3: Merged peak volume predictions from simulations of each water model, TIP3P and TIP4P-Ew for 800MHz $^2\text{H}_2\text{O}$. See Table 5.2 caption for further detail.

TIP3P						TIP4P-Ew					
-0.00281	HA	VAL 24	HB3	ASN 27	U	-0.00287	HB	ALA 21	HG	VAL 24	-5950
-0.00239	HB3	ASP 23	HB2	SER 26	-1456	-0.00141	HG3	GLU 22	HG	VAL 24	-4090
-0.00147	HA	GLU 22	HA	ASN 27	0	-0.00113	HA	GLU 22	HA	ASN 27	0
-0.00138	HB	ALA 21	HG	VAL 24	-5950	-0.00108	HA	VAL 24	HB2	ASN 27	-1123
-0.00121	HG3	GLU 22	HG	VAL 24	-4090	-0.00101	HB3	ASP 23	HB2	SER 26	-1456
-0.00119	HA	VAL 24	HB2	ASN 27	-1123	-0.00097	HA	VAL 24	HB3	ASN 27	U
-0.00073	HB3	GLU 22	HG	VAL 24	-3028	-0.00096	HB2	GLU 22	HG	VAL 24	-1894
-0.00073	HG	VAL 24	HB3	ASN 27	-1403	-0.00082	HB3	GLU 22	HG	VAL 24	-3028
-0.00072	HG2	GLU 22	HG	VAL 24	-3402	-0.00078	HG2	GLU 22	HG	VAL 24	-3402
-0.00070	HA	ASP 23	HA	LYS 28	0	-0.00078	HA	GLU 22	HG	VAL 24	-647
-0.00065	HB3	ASP 23	HE2	LYS 28	0	-0.00066	HB	VAL 24	HB2	ASN 27	0
-0.00061	HG2	GLU 22	HA	ASN 27	0	-0.00057	HG	VAL 24	HB2	SER 26	-1068
-0.00061	HG	VAL 24	HB2	ASN 27	-1283	-0.00055	HG	VAL 24	HB3	ASN 27	-1403
-0.00060	HB2	GLU 22	HG	VAL 24	-1894	-0.00054	HG	VAL 24	HE2	LYS 28	-2498
-0.00058	HB	ALA 21	HA	ASN 27	0	-0.00053	HB3	ASP 23	HA	LYS 28	0
-0.00056	HA	VAL 24	HA	ASN 27	0	-0.00051	HG	VAL 24	HA	SER 26	0
-0.00053	HB	ALA 21	HB2	SER 26	0	-0.00051	HG	VAL 24	HB2	ASN 27	-1283
-0.00051	HG	VAL 24	HA	GLY 29	0	-0.00050	HA	VAL 24	HB2	LYS 28	0
-0.00049	HA	GLY 25	HA	GLY 29	0	-0.00050	HG2	GLU 22	HA	ASN 27	0
-0.00047	HG	VAL 24	HE2	LYS 28	-2498	-0.00048	HB	ALA 21	HB2	SER 26	0
-0.00045	HB3	ASP 23	HG3	LYS 28	0	-0.00048	HA	SER 26	HE2	LYS 28	0
-0.00044	HA	GLU 22	HA	GLY 29	0	-0.00047	HB	ALA 21	HB3	ASP 23	0
-0.00043	HG	VAL 24	HB2	SER 26	-1068	-0.00046	HB	ALA 21	HA	ASP 23	0
-0.00042	HG3	GLU 22	HB2	SER 26	0	-0.00045	HG	VAL 24	HA	ASN 27	0
-0.00042	HG3	LYS 28	HB	ALA 30	0	-0.00044	HA	GLY 25	HE2	LYS 28	0
-0.00039	HG2	GLU 22	HB2	ASN 27	0	-0.00039	HG	VAL 24	HA	LYS 28	-699
-0.00039	HG	VAL 24	HD2	LYS 28	0	-0.00039	HB2	SER 26	HE2	LYS 28	0
-0.00039	HA	GLU 22	HG	VAL 24	-647	-0.00037	HA	ASP 23	HB2	SER 26	0
-0.00038	HG3	GLU 22	HA	ASN 27	0	-0.00037	HB	ALA 21	HA	SER 26	0
-0.00036	HG	VAL 24	HB3	SER 26	0	-0.00037	HB2	ASN 27	HB	ALA 30	-1991
-0.00035	HB	ALA 21	HB	ALA 30	0	-0.00036	HG	VAL 24	HB3	SER 26	0
-0.00035	HB2	SER 26	HE2	LYS 28	0						
-0.00035	HA	GLY 25	HE2	LYS 28	0						
-0.00034	HB2	GLU 22	HA	GLY 29	0	-0.00032	HA	GLY 25	HD2	LYS 28	-1434
-0.00032	HA	ASP 23	HB2	SER 26	0	-0.00013	HB3	SER 26	HG3	LYS 28	-825
-0.00032	HA	GLY 25	HD2	LYS 28	-1434	-0.00003	HB3	ASP 23	HB3	SER 26	-1745
-0.00031	HB3	ASP 23	HB2	LYS 28	0	+0.00003	HB2	ASP 23	HB3	SER 26	-678
-0.00031	HB	VAL 24	HB3	ASN 27	0	+0.00004	HB3	ASN 27	HB	ALA 30	-3117
-0.00030	HB	ALA 21	HB2	ASN 27	0	+0.00011	HB2	ASP 23	HB2	SER 26	-1345
-0.00029	HA	SER 26	HE2	LYS 28	0						
-0.00029	HG	VAL 24	HA	ASN 27	0						
-0.00029	HA	GLU 22	HA	LYS 28	0						
-0.00029	HB3	ASN 27	HB	ALA 30	-3117						
-0.00029	HB	ALA 21	HB3	ASP 23	0						
-0.00029	HB3	ASP 23	HB3	SER 26	-1745						
-0.00028	HB3	ASP 23	HD2	LYS 28	0						
-0.00028	HB2	SER 26	HG3	LYS 28	0						
-0.00027	HB3	ASP 23	HA	GLY 25	0						
-0.00027	HE2	LYS 28	HB	ALA 30	0						
-0.00026	HB	ALA 21	HB2	LYS 28	0						
-0.00026	HB3	GLU 22	HA	ASN 27	0						
-0.00026	HB3	ASP 23	HG2	LYS 28	0						
-0.00026	HG	VAL 24	HA	SER 26	0						
-0.00025	HA	VAL 24	HE2	LYS 28	0						
-0.00025	HB	ALA 21	HA	ASP 23	0						
-0.00025	HG3	GLU 22	HB2	ASN 27	0						
-0.00025	HB	ALA 21	HA	ALA 30	0						
-0.00024	HB3	GLU 22	HB2	ASN 27	0						
-0.00024	HB	ALA 21	HA	VAL 24	0						
-0.00024	HG2	GLU 22	HA	GLY 29	0						
-0.00021	HG	VAL 24	HA	LYS 28	-699						
-0.00007	HB3	SER 26	HG3	LYS 28	-825						
-0.00002	HB2	ASN 27	HB	ALA 30	-1991						
-0.00001	HB2	ASP 23	HB3	SER 26	-678						
+0.00016	HB2	ASP 23	HB2	SER 26	-1345						

Peptide structure from experiments and simulations

Determining a single structure from multiple weak, medium range ROE experimental crosspeaks can be misleading since there is no guarantee that the underlying peptide adopts only a single structure instead of multiple distinct populations of structures

or perhaps a fully heterogeneous ensemble. Typical structure calculations on structured proteins assume that weak ROE cross peaks correspond to large ($\sim 4.5\text{-}6.0\text{\AA}$) distance restraints on a single well-defined structural population, and hence all the restraints should be applied simultaneously. Though we will show that the peptide ensemble involves significant disorder and hence it is inappropriate to use the standard structure determination methods[187-189], we calculate a single best-fit structure for purposes of comparison.

The set of restraints were used to calculate 1000 structures of $\text{A}\beta_{21-30}$ and the 50 lowest energy structures were aligned. For the entire peptide (Ala21 to Ala30), the superposition of the final 50 structures has an r.m.s. deviation of $0.81 \pm 0.42 \text{\AA}$ for the backbone atoms and $1.15 \pm 0.61 \text{\AA}$ for all heavy atoms. Sixteen of the twenty structures are within 1.0\AA r.m.s deviation for all heavy atoms, forming the dominant cluster depicted in Figure 5.6. The lowest energy structures from this minimization against all the ROESY restraints have three major bends, pinching together Asp23 to the Ser26, Gly25 to Lys28 and Asn27 to Ala30. These bends are created by 12 unique (i, i+3) and (i,i+4) ROE interactions between residues Asp23-Ser26, Val24-Asn27, Val24-Lys28, Gly25-Lys28, and Asn27-Ala30. Many of the crosspeaks for these interactions are very weak and detectable in the ROESY spectrum only when the peptide is dissolved in 100% $^2\text{H}_2\text{O}$. However, two or more ROE cross peaks were present for all of these (i, i+3) interactions, except for the interaction between Gly25-Lys28, for which only one cross peak between Gly25 HAs and Lys28 HDs was detected.

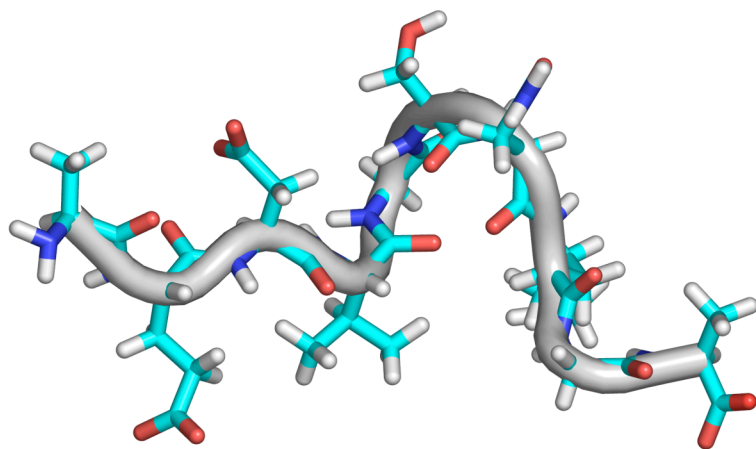


Figure 5.6: Representative structure from restraint energy minimized ensemble, simultaneously applying all observed ROE interactions.

Given the good quality of the TIP4P-Ew simulated experimental observables presented above, we can refine the experimental picture by analyzing the underlying simulated ensemble for structural populations. Standard clustering by RMSD is not informative due to the lack of order in much of the ensemble. The two first principal components in a PCA analysis yielded only a single large population, giving little information about the underlying structure. Since hydrophobic collapse is unlikely to be the dominant structuring force in a peptide that has only a single large aliphatic and no aromatic residues, hydrogen bond interactions are hypothesized to stabilize the structure that gives rise to the ROE interactions. Thus the most useful tool for partitioning the structures is the patterning of the hydrogen bonding and electrostatic interactions that may persist in sub-populations. It is important to emphasize that 60-65% of the TIP4P-Ew population is unstructured by these metrics, however the remaining 35-40% of the population does explain the presence of the medium-range restraints observed in the experimental ROESY crosspeaks with a large degree of success.

In the TIP4P-Ew simulations, the most populated interaction is between the sidechain carboxyl oxygens of Asp23 and the backbone amide of G25 (Figure 5.7A). This interaction is found in 17% of the ensemble. This interaction stabilizes the backbone dihedral angles near to those of a classic type I β -turn for residues Asp23 to Ser26. This type I β -turn is classically defined by backbone hydrogen bonding between residues i (Asp23) carbonyl oxygen with residue $i+3$ (Ser26) amide hydrogen, with the amide hydrogen of $i+2$ (Gly25) pointing toward the sidechain of i , precisely this most populated hydrogen bond we found. The classic type I β -turn, which is found in 5% of the

population, is therefore stabilized not only by the i to $i+3$ backbone hydrogen bonds, but also by the additional favorable interactions of Gly25 backbone amide hydrogen with the sidechain of Asp23 (Figure 5.7B).

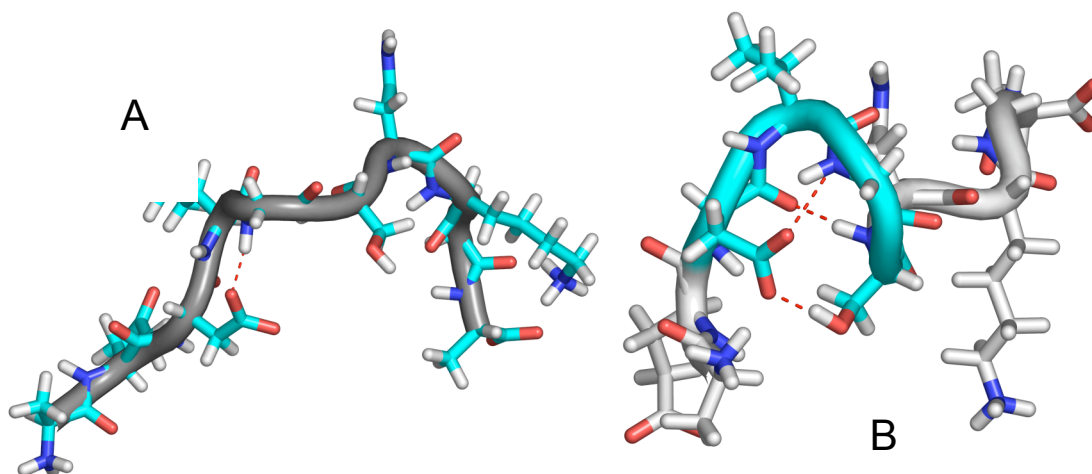


Figure 5.7: Representative structures showing A) the most populated N-terminal hydrogen bond (dotted red) and B) the hydrogen bonds and electrostatic interactions (dotted red) stabilizing the type I β -turn (cyan) centered at Val24 and Gly25.

The peptide maintains a turn with very similar backbone dihedrals if interactions between the carboxyl oxygens of Asp23 to the Ser26 backbone amide or with the sidechain hydroxyl of Ser26 are found, consistent with the observed ROEs which brings together the sidechain hydrogens of Asp23 and Ser26. If structures with at least one of these Asp23 to Ser26 interactions are also considered, the turn increases from 5% to 14% of the ensemble, much higher than any other turn region in the peptide. The TIP3P structural ensemble shows this turn in approximately 35% of the ensemble, and this larger amount of structured population contributes to the poor agreement of ROE cross peak volumes when compared to experiment.

Despite the prevalence of structures with type I β -turn structure, the β -turn does not nucleate a β -sheet, which would be characterized by backbone contacts between Glu22 and Asn27. Instead the backbone carbonyl oxygen of Asp23 to Asn27 backbone amide is an additional bond that often contributes to stabilizing the turn, precluding the formation of β -sheet structure flanking the β -turn. This interaction brings the Val24 methyl hydrogens near the Asn27 sidechain hydrogens, accounting for those observed ROE interactions. Other smaller groups of covarying hydrogen bonds are observed, including simultaneous interactions between the backbone carbonyl of Val24 with the backbone amide of both Asn27 and Lys28, bringing in proximity the Val24 sidechain with Asn27 sidechain as observed in the ROESY spectra.

Salt bridge formation between Asp23 and Lys28, observed in the solid state NMR structure of the $A\beta_{1-40}$ fibril, is found in 7% of the ensemble (Figure 5.8), while the competing salt bridge between Glu22 and Lys28 is found 1.5% of the time. Together these salt bridge structures are observed as often as the turn populations, but the salt-bridge contacts do not stabilize either hydrogen bonding structure or close proximity of other protons in the intervening region, suggesting these populations are not experimentally observable ROEs. A close contact involving the basic lysine amine occurring in only 7% may be observed in a ROESY spectrum, but salt bridges, unless stable enough to prevent proton exchange at neutral pH of the basic amine on the NMR experiment timescale, do not bring together NMR visible protons.

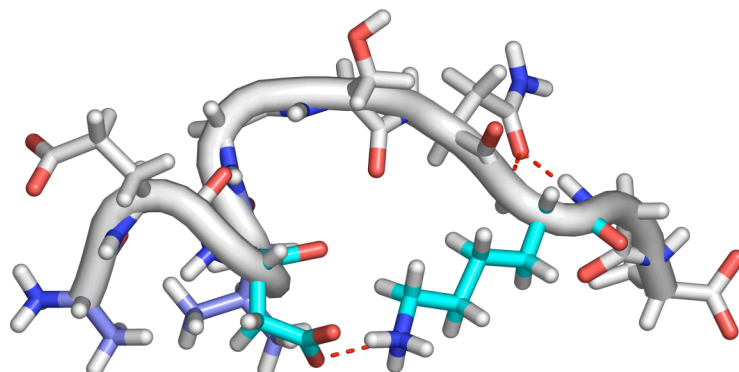


Figure 5.8: Asp23 to Lys28 salt-bridge (shown by red dotted line between aqua amino acids) found in 7% of the TIP4P-Ew ensemble, and Ala21 HB to Val24 HG (both in light blue) van der Waals contacts which give rise to the strongest observed ROE interaction.

5.4 Discussion

$A\beta_{21-30}$ shows no long-range and only weak medium-range ROE interactions, demonstrating none of the features of a protein with a single native state. We therefore conclude that the presence of a singly populated collapsed structure incorporating a unique bend due to a $i,i+8$ Glu22 H_α to Ala30 HN cross-peak and $i,i+6$ Glu22 sidechain to Lys28 sidechain hydrogens reported by Lazo *et al.* and Grant *et al.* is incorrect on two levels. The first is a problem of misassignment in their lower resolution ROESY spectra in which they determine a $i,i+8$ interaction that is instead revealed to be a weak $i,i+2$ interaction with the higher resolution 800 and 900 MHz spectra used here. We see no evidence of Glu22 to Lys28 $i,i+6$ interactions and instead suggest that the chemical shift overlap between HB3 Lys28 and HB2 E22 resulted in peak misassignment. More significantly, peptides and disordered protein systems should not conform to a single dominant structure, and should only be described by appropriate ensembles.

This poor quality of a single structure becomes evident when simultaneously applying all ROE interactions as distance restraints to give a minimized structure with surprisingly few favorable inter-residue interactions. One of the three bends in the minimized structure backbone is stabilized by one backbone hydrogen bond, and a second by a single Lys28 NH_3^+ interaction, but most of the pair distances for which ROE interactions are observed are not restrained by any favorable intermolecular interactions. Since there are only a few restraints that are all “weak” and hence provide only a loose upper bound on the distance, all of the restraints are satisfied by pair distances near this bound. This loose bounding results in a dominant structure with no consensus stabilizing contacts, hydrogen bonds, regular secondary structures or reverse turns.

We determine a good match between the TIP4P-Ew/ff99SB simulated and experimentally observed structure and dynamics, as measured by ROE cross peaks and ^{13}C relaxation, indicating that these simulations faithfully approximate the ensemble of structures interrogated by the experiments, but that are impossible to directly observe by experiment. We find that the structural ensemble of the $\text{A}\beta_{21-30}$ involves a majority (~60%) of unstructured population according to lack of any DSSP secondary structure assignment. However the remaining minority population involves ~14% population of β -turn structure centered at Val24 and G25 bringing together Asp23 and S26. A separate set of structures populated only by a few percent brings together the Val24 and Asn27 regions. The simulations also indicate that the Asp23 to Lys28 salt bridge, important to the fibril structure, is formed in ~7% of the ensemble.

The low population of structure in the A β_{21-30} peptide determined by both NMR experiments and molecular simulations is in contradiction to previous NMR and protease-resistance study and followup where the authors conclude that A β_{21-30} adopts a single predominant folded conformation. We have thoroughly addressed the lack of evidence for this conclusion by NMR and we note that absolute proteolysis is likely an insensitive tool to distinguish a partially ordered ensemble from a highly folded population.

5.5 Conclusion

The interplay of molecular simulation and a variety of NMR observables provides the correct tools for characterizing the structural ensemble for disordered systems, as measured by chemical shifts, scalar couplings, ROE cross peaks, and ^{13}C relaxation. The recent advent of improved empirical force fields, enhanced sampling techniques, and NMR relaxation matrix calculations allow us to predict well the high quality NMR experimental observables generated here. The ff99SB/TIP4P-Ew simulations do not overpredict contacts between regions where there are no experimentally observed ROEs and thus there are few false positive peaks, while false negatives are weak but often just buried in the noise. By contrast, the TIP3P simulations involve structural ensembles that are too collapsed, resulting in far too strong, and have far too many (false positive) ROE cross peaks. Additionally the solution dynamics are much too fast to correctly reproduce the observed ^{13}C relaxation times.

At the same time the experimental and predicted peaks do not match perfectly for ff99SB/TIP4P-EW, and the reasons for the discrepancy vary. The severity of $1/r^6$

averaging means that slight average distance changes can have greater than two-fold effects on the peak volume; therefore, regions of the chain brought together in close proximity may emphasize local geometric rotomers that are imperfectly captured by the empirical force fields. Potential experimental artifacts such as through-bond TOCSY transfer mechanisms not accounted for in the back calculation of ROESY cross-peaks may additionally contribute to prediction and experiment discrepancies. Even so, these structural and dynamical predictions match better than any previous predictions for peptides of this length range, a success that we attribute in particular to the advent of new generation force fields.

Faithfully predicting ROESY cross peaks is challenging since the nature of the ROESY experiment singles out the minority populations of close range interactions (if they exist) through a steep power dependence on distance, making them both more structurally informative but harder to converge in the simulated equilibrium ensemble. By contrast, each member of the structural ensemble contributes equally to population averaged quantities such as chemical shift and scalar coupling, making prediction of these quantities from simulations much easier to converge, but far less structurally informative. If we were only to have examined the chemical shift and scalar coupling ($^3J_{\text{HNH}\alpha}$) measures, we would have found that either the TIP3P or the TIP4P-Ew solvated ff99SB peptide would have equivalently reproduced the experimental data. For a system like this where the peptide exchanges conformations on timescales faster than the experimental timescale, the scalar coupling and chemical shift values provide direct Boltzmann weighted information dominated by the significant percentage of “random

coil” structure. Conversely, predicted ROE interactions are sensitive to minor populations of close contacts that distinguish a true partially structured ensemble from other partially structured, or even completely unstructured, ensembles. Predicting the ROE interactions is therefore a more stringent test for simulations which must distinguish a smaller population of heterogeneous structure from a high percentage background of unstructured or random coil like structure in the ensemble. As such this work serves as a validation study for ROESY characterization of $A\beta_{1-40}$ and $A\beta_{1-42}$ by molecular simulation, where collection of as detailed NMR data will be more challenging due to aggregation and fibril formation on experimental timescales at physiological conditions.

5.6 Acknowledgements

We thank Dr. Jeff Pelton and the Central California 900 MHz Facility for experimental resources and assistance. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Conclusion

This series of molecular simulation studies paired with experimental characterization have elucidated mechanisms by which the properties of the monomer unit influence the aggregation outcome and kinetics. Based on our detailed characterization of the wild type sequence $A\beta_{21-30}$ by simulation and NMR, we see no evidence of a folding nucleus. The presence or absence of a folding nucleus in $A\beta$ and its FAD mutants had previously been suggested to be the structural unit that seeds aggregate formation, explaining the differences in FAD aggregation characteristics[23, 24]. Instead, we believe that the effects of mutations on the fibril state, as discussed in chapter 4, are more likely the origin of differential fibril formation and morphology differences than any difference in structure in the monomer states. We hypothesize that mutational and other sequence effects on the structural stability of the ordered aggregate will account for most aggregation outcome differences in disease proteins, especially those involving disordered peptides or segments aggregation, including AD as well as polyglutamine diseases (e.g. Huntington's Disease) and Parkinson's Disease. These unstructured peptides lack the folding characteristics protecting non-disease proteins from aggregation. Not only are they missing a stable globular native state[190, 191], they lack fast forming intermediates or significant contacts and hence partial order in the denatured state, which contributed to the reduced aggregation of protein G relative to protein L. With our success in reproducing and interpreting structural and dynamical experimental observables with all-atom explicit simulations of $A\beta_{21-30}$, we believe we have validated this combination of empirical potentials for future studies of the full-length $A\beta$ peptides

and other fibril forming disease systems where the rapid sequestration of monomers into aggregates will make detailed experimental characterization more challenging. These highly detailed simulations may then provide input for the parameterization of new coarse-grained models. This combined work in coarse-grained and fully atomistic simulation of the early steps in protein aggregation paired with a tight connection to experimental observables has demonstrated how chemical and biophysical properties of the monomers result in aggregation protection for non-disease proteins and determine aggregation outcomes for disease proteins.

Bibliography

1. Pawar, A.P., et al., *Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases*. J Mol Biol, 2005. **350**(2): p. 379-92.
2. Goedert, M. and M.G. Spillantini, *A century of Alzheimer's disease*. Science, 2006. **314**(5800): p. 777-781.
3. Cellmer, T., et al., *Thermodynamics of folding and association of lattice-model proteins*. J Chem Phys, 2005. **122**(17): p. 174908.
4. Clark, E.D., *Protein refolding for industrial processes [Review]*. Current Opinion in Biotechnology, 2001. **12**(2): p. 202-207.
5. Dobson, C.M., *Protein folding and misfolding [Review]*. Nature, 2003. **426**(6968): p. 884-890.
6. Tycko, R., et al., *Probing the molecular structure of amyloid fibrils with solid state NMR*. Biophysical Journal, 2002. **82**(1): p. 187A-187A.
7. Brown, S. and T. Head-Gordon, *Intermediates and the folding of proteins L and G*. Protein Science, 2004. **13**(4): p. 958-970.
8. Brown, S., N. Fawzi, and T. Head-Gordon, *Coarse-Grained Sequences for Protein Folding and Design*. Proc Natl Acad Sci U S A, 2003. **100**: p. 10712-10717.
9. Ramirez-Alvarado, M., M.J. Cocco, and L. Regan, *Mutations in the B1 domain of protein G that delay the onset of amyloid fibril formation in vitro*. Protein Science, 2003. **12**(3): p. 567-576.
10. Cellmer, T., et al., *Kinetic studies of protein L aggregation and disaggregation*. Biophys Chem, 2007. **125**(2-3): p. 350-9.
11. Stefani, M. and C.M. Dobson, *Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution*. J Mol Med, 2003. **81**(11): p. 678-99.
12. Bucciantini, M., et al., *Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases*. Nature, 2002. **416**(6880): p. 507-11.
13. Petkova, a.T., W.M. Yau, and R. Tycko, *Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils*. Biochemistry, 2006. **45**(2): p. 498-512.
14. Hendriks, L., et al., *PRESENILE-DEMENTIA AND CEREBRAL-HEMORRHAGE LINKED TO A MUTATION AT CODON-692 OF THE BETA-AMYLOID PRECURSOR PROTEIN GENE*. Nature Genetics, 1992. **1**(3): p. 218-221.
15. Kamino, K., et al., *LINKAGE AND MUTATIONAL ANALYSIS OF FAMILIAL ALZHEIMER-DISEASE KINDREDS FOR THE APP GENE REGION*. American Journal of Human Genetics, 1992. **51**(5): p. 998-1014.
16. Nilsberth, C., et al., *The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced A beta protofibril formation*. Nature Neuroscience, 2001. **4**(9): p. 887-893.

17. Walsh, D.M., et al., *In vitro studies of amyloid beta-protein fibril assembly and toxicity provide clues to the aetiology of Flemish variant (Ala(692) -> Gly) Alzheimer's disease*. *Biochemical Journal*, 2001. **355**: p. 869-877.
18. Levy, E., et al., *Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type*. *Science*, 1990. **248**(4959): p. 1124-6.
19. Harper, J.D. and P.T. Lansbury, *Models of amyloid seeding in Alzheimer's disease and scrapie: Mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins*. *Annual Review of Biochemistry*, 1997. **66**: p. 385-407.
20. Hortschansky, P., et al., *The aggregation kinetics of Alzheimer's beta-amyloid peptide is controlled by stochastic nucleation*. *Protein Science*, 2005. **14**(7): p. 1753-1759.
21. Lomakin, a., et al., *Kinetic theory of fibrillogenesis of amyloid beta-protein*. *Proceedings of the National Academy of Sciences of the United States of America*, 1997. **94**(15): p. 7942-7947.
22. Meinhardt, J., et al., *Similarities in the thermodynamics and kinetics of aggregation of disease-related A beta(1-40) peptides*. *Protein Science*, 2007. **16**(6): p. 1214-1222.
23. Grant, M.A., et al., *Familial Alzheimer's disease mutations alter the stability of the amyloid beta-protein monomer folding nucleus*. *Proc Natl Acad Sci U S A*, 2007. **104**(42): p. 16522-7.
24. Lazo, N.D., et al., *On the nucleation of amyloid beta-protein monomer folding*. *Protein Sci*, 2005. **14**(6): p. 1581-96.
25. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters*. *Proteins*, 2006. **65**(3): p. 712-25.
26. Horn, H.W., et al., *Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew*. *J Chem Phys*, 2004. **120**(20): p. 9665-78.
27. Hukushima, K. and K. Nemoto, *Exchange Monte Carlo method and application to spin glass simulations*. *Journal of the Physical Society of Japan*, 1996. **65**(6): p. 1604-1608.
28. King, J., et al., *THERMOLABILE FOLDING INTERMEDIATES - INCLUSION BODY PRECURSORS AND CHAPERONIN SUBSTRATES [Review]*. *FASEB Journal*, 1996. **10**(1): p. 57-66.
29. Safar, J., et al., *Scrapie Amyloid (Prion) Protein Has the Conformational Characteristics of an Aggregated Molten Globule Folding Intermediate*. *Biochemistry*, 1994. **33**(27): p. 8375-8383.
30. Fink, A.L., *PROTEIN AGGREGATION - FOLDING AGGREGATES, INCLUSION BODIES AND AMYLOID [Review]*. *Folding & Design*, 1998. **3**(1): p. R 9-R 23.
31. Uversky, V.N., et al., *Association-Induced Folding of Globular Proteins*. *Proceedings of the National Academy of Sciences of the United States of America*, 1998. **95**(10): p. 5480-5483.
32. Silow, M., et al., *Formation of short-lived protein aggregates directly from the coil in two-state folding*. *Biochemistry*, 1999. **38**(40): p. 13006-13012.

33. Chiti, F., et al., *Designing conditions for in vitro formation of amyloid protofilaments and fibrils*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(7): p. 3590-3594.
34. Taddei, N., et al., *Stabilisation of alpha-helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding*. Journal of Molecular Biology, 2000. **300**(3): p. 633-647.
35. Gu, H.D., et al., *A Phage Display System for Studying the Sequence Determinants of Protein Folding*. Protein Science, 1995. **V4**(N6): p. 1108-1117.
36. Gu, H., Kim D., Baker, D., *Constrasting roles for symmetrically disposed beta-turns in the folding of a small protein*. Journal of Molecular Biology, 1997. **274**: p. 588-596.
37. Kim, D.E., H.D. Gu, and D. Baker, *The Sequences of Small Proteins Are Not Extensively Optimized for Rapid Folding by Natural Selection*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(9): p. 4982-4986.
38. Scalley, M.L., et al., *Kinetics of folding of the IgG binding domain of peptostreptococcal protein L*. Biochemistry, 1997. **V36**(N11): p. 3373-3382.
39. McCallister, E.L., E. Alm, and D. Baker, *Critical role of beta-hairpin formation in protein G folding*. Nature Structural Biology, 2000. **7**(8): p. 669-673.
40. Park, S.H., K.T. Oneil, and H. Roder, *An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core*. Biochemistry, 1997. **V36**(N47): p. 14277-14283.
41. Park, S.H., M.C.R. Shastry, and H. Roder, *Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing*. Nature Structural Biology, 1999. **6**(10): p. 943-947.
42. DeLano, W.L., *The PyMOL User's Manual*. 2002, San Carlos, California USA: DeLano Scientific.
43. Ramirez-Alvarado, M., J.S. Merkel, and L. Regan, *A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(16): p. 8979-8984.
44. Cellmer, T. and H. Blanch. 2004.
45. Dobson, C.M., *Protein folding and disease: a view from the first Horizon Symposium*. Nature Reviews, 2003. **Drug Discovery**. **2**(2): p. 154-160.
46. Polverino de Lauroto, P., et al., *Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis*. J Mol Biol, 2003. **334**(1): p. 129-41.
47. Yi, Q., et al., *NMR characterization of residual structure in the denatured state of protein L*. Journal of Molecular Biology, 2000. **299**(5): p. 1341-1351.
48. Horwich, A., *Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions*. Journal of Clinical Investigation, 2002. **110**(9): p. 1221-1232.

49. Mok, Y.K., et al., *NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions*. Journal of Molecular Biology, 1999. **289**(3): p. 619-638.
50. Kortemme, T., et al., *Similarities between the spectrin SH3 domain denatured state and its folding transition state*. Journal of Molecular Biology, 2000. **297**(5): p. 1217-1229.
51. Crowhurst, K.A., et al., *Corrigendum to the paper by Mok et al (1999)'NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions'*. Journal of Molecular Biology, 2003. **329**: p. 185-187.
52. Daggett, V. and A.R. Fersht, *Is there a unifying mechanism for protein folding? [Review]*. Trends in Biochemical Sciences, 2003. **28**(1): p. 18-25.
53. Capaldi, A.P., C. Kleanthous, and S.E. Radford, *Im7 folding mechanism: misfolding on a path to the native state*. Nature Structural Biology, 2002. **9**(3): p. 209-216.
54. Capaldi, A.P., et al., *Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate*. Nature Structural Biology, 2001. **8**(1): p. 68-72.
55. Sanchez, I.E. and T. Kiefhaber, *Hammond behavior versus ground state effects in protein folding: Evidence for narrow free energy barriers and residual structure in unfolded states*. Journal of Molecular Biology, 2003. **327**(4): p. 867-884.
56. Wright, C.F., et al., *Parallel protein-unfolding pathways revealed and mapped*. Nature Structural Biology, 2003. **10**(8): p. 658-662.
57. Lindberg, M., J. Tangrot, and M. Oliveberg, *Complete change of the protein folding transition state upon circular permutation*. Nature Structural Biology, 2002. **9**(11): p. 818-822.
58. Head-Gordon, T. and S. Brown, *Minimalist models for protein folding and design*. Current Opinion in Structural Biology, 2003. **13**(2): p. 160-167.
59. Sorenson, J. and T. Head-Gordon, *Matching simulation and experiment: A new simplified model for protein folding*. Journal of Computational Biology, 2000. **7**(3/4): p. 469-481.
60. Alm, E., et al., *Simple physical models connect theory and experiment in protein folding kinetics*. Journal of Molecular Biology, 2002. **322**(2): p. 463-476.
61. Plaxco, K.W., K.T. Simons, and D. Baker, *Contact order, transition state placement and the refolding rates of single domain proteins*. Journal of Molecular Biology, 1998. **V277**(N4): p. 985-994.
62. Broome, B.M. and M.H. Hecht, *Nature disfavors sequences of alternating polar and non-polar amino acids: Implications for amyloidogenesis*. Journal of Molecular Biology, 2000. **296**(4): p. 961-968.
63. Schwartz, R., S. Istrail, and J. King, *Frequencies of amino acid strings in globular protein sequences indicate suppression of blocks of consecutive hydrophobic residues*. Protein Science, 2001. **10**(5): p. 1023-1031.
64. West, M.W., et al., *De novo amyloid proteins from designed combinatorial libraries*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(20): p. 11211-11216.

65. Ferguson, D.M. and D.G. Garrett, *Simulated annealing - Optimal histogram methods*. Monte Carlo Methods in Chemical Physics, 1999. **105**: p. 311-336.
66. Kumar, S., et al., *Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method*. Journal of Computational Chemistry, 1995. **V16(N11)**: p. 1339-1350.
67. Du, R., et al., *On the transition coordinate for protein folding*. Journal of Chemical Physics, 1998. **V108(N1)**: p. 334-350.
68. Dobson, C.M., *Protein folding and misfolding*. Nature, 2003. **426(6968)**: p. 884-890.
69. Dobson, C.M., *Principles of protein folding, misfolding and aggregation*. Semin Cell Dev Biol, 2004. **15(1)**: p. 3-16.
70. Cohen, F.E. and J.W. Kelly, *Therapeutic approaches to protein-misfolding diseases*. Nature, 2003. **426(6968)**: p. 905-9.
71. Harper, J.D. and P.T. Lansbury, Jr., *Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins*. Annu Rev Biochem, 1997. **66**: p. 385-407.
72. Jarrett, J.T. and P.T. Lansbury, Jr., *Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie?* Cell, 1993. **73(6)**: p. 1055-8.
73. Kim, W. and M.H. Hecht, *Sequence determinants of enhanced amyloidogenicity of Alzheimer A β ₄₂ peptide relative to A β ₄₀*. J Biol Chem, 2005. **280(41)**: p. 35069-76.
74. Lomakin, a., et al., *On the nucleation and growth of amyloid beta-protein fibrils: Detection of nuclei and quantitation of rate constants*. Proceedings of the National Academy of Sciences of the United States of America, 1996. **93(3)**: p. 1125-1129.
75. Crescenzi, O., et al., *Solution structure of the Alzheimer amyloid beta-peptide (1-42) in an apolar microenvironment. Similarity with a virus fusion domain*. Eur J Biochem, 2002. **269(22)**: p. 5642-8.
76. Tomaselli, S., et al., *The alpha-to-beta conformational transition of Alzheimer's A β ₍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)**: p. 257-67.
77. Baumketner, A., et al., *Amyloid beta-protein monomer structure: a computational and experimental study*. Protein Sci, 2006. **15(3)**: p. 420-8.
78. Bernstein, S.L., et al., *Amyloid beta-protein: Monomer structure and early aggregation states of A β ₄₂ and its Pro(19) alloform*. Journal of the American Chemical Society, 2005. **127(7)**: p. 2075-2084.
79. Lazo, N.D., et al., *On the nucleation of amyloid beta-protein monomer folding*. Protein Science, 2005. **14(6)**: p. 1581-1596.
80. Sciarretta, K.L., et al., *A beta 40-Lactam(D23/K28) models a conformation highly favorable for nucleation of amyloid*. Biochemistry, 2005. **44(16)**: p. 6003-6014.

81. Petkova, A.T., et al., *Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils*. Science, 2005. **307**(5707): p. 262-265.
82. Petkova, A.T., et al., *A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(26): p. 16742-16747.
83. Balbach, J.J., et al., *Supramolecular structure in full-length Alzheimer's beta-amyloid fibrils: Evidence for a parallel beta-sheet organization from solid-state nuclear magnetic resonance*. Biophysical Journal, 2002. **83**(2): p. 1205-1216.
84. Buchete, N.V., R. Tycko, and G. Hummer, *Molecular dynamics simulations of Alzheimer's beta-amyloid protofilaments*. J Mol Biol, 2005. **353**(4): p. 804-21.
85. Dolphin, G.T., P. Dumy, and J. Garcia, *Control of Amyloid beta-Peptide Protofibril Formation by a Designed Template Assembly*. Angew Chem Int Ed Engl, 2006. **45**(17): p. 2699-2702.
86. Ferrone, F., *Analysis of protein aggregation kinetics*. Methods Enzymol, 1999. **309**: p. 256-74.
87. Favrin, G., A. Irback, and S. Mohanty, *Oligomerization of amyloid A beta(16-22) peptides using hydrogen bonds and hydrophobicity forces*. Biophysical Journal, 2004. **87**(6): p. 3657-3664.
88. Klimov, D.K., J.E. Straub, and D. Thirumalai, *Aqueous urea solution destabilizes A beta(16-22) oligomers*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(41): p. 14760-14765.
89. Klimov, D.K. and D. Thirumalai, *Dissecting the assembly of A beta(16-22) amyloid peptides into antiparallel beta sheets*. Structure, 2003. **11**(3): p. 295-307.
90. Urbanc, B., et al., *Molecular dynamics simulation of amyloid beta dimer formation*. Biophysical Journal, 2004. **87**(4): p. 2310-2321.
91. Urbanc, B., et al., *In silico study of amyloid beta-protein folding and oligomerization*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(50): p. 17345-17350.
92. Brown, S., N.J. Fawzi, and T. Head-Gordon, *Coarse-grained sequences for protein folding and design*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(19): p. 10712-10717.
93. Fawzi, N.L., et al., *Influence of denatured and intermediate states of folding on protein aggregation*. Protein Science, 2005. **14**(4): p. 993-1003.
94. Sorenson, J.M. and T. Head-Gordon, *Matching simulation and experiment: A new simplified model for simulating protein folding*. Journal of Computational Biology, 2000. **7**(3-4): p. 469-481.
95. Sorenson, J.M. and T. Head-Gordon, *Protein engineering study of protein L by simulation*. Journal of Computational Biology, 2002. **9**(1): p. 35-54.
96. Sorenson, J.M. and T. Head-Gordon, *Redesigning the hydrophobic core of a model beta-sheet protein: Destabilizing traps through a threading approach*. Proteins-Structure Function and Genetics, 1999. **37**(4): p. 582-591.
97. Ban, T., et al., *Direct observation of Abeta amyloid fibril growth and inhibition*. J Mol Biol, 2004. **344**(3): p. 757-67.

98. Baumketner, A., et al., *Amyloid beta-protein monomer structure: A computational and experimental study*. Protein Science, 2006. **15**(3): p. 420-428.
99. Baumketner, A., et al., *Structure of the 21-30 fragment of amyloid beta-protein*. Protein Science, 2006. **15**(6): p. 1239-1247.
100. Lazo, N., et al., *Probing nucleation of A β monomer folding*. Neurobiology of Aging, 2004. **25**: p. S149-S149.
101. Roher, A.E., et al., *beta-Amyloid-(1-42) is a major component of cerebrovascular amyloid deposits: implications for the pathology of Alzheimer disease*. Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10836-40.
102. Gravina, S.A., et al., *Amyloid beta protein (A β) in Alzheimer's disease brain. Biochemical and immunocytochemical analysis with antibodies specific for forms ending at A β 40 or A β 42(43)*. J Biol Chem, 1995. **270**(13): p. 7013-6.
103. Selkoe, D.J., *The molecular pathology of Alzheimer's disease*. Neuron, 1991. **6**(4): p. 487-98.
104. Luhrs, T., et al., *3D structure of Alzheimer's amyloid-beta(1-42) fibrils*. Proc Natl Acad Sci U S A, 2005. **102**(48): p. 17342-7.
105. Kaye, R., et al., *Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis*. Science, 2003. **300**(5618): p. 486-9.
106. Ben-Naim, A., *Statistical Mechanics of Waterlike Particles in 2 Dimension. I. Physical Model and Application of Percus-Yevick Equation*. Journal of Chemical Physics, 1971. **54**(9): p. 3682.
107. Silverstein, K.A.T., A.D.J. Haymet, and K.A. Dill, *A simple model of water and the hydrophobic effect*. Journal of the American Chemical Society, 1998. **120**(13): p. 3166-3175.
108. Andersen, H.C., *RATTLE - A VELOCITY VERSION OF THE SHAKE ALGORITHM FOR MOLECULAR-DYNAMICS CALCULATIONS*. Journal of Computational Physics, 1983. **52**(1): p. 24-34.
109. DeLano, W.L., *The PyMOL Molecular Graphics System*. 2002, DeLano Scientific: San Carlos, CA, USA.
110. Bugiani, O., et al., *An Italian type of HCHWA*. Neurobiol Aging, 1998. **19**: p. S238.
111. Tagliavini, F., et al., *A new betaPP mutation related to hereditary cerebral hemorrhage*. Alzheimer's reports: vascular factors in Alzheimer's disease, 1999. **2**(Suppl1): p. S28.
112. Grabowski, T.J., et al., *Novel amyloid precursor protein mutation in an Iowa family with dementia and severe cerebral amyloid angiopathy*. Annals of Neurology, 2001. **49**(6): p. 697-705.
113. Van Nostrand, W.E., et al., *Pathogenic effects of cerebral amyloid angiopathy mutations in the amyloid beta-protein precursor, in Alzheimer's Disease: Vascular Etiology and Pathology*. 2002, New York Acad Sciences: New York. p. 258-265.
114. Morelli, L., et al., *Differential degradation of amyloid beta genetic variants associated with hereditary dementia or stroke by insulin-degrading enzyme*. Journal of Biological Chemistry, 2003. **278**(26): p. 23221-23226.

115. Stern, E.A., et al., *Cortical synaptic integration in vivo is disrupted by amyloid-beta plaques*. Journal of Neuroscience, 2004. **24**(19): p. 4535-4540.
116. Bucciantini, M., et al., *Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases*. Nature, 2002. **416**(6880): p. 507-511.
117. Stefani, M. and C.M. Dobson, *Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution*. Journal of Molecular Medicine-Jmm, 2003. **81**(11): p. 678-699.
118. Braak, H. and E. Braak, *Evolution of neuronal changes in the course of Alzheimer's disease*. Journal of Neural Transmission-Supplement, 1998(53): p. 127-140.
119. Lambert, M.P., et al., *Diffusible, nonfibrillar ligands derived from A beta(1-42) are potent central nervous system neurotoxins*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(11): p. 6448-6453.
120. Stine, W.B., et al., *In vitro characterization of conditions for amyloid-beta peptide oligomerization and fibrillogenesis*. Journal of Biological Chemistry, 2003. **278**(13): p. 11612-11622.
121. Klyubin, I., et al., *Soluble Arctic amyloid beta protein inhibits hippocampal long-term potentiation in vivo*. European Journal of Neuroscience, 2004. **19**(10): p. 2839-2846.
122. Dolphin, G.T., P. Dumy, and J. Garcia, *Control of amyloid beta-peptide protofibril formation by a designed template assembly*. Angewandte Chemie-International Edition, 2006. **45**(17): p. 2699-2702.
123. Ferrone, F., *Analysis of protein aggregation kinetics*, in *Amyloid, Prions, and Other Protein Aggregates*. 1999, Academic Press Inc: San Diego. p. 256-274.
124. Nguyen, P.H., et al., *Monomer adds to preformed structured oligomers of A beta-peptides by a two-stage dock-lock mechanism*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(1): p. 111-116.
125. Tarus, B., J.E. Straub, and D. Thirumalai, *Probing the initial stage of aggregation of the A beta(10-35)-protein: Assessing the propensity for peptide dimerization*. Journal of Molecular Biology, 2005. **345**(5): p. 1141-1156.
126. Baumketner, A. and J.E. Shea, *Folding landscapes of the Alzheimer amyloid-beta(12-28) peptide*. Journal of Molecular Biology, 2006. **362**(3): p. 567-579.
127. Buchete, N.V. and G. Hummer, *Structure and dynamics of parallel beta-sheets, hydrophobic core, and loops in Alzheimer's A beta fibrils*. Biophysical Journal, 2007. **92**(9): p. 3032-3039.
128. Hills, R.D. and C.L. Brooks, *Hydrophobic cooperativity as a mechanism for amyloid nucleation*. Journal of Molecular Biology, 2007. **368**(3): p. 894-901.
129. Huet, A. and P. Derreumaux, *Impact of the mutation A21G (Flemish variant) on Alzheimer's beta-amyloid dimers by molecular dynamics simulations*. Biophysical Journal, 2006. **91**(10): p. 3829-3840.
130. Ma, B.Y. and R. Nussinov, *Stabilities and conformations of Alzheimer's beta-amyloid peptide oligomers (A beta(16-22) A beta(16-35) and A beta(10-35))*:

- Sequence effects*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(22): p. 14126-14131.
131. Tarus, B., J.E. Straub, and D. Thirumalai, *Dynamics of Asp23-Lys28 salt-bridge formation in A beta(10-35) monomers*. Journal of the American Chemical Society, 2006. **128**(50): p. 16159-16168.
 132. Xu, Y.C., et al., *Conformational transition of amyloid beta-peptide*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(15): p. 5403-5407.
 133. Wei, G.H. and J.E. Shea, *Effects of solvent on the structure of the Alzheimer amyloid-beta(25-35) peptide*. Biophysical Journal, 2006. **91**(5): p. 1638-1647.
 134. Zheng, J., et al., *Modeling the Alzheimer A{beta}17-42 Fibril Architecture: Tight Intermolecular Sheet-Sheet Association and Intramolecular Hydrated Cavities*. Biophys J, 2007. **93**(9): p. 3046-57.
 135. Yap, E.H., N.L. Fawzi, and T. Head-Gordon, *A Coarse-Grained alpha-Carbon Protein Model with Anisotropic Hydrogen-Bonding*. Proteins-Structure Function and Bioinformatics, 2007. **in press**.
 136. Fawzi, N.L., et al., *Determining the critical nucleus and mechanism of fibril elongation of the Alzheimer's A beta(1-40) peptide*. Journal of Molecular Biology, 2007. **365**(2): p. 535-550.
 137. Westermarck, G.T., K.H. Johnson, and P. Westermarck, *Staining methods for identification of amyloid in tissue*, in *Amyloid, Prions, and Other Protein Aggregates*. 1999, Academic Press Inc: San Diego. p. 3-25.
 138. Sikorski, P. and E. Atkins, *New model for crystalline polyglutamine assemblies and their connection with amyloid fibrils*. Biomacromolecules, 2005. **6**(1): p. 425-432.
 139. Bucciantini, M., et al., *Prefibrillar amyloid protein aggregates share common features of cytotoxicity*. Journal of Biological Chemistry, 2004. **279**(30): p. 31374-31382.
 140. Petkova, A.T., et al., *Structural investigations of Alzheimer's beta-amyloid fibrils by solid state NMR*. Biophysical Journal, 2004. **86**(1): p. 506A-506A.
 141. Coles, M., et al., *Solution structure of amyloid beta-peptide(1-40) in a water-micelle environment. Is the membrane-spanning domain where we think it is?* Biochemistry, 1998. **37**(31): p. 11064-77.
 142. Shao, H., et al., *Solution structures of micelle-bound amyloid beta-(1-40) and beta-(1-42) peptides of Alzheimer's disease*. J Mol Biol, 1999. **285**(2): p. 755-73.
 143. Sticht, H., et al., *Structure of amyloid A4-(1-40)-peptide of Alzheimer's disease*. Eur J Biochem, 1995. **233**(1): p. 293-8.
 144. Sgourakis, N.G., et al., *The Alzheimer's peptides Abeta40 and 42 adopt distinct conformations in water: a combined MD / NMR study*. J Mol Biol, 2007. **368**(5): p. 1448-57.
 145. Yan, Y., et al., *Methyl dynamics of the amyloid-beta peptides Abeta40 and Abeta42*. Biochem Biophys Res Commun, 2007. **362**(2): p. 410-4.

146. Yan, Y. and C. Wang, *Abeta42 is more rigid than Abeta40 at the C terminus: implications for Abeta aggregation and toxicity*. J Mol Biol, 2006. **364**(5): p. 853-62.
147. Hendriks, L., et al., *Presenile-dementia and cerebral-hemorrhage linked to a mutation at codon-492 of the beta-amyloid precursor protein gene*. Nature Genetics, 1992. **1**(3): p. 218-221.
148. Van Nostrand, W.E., et al., *Localization of a fibrillar amyloid beta-protein binding domain on its precursor*. Journal of Biological Chemistry, 2002. **277**(39): p. 36392-36398.
149. Baumketner, A., et al., *Structure of the 21-30 fragment of amyloid beta-protein*. Protein Sci, 2006. **15**(6): p. 1239-47.
150. Borreguero, J.M., et al., *Folding events in the 21-30 region of amyloid beta-protein (Abeta) studied in silico*. Proc Natl Acad Sci U S A, 2005. **102**(17): p. 6015-20.
151. Cruz, L., et al., *Solvent and mutation effects on the nucleation of amyloid beta-protein folding*. Proc Natl Acad Sci U S A, 2005. **102**(51): p. 18258-63.
152. Chen, W., N. Mousseau, and P. Derreumaux, *The conformations of the amyloid-beta (21-30) fragment can be described by three families in solution*. J Chem Phys, 2006. **125**(8): p. 084911.
153. Marsh, J.A., et al., *Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation*. Protein Sci, 2006. **15**(12): p. 2795-804.
154. Marsh, J.A., et al., *Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure*. J Mol Biol, 2007. **367**(5): p. 1494-510.
155. Choy, W.Y. and J.D. Forman-Kay, *Calculation of ensembles of structures representing the unfolded state of an SH3 domain*. J Mol Biol, 2001. **308**(5): p. 1011-32.
156. Bezsonova, I., et al., *Oxygen as a paramagnetic probe of clustering and solvent exposure in folded and unfolded states of an SH3 domain*. J Am Chem Soc, 2007. **129**(6): p. 1826-35.
157. Mu, Y.G., et al., *Conformational dynamics of trialanine in water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments*. Journal of Physical Chemistry B, 2003. **107**(21): p. 5064-5073.
158. Mu, Y.G., G. Stock, and U. W. Schiller, *Conformational dynamics of trialanine in water: A molecular dynamics study*. Journal of Physical Chemistry B, 2002. **106**(20): p. 5294-5301.
159. Case, D.A., *Molecular dynamics and NMR spin relaxation in proteins*. Acc Chem Res, 2002. **35**(6): p. 325-31.
160. Showalter, S.A., et al., *Toward Quantitative Interpretation of Methyl Side-Chain Dynamics from NMR by Molecular Dynamics Simulations*. J Am Chem Soc, 2007.

161. Graf, J., et al., *Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study*. J Am Chem Soc, 2007. **129**(5): p. 1179-89.
162. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes*. J Biomol NMR, 1995. **6**(3): p. 277-93.
163. Johnson, B.A., R.A. Blevins, and Pe, *NMR VIEW - A computer-program for the visualization and analysis of NMR data*. Journal of Biomolecular NMR, 1994. **4**(5): p. 603-614.
164. Keller, R., *The Computer Aided Resonance Assignment Tutorial*. 2004: CANTINA Verlag.
165. Hwang, T.L., A.J. Shaka, and Hm, *Cross relaxation without TOCSY - Transverse rotating-frame Overhauser effect spectroscopy*. Journal of the American Chemical Society, 1992. **114**(8): p. 3157-3159.
166. Guntert, P., *Automated NMR structure calculation with CYANA*. Methods Mol Biol, 2004. **278**: p. 353-78.
167. DeLano, W.L., *The PyMOL Molecular Graphics System*. 2002, San Carlos, CA, USA.: DeLano Scientific.
168. Case, D.A., et al., *The Amber biomolecular simulation programs*. J Comput Chem, 2005. **26**(16): p. 1668-88.
169. Palmer, A.G., M. Rance, and P.E. Wright, *Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-13 Heteronuclear NMR Spectroscopy*. Journal of the American Chemical Society, 1991. **113**(12): p. 4371-4380.
170. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water*. Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
171. Darden, T., D. York, and L. Pedersen, *Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems*. Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.
172. Geyer, C.J. and E.a. Thompson, *Annealing Markov-Chain Monte-Carlo with Applications to Ancestral Inference*. Journal of the American Statistical Association, 1995. **90**(431): p. 909-920.
173. Andersen, H.C., *Molecular-Dynamics Simulations at Constant Pressure and-or Temperature*. Journal of Chemical Physics, 1980. **72**(4): p. 2384-2393.
174. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features*. Biopolymers, 1983. **22**(12): p. 2577-637.
175. Moon, S. and D.A. Case, *A new model for chemical shifts of amide hydrogens in proteins*. J Biomol NMR, 2007. **38**(2): p. 139-50.
176. Xu, X.P. and D.A. Case, *Automated prediction of 15N, 13Calpha, 13Cbeta and 13C' chemical shifts in proteins using a density functional database*. J Biomol NMR, 2001. **21**(4): p. 321-33.
177. Osapay, K. and D.A. Case, *Analysis of proton chemical shifts in regular secondary structure of proteins*. J Biomol NMR, 1994. **4**(2): p. 215-30.

178. Neal, S., et al., *Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts*. J Biomol NMR, 2003. **26**(3): p. 215-40.
179. Vuister, G.W. and A. BAX, *Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins*. Journal of the American Chemical Society, 1993. **115**(17): p. 7772-7777.
180. Peter, C., X. Daura, and W.F. van Gunsteren, *Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations*. J Biomol NMR, 2001. **20**(4): p. 297-310.
181. Lipari, G. and A. Szabo, *Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules.2. Analysis of Experimental Results*. Journal of the American Chemical Society, 1982. **104**(17): p. 4559-4570.
182. Lipari, G. and A. Szabo, *Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules.1. Theory and Range of Validity*. Journal of the American Chemical Society, 1982. **104**(17): p. 4546-4559.
183. Daura, X., et al., *Unfolded state of peptides*. Adv Protein Chem, 2002. **62**: p. 341-60.
184. Wishart, D.S., et al., *H-1, C-13 and N-15 Chemical-Shift Referencing in Biomolecular Nmr*. Journal of Biomolecular NMR, 1995. **6**(2): p. 135-140.
185. Merutka, G., H.J. Dyson, and P.E. Wright, *Random Coil H-1 Chemical-Shifts Obtained as a Function of Temperature and Trifluoroethanol Concentration for the Peptide Series Ggxxg*. Journal of Biomolecular Nmr, 1995. **5**(1): p. 14-24.
186. Feenstra, K.A., et al., *A comparison of methods for calculating NMR cross-relaxation rates (NOESY and ROESY intensities) in small peptides*. Journal of Biomolecular Nmr, 2002. **23**(3): p. 181-194.
187. Nanzer, A.P., W.F. van Gunsteren, and A.E. Torda, *Parameterization of time-averaged distance restraints in MD simulations*. Journal of Biomolecular NMR, 1995. **6**(3): p. 313-320.
188. Torda, A.E., R.M. Scheek, and W.F. van Gunsteren, *Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat*. Journal of Molecular Biology, 1990. **214**(1): p. 223-235.
189. Torda, A.E., R.M. Scheek, and W.F. van Gunsteren, *Time-dependent distance restraints in molecular-dynamics simulations*. Chemical Physics Letters, 1989. **157**(4): p. 289-294.
190. Marchut, A.J. and C.K. Hall, *Effects of chain length on the aggregation of model polyglutamine peptides: molecular dynamics simulations*. Proteins, 2007. **66**(1): p. 96-109.
191. Fink, A.L., *The aggregation and fibrillation of alpha-synuclein*. Acc Chem Res, 2006. **39**(9): p. 628-34.

Publishing Agreement

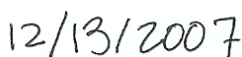
It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature



Date